

ROUTLEDGE



ROUTLEDGE
HANDBOOKS

The top half of the cover features a vibrant, abstract composition of colorful ink splashes. The colors include bright blue, deep red, fiery orange, and lush green, all set against a stark white background. The splashes are fluid and organic, creating a sense of movement and blending. The bottom half of the cover is a solid black rectangle, providing a high-contrast background for the white text.

The Routledge Handbook of Language Contact

Edited by Evangelia Adamou and Yaron Matras

The Routledge Handbook of Language Contact

The Routledge Handbook of Language Contact provides an overview of the state of the art of current research in contact linguistics. Presenting contact linguistics as an established field of investigation in its own right and featuring 26 chapters, this handbook brings together a broad range of approaches to contact linguistics, including:

- experimental and observational approaches and formal theories;
- a focus on social and cognitive factors that impact the outcome of language contact situations and bilingual language processing;
- the emergence of new languages and speech varieties in contact situations, and contact linguistic phenomena in urban speech and linguistic landscapes.

With contributions from an international range of leading and emerging scholars in their fields, the four sections of this text deal with methodological and theoretical approaches, the factors that condition and shape language contact, the impact of language contact on individuals, and language change, repertoires, and formation.

This handbook is an essential reference for anyone with an interest in language contact in particular regions of the world, including Anatolia, Eastern Polynesia, the Balkans, Asia, Melanesia, North America, and West Africa.

Evangelia Adamou is Senior Researcher at the CNRS (France). She specializes in the analysis of endangered languages with a focus on language contact and bilingualism, combining corpus and experimental methods. Recent publications include *A Corpus-Driven Approach to Language Contact* (2016, De Gruyter Mouton) and *The Adaptive Bilingual Mind* (under contract, Cambridge University Press).

Yaron Matras is Professor of Linguistics at the University of Manchester. His research interests include contact linguistics, urban multilingualism, typology, and language documentation. He has worked on dialects of Romani, German, Kurdish, and other languages, and is the founder of the Multilingual Manchester research unit.

Routledge Handbooks in Linguistics

Routledge Handbooks in Linguistics provide overviews of a whole subject area or sub-discipline in linguistics, and survey the state of the discipline including emerging and cutting edge areas. Edited by leading scholars, these volumes include contributions from key academics from around the world and are essential reading for both advanced undergraduate and postgraduate students.

The Routledge Handbook of Discourse Processes, Second Edition
Edited by Michael F. Schober, David N. Rapp, and M. Anne Britt

The Routledge Handbook of Phonetics
Edited by William F. Katz and Peter F. Assmann

The Routledge Handbook of Vocabulary Studies
Edited by Stuart Webb

The Routledge Handbook of North American Languages
Edited by Daniel Siddiqi, Michael Barrie, Carrie Gillon, Jason D. Haugen and Éric Mathieu

The Routledge Handbook of Language and Science
Edited by David R. Gruber and Lynda Walsh

The Routledge Handbook of Language and Emotion
Edited by Sonya E. Pritzker, Janina Fenigsen, and James M. Wilce

The Routledge Handbook of Language Contact
Edited by Evangelia Adamou and Yaron Matras

Further titles in this series can be found online at www.routledge.com/series/RHIL

The Routledge Handbook of Language Contact

*Edited by Evangelia Adamou
and Yaron Matras*

First published 2021
by Routledge
2 Park Square, Milton Park, Abingdon, Oxon OX14 4RN

and by Routledge
52 Vanderbilt Avenue, New York, NY 10017

Routledge is an imprint of the Taylor & Francis Group, an informa business

© 2021 selection and editorial matter, Evangelia Adamou and Yaron Matras; individual chapters, the contributors

The right of Evangelia Adamou and Yaron Matras to be identified as the author of the editorial material, and of the authors for their individual chapters, has been asserted in accordance with sections 77 and 78 of the Copyright, Designs and Patents Act 1988.

All rights reserved. No part of this book may be reprinted or reproduced or utilised in any form or by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying and recording, or in any information storage or retrieval system, without permission in writing from the publishers.

Trademark notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

British Library Cataloguing-in-Publication Data

A catalogue record for this book is available from the British Library

Library of Congress Cataloging-in-Publication Data

Names: Adamou, Evangelia, editor. | Matras, Yaron, 1963– editor.

Title: The Routledge handbook of language contact / edited by Evangelia Adamou and Yaron Matras.

Description: London ; New York : Routledge, 2020. | Series: Routledge handbooks in linguistics | Includes bibliographical references and index. |

Identifiers: LCCN 2020001467 (print) | LCCN 2020001468 (ebook) | ISBN 9780815363552 (hardback) | ISBN 9781351109154 (ebook)

Subjects: LCSH: Languages in contact.

Classification: LCC P130.5 .R685 2020 (print) | LCC P130.5 (ebook) | DDC 306.44—dc23

LC record available at <https://lcn.loc.gov/2020001467>

LC ebook record available at <https://lcn.loc.gov/2020001468>

ISBN: 978-0-8153-6355-2 (hbk)

ISBN: 978-1-351-10915-4 (ebk)

Typeset in Times New Roman
by Apex CoVantage, LLC

Contents

| | |
|---|----------|
| <i>List of figures</i> | viii |
| <i>List of tables</i> | x |
| <i>List of contributors</i> | xi |
| | |
| Introduction to <i>The Routledge Handbook of Language Contact</i> <i>Evangelia Adamou and Yaron Matras</i> | 1 |
| | |
| PART 1 | |
| Methods and theoretical approaches | 5 |
| | |
| 1 Processing multilingual data <i>Barbara E. Bullock, Almeida Jacqueline Toribio, Jacqueline Serigos, and Gualberto Guzmán</i> | 7 |
| | |
| 2 Language contact in the lab <i>Paola E. Dussias, Judith F. Kroll, Melinda Fricke, and Michael A. Johns</i> | 28 |
| | |
| 3 A variationist perspective on language contact <i>Shana Poplack</i> | 46 |
| | |
| 4 The 4-M model: different routes in production for different morphemes <i>Janice L. Jake and Carol Myers-Scotton</i> | 63 |
| | |
| 5 Theoretical approaches to the grammar of code-switching <i>Jeff MacSwan</i> | 88 |
| | |
| 6 Usage-based approaches <i>Ad Backus</i> | 110 |

| | |
|--|------------|
| PART 2 | |
| Processes and dimensions | 127 |
| 7 Social factors <i>Kofi Yakpo</i> | 129 |
| 8 Language contact: pragmatic factors <i>Peter Auer</i> | 147 |
| 9 Cognitive factors of language contact <i>Kees de Bot and Lars Bülow</i> | 168 |
| 10 Typological factors <i>Felicity Meakins</i> | 185 |
| 11 Cross-language contact in the developing grammars of bilingual children <i>Jennifer Austin</i> | 201 |
| 12 First language attrition in the twenty-first century: how continued L1 contact in the digital age fuels language attrition theorizing <i>Merel Keijzer</i> | 221 |
| PART 3 | |
| Outcomes | 235 |
| 13 Borrowing <i>Yaron Matras and Evangelia Adamou</i> | 237 |
| 14 Code-switching and bilinguals' grammars <i>Rena Torres Cacoullos and Catherine E. Travis</i> | 252 |
| 15 Convergence <i>Björn Wiemer</i> | 276 |
| 16 Creoles and pidgins: why the latter are not the ancestors of the former <i>Salikoko S. Mufwene</i> | 300 |
| 17 Mixed Languages <i>Carmel O'Shannessy</i> | 325 |
| 18 Linguistic landscape and urban multilingualism <i>Carla Bagna, Monica Barni, and Martina Bellinzona</i> | 349 |

| | | |
|---------------|--|------------|
| 19 | Urban youth speech styles in multilingual settings <i>Margreet Dorleijn, Maarten Kossmann, and Jacomine Nortier</i> | 366 |
| PART 4 | | |
| | Linguistic areas | 383 |
| 20 | The Balkans <i>Victor A. Friedman</i> | 385 |
| 21 | Anatolia <i>Anaïd Donabedian and Ioanna Sitaridou</i> | 404 |
| 22 | Language contact in the Asian region <i>Umberto Ansaldo and Lisa Lim</i> | 434 |
| 23 | Eastern Polynesia <i>Mary Walworth</i> | 462 |
| 24 | Linguistic Melanesia <i>Antoinette Schapper</i> | 480 |
| 25 | Language contact in North America <i>Marianne Mithun</i> | 503 |
| 26 | Language contact in West Africa <i>Friederike Lüpke and Rachel Watson</i> | 528 |
| | <i>Index</i> | 551 |

Figures

| | | |
|-------|--|-----|
| 1.1 | Ordered distribution of spans in Czech-English conversation | 17 |
| 1.2 | Three-dimensional plot of corpora by burstiness * M-Index * I-Index | 21 |
| 5.1 | A minimalist model of bilingual code-switching | 101 |
| 10.1 | Myers-Scotton's 4-M model | 189 |
| 14.1 | Rates of word-initial /d/ lenition (left panel) and contextual distribution (right panel) according to frequency of occurrence in contexts favourable to lenition, for cognate and non-cognate words in a US Spanish variety | 256 |
| 14.2 | Shared social and cognitive properties of single-word insertions and multiword code-switching strings | 257 |
| 14.3 | Divergent structural linguistic properties of single-word insertions and multiword code-switching strings | 259 |
| 14.4 | Distribution of clauses by subject person in bilinguals' Spanish and English | 260 |
| 14.5 | Distribution of a linguistic variable according to the proximate presence of code-switching in a bilingual speech corpus | 264 |
| 14.6 | Interlingual similarities and differences in variation patterns: subject pronoun probabilities according to accessibility, in monolingual and bilingual varieties, and in presence vs. absence of code-switching | 267 |
| 14.7 | Priming across vs. within languages: subject pronoun probabilities in presence of CS, according to priming (realisation of the previous coreferential mention) | 269 |
| 21.1 | A schematic representation of the main dividing ethno-linguistic lines in Anatolia before the massive shifts of the early twentieth century | 408 |
| 21.2 | Four zones with epicentres and fade-outs | 411 |
| 23.1 | Map of the Polynesian Triangle | 463 |
| 23.2 | Eastern Polynesian linguistic relationships | 464 |
| 24.1 | Approximate location of Linguistic Melanesia | 481 |
| 24.2 | Distribution of Papuan languages on and around New Guinea | 481 |
| 24.3 | Presence of tonal contrasts in Linguistic Melanesia | 483 |
| 24.4 | Presence of one liquid phoneme in Linguistic Melanesia | 485 |
| 24.5 | Absence of velar nasal phonemes in Linguistic Melanesia | 486 |
| 24.6 | Dominant order of subject, verb, and object in Linguistic Melanesia | 487 |
| 24.7 | Dominant order of noun and genitive in Linguistic Melanesia | 488 |
| 24.8 | Dominant order of noun and numeral in Linguistic Melanesia | 488 |
| 24.9 | Colexification of 'fire' and 'firewood' in Linguistic Melanesia | 489 |
| 24.10 | Comparative constructions in Linguistic Melanesia | 490 |
| 24.11 | Numeral bases and counting systems in Linguistic Melanesia | 491 |

| | | |
|------|---|-----|
| 25.1 | Northwest Coast | 507 |
| 25.2 | California | 508 |
| 25.3 | Southeast | 514 |
| 26.1 | Roles of named languages in GS's repertoires in 1995 | 534 |
| 26.2 | Roles of named languages in GS's repertoires in 2015 | 534 |
| 26.3 | Influence of perspective on linguistic categorization | 541 |

Tables

| | | |
|------|--|-----|
| 1.1 | Komi-Russian annotations | 13 |
| 1.2 | Czech-English data | 16 |
| 19.1 | Examples from Yuba UYSS | 373 |
| 20.1 | The differences between a language family and a sprachbund | 390 |
| 21.1 | Pairs of phonemic isoglosses defining Anatolian sub-areas | 413 |
| 21.2 | Discourse particles in Western Anatolian languages | 421 |
| 22.1 | Recent Lankan contact features | 437 |
| 22.2 | Core case in SLM, Sinhala and Lankan Tamil | 440 |
| 23.1 | Consonant inventories of some Eastern Polynesian languages | 464 |
| 26.1 | Joola greetings: ‘peace only’ | 535 |
| 26.2 | Forms for ‘water’ in Joola Kujireray | 537 |

Contributors

Evangelia Adamou is Senior Researcher at the CNRS (France). She specializes in the analysis of endangered languages with a focus on language contact and bilingualism, combining corpus and experimental methods. Recent publications include *A Corpus-Driven Approach to Language Contact* (2016, De Gruyter Mouton) and *The Adaptive Bilingual Mind* (under contract, Cambridge University Press).

Umberto Ansaldo is Professor of Linguistics and Head of the School of Literature, Art, and Media at The University of Sydney. His interests cover languages and societies of the Asian region, linguistic diversity and human migration, and language contact and change. He is founding co-editor of the journal *Language Ecology*.

Peter Auer is Full Professor of Germanic Linguistics at the University of Freiburg (Germany). His research interests include bilingualism, sociolinguistic variation, interactional linguistics, prosody and syntax of conversational language, and gaze in conversation.

Jennifer Austin is a professor in the Department of Spanish and Portuguese Studies at Rutgers University, Newark. Her research interests include language acquisition, bilingualism, and the effects of language contact, and she has published articles on the acquisition of Basque, English, and Spanish. She is also a co-author of the book *Bilingualism in the Spanish-Speaking World* (Cambridge University Press, 2015) together with María Blume and Liliana Sánchez.

Ad Backus is a professor of linguistics and sociolinguistics at Tilburg University in the Netherlands. He has worked mostly on contact effects in Turkish as spoken in the Netherlands and tries to combine contact linguistics and usage-based linguistics. He has recently published an edited volume (Zenner, Backus and Winter-Froemel, 2019) and is the chief editor of *International Journal of Bilingualism*.

Carla Bagna is Associate Professor in Educational Linguistics at the University for Foreigners of Siena. Her main research interests are second language acquisition, teaching and learning of Italian as a foreign language (in Italy and abroad), and immigrant languages in the Italian Linguistic Landscape.

Monica Barni is Full Professor in Educational Linguistics at the University for Foreigners of Siena. Her research activity is mainly oriented towards theoretical and methodological issues of language testing, language contact, and immigrant languages in Italian society and school, also through the approach of Linguistic Landscape.

Martina Bellinzona is a PhD researcher at the University for Foreigners of Siena. Her current research focuses on language policy and linguistic schoolscapes in the Italian school context. Her research interests include mainly educational linguistics, migratory flows, and Linguistic Landscape.

Barbara E. Bullock is Professor of Linguistics in the Department of French and Italian at the University of Texas. Her scholarship has centred on the structural effects of bilingualism and language contact at all levels of linguistic analysis.

Lars Bülow is Professor at the University of Vienna. His current research focuses on Bavarian dialects of Austria, exploring variation, language change, and contact.

Kees de Bot was Chair of Applied Linguistics at the University of Groningen. He is now professor at the University of Pannonia. His research ranges from psycholinguistic aspects of multilingualism to language policy. In more recent work, he focuses on the application of complex dynamic systems theory in applied linguistics.

Anaïd Donabedian is Professor in Armenian Linguistics at the Institut National des Langues et Civilisations Orientales, member of the research laboratory Structure et Dynamique des Langues (SeDyL UMR8202), and editor of *Faits de Langues* (Brill). Her research focuses on the typology and variation of Modern Armenian, particularly the discourse-syntax interface.

Margreet Dorleijn (PhD Amsterdam 1996) is Assistant Professor at the University of Amsterdam. In the past, she worked on Turkish-Kurdish Language contact; presently she focuses on Turkish-Dutch language contact and on linguistic practices of young people with a migrant background in the Netherlands.

Paola E. Dussias is Professor of Spanish, Linguistics and Psychology at Pennsylvania State University. The research that she and her students conduct employs methodological tools from linguistics, experimental psycholinguistics, and cognitive neuroscience to examine the way in which bilingual readers and speakers negotiate the presence of two languages in a single mind.

Melinda Fricke is an assistant professor in the Department of Linguistics at the University of Pittsburgh. Her work combines corpus and experimental methods to study the psycholinguistics of speech production and perception, often focusing on phonetic variation in bilingual speakers.

Victor A. Friedman (PhD in both linguistics and in Slavic languages and literatures, University of Chicago, 1975) is Andrew W. Mellon Distinguished Service Professor Emeritus in the Humanities at the University of Chicago and Honorary Adjunct at La Trobe University. His research interests include all aspects of the languages of the Balkans and the Caucasus.

Gualberto Guzmán is a graduate student in computer science at the University of Texas at Austin. He focuses on artificial intelligence and software design.

Janice L. Jake works primarily on language contact, specifically code-switching. She is interested in contact situations where there may be language shift and where this shift has

consequences for the grammars of the languages in contact. She is Chair of the English Department at Midlands Technical College, in Columbia, SC, USA.

Michael A. Johns is a PhD candidate and NSF graduate fellow in the Department of Spanish, Italian and Portuguese at Penn State University. His research examines bilingual code-switching, testing sociolinguistic and psycholinguistic theories of language production and processing by incorporating insights from sociolinguistic data in the design and implementation of lab-based experiments.

Merel Keijzer is a full professor of English linguistics and English as a second language at the University of Groningen. She is the PI of the Bilingualism and Aging Lab (BALAB). Her research interests focus on psycholinguistic approaches to bilingual language development across the lifespan, most notably in the third-age.

Maarten Kossmann (PhD, Leiden, 1994) is professor of Berber studies at Leiden University. He has published extensively about Berber synchronic and diachronic linguistics, about Berber in contact, and about Dutch as used by Moroccan heritage youth in the Netherlands.

Judith F. Kroll is Distinguished Professor of Psychology at the University of California, Riverside. Her research used the methods of cognitive neuroscience to examine the interplay between two languages during second language learning and once speakers are proficient bilinguals.

Lisa Lim is Associate Professor at the University of Sydney. Her research encompasses Asia, contact, shift, endangerment, postvernacularity, and the sociolinguistics of globalization. She is co-author of *Languages in Contact*, co-editor of *The Multilingual Citizen*, founding co-editor of *Language Ecology*, and ‘Language Matters’ columnist for the *South China Morning Post*.

Friederike Lüpke is Professor of Language Documentation and Description at SOAS, University of London. She has worked on the description of Mande and Atlantic languages. Currently, the focus of her research is on language contact and the description and documentation of one of West Africa’s most complex multilingual settings in Casamance, Senegal.

Jeff MacSwan is Professor of Applied Linguistics, Linguistics, and Neuroscience and Cognitive Science at the University of Maryland. His research program focuses on the linguistic study of bilingualism and code-switching and its implications for language education policy. He is a fellow of the American Educational Research Association.

Yaron Matras is Professor of Linguistics at the University of Manchester. His research interests include contact linguistics, urban multilingualism, typology, and language documentation. He has worked on dialects of Romani, German, Kurdish, and other languages, and is the founder of the Multilingual Manchester research unit.

Felicity Meakins, Associate Professor, is an Australian Research Council (ARC) Future Fellow in Linguistics at the University of Queensland and a chief investigator on the ARC Centre of Excellence for the Dynamics of Language. She is a field linguist who specializes in the documentation of Australian languages and the effect of English on these languages.

Marianne Mithun is Professor of Linguistics at the University of California, Santa Barbara. Her work ranges over morphology, syntax, discourse, prosody, and their interrelations; language contact and language change; typology and universals; language documentation; and particularly languages indigenous to North America and Austronesia.

Salikoko S. Mufwene is the Frank J. McLoraine Distinguished Service Professor of Linguistics and the College at the University of Chicago. He practises evolutionary linguistics, focused on the phylogenetic emergence of languages and language speciation, including the emergence of creoles and other indigenized varieties of European languages in the colonies. His books include *The Ecology of Language Evolution* (Cambridge University Press, 2001) and *Language Evolution: Contact, Competition and Change* (Continuum Press, 2008).

Carol Myers-Scotton developed the Matrix Language Frame model to account for the unequal nature of bilingual structures. In addition, she has written about socio-pragmatics, especially with respect to code-switching. She is a distinguished emerita professor at the University of South Carolina and presently an adjunct professor at Michigan State University.

Jacomine Nortier (PhD, Amsterdam, 1989) is Associate Professor in Sociolinguistics, Multilingualism and Language Contact at Utrecht University. Earlier publications were concerned with code-switching between Dutch and Moroccan Arabic; more recently she publishes mostly on young people's language and identity matters.

Carmel O'Shannessy is a lecturer in linguistics in the College of Arts and Social Sciences at the Australian National University. Her research focuses on Australian languages, investigating questions of language contact, change and multilingual acquisition, especially with regard to change in Warlpiri and the emergence of Light Warlpiri.

Shana Poplack is Distinguished University Professor and Canada Research Chair in Linguistics, and Director of the Sociolinguistics Laboratory at the University of Ottawa. Her work applies theoretical and methodological insights gained from the study of linguistic variation and change to a variety of fields, including language contact, Romance linguistics, African American vernacular English, prescription and praxis, and the role of the school in impeding linguistic change.

Antoinette Schapper is a researcher in linguistics at Langues et Civilisations à Tradition Orale (Lacito), a unit of the CNRS. Her research interests centre on the description, typology, and history of the languages of Wallacea and western New Guinea.

Jacqueline Serigos, Assistant Professor of Spanish Linguistics in the Department of Modern and Classical Languages at George Mason University, investigates contact phenomena such as borrowing and calquing, drawing on insights from multiple linguistic subfields, including corpus linguistics, computational linguistics, sociolinguistics, and semantics, as well as the discipline of statistics.

Ioanna Sitaridou is Reader in Spanish and Historical Linguistics at the University of Cambridge and Fellow and Director of Studies in Linguistics and Modern and Medieval Languages at Queens' College, Cambridge. Her current research focuses on the Romance languages and

Asia Minor Greek (Romeyka, in particular), language change, language contact, and cue-based syntactic reconstruction.

Almeida Jacqueline Toribio is Professor of Linguistics in the Department of Spanish and Portuguese at the University of Texas. She pursues research on linguistic variation among monolinguals in rural communities and bilinguals in immigrant settings, with field and laboratory approaches and, most recently, methods of corpus linguistics.

Rena Torres Cacoullos is Professor of Spanish and Linguistics at the Pennsylvania State University and editor-in-chief of *Language Variation and Change*. She blends variationist and usage-based perspectives in quantitative analysis of natural production data in its social context. Her research is funded by the National Science Foundation.

Catherine E. Travis is Professor of Modern European Languages at the Australian National University, and a chief investigator in the Centre of Excellence for the Dynamics of Language. Her research addresses questions related to linguistic and social factors impacting on language variation and change, in particular in socially diverse communities.

Mary Walworth co-directs the Comparative Oceanic Languages project at the Max Planck Institute for the Science of Human History. Her research focuses on the evolution and interaction of the languages of Polynesia and Vanuatu and the socio-historical relationships between the speakers of these languages.

Rachel Watson is a postdoctoral research associate at SOAS, University of London. Her doctoral research comprised a description and documentation of Joola Kujireray, and her current research interests include theoretical and methodological issues involved in linguistic description in multilingual contexts.

Björn Wiemer is Professor of Slavic Linguistics at the University of Mainz, with expertise also in Lithuanian. Apart from areal and contact linguistics, he specializes in verbal categories, modality and evidentiality, clausal complementation, and lexical semantics. In all these fields, his research also covers diachrony and change.

Kofi Yakpo is Associate Professor of Linguistics at the University of Hong Kong. His research addresses the complex interaction of genealogical, areal, typological, and social forces in the evolution of contact languages, with a focus on the English-lexifier creoles of Africa and the Americas, and the Asian diaspora languages of the Caribbean.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Introduction to *The Routledge Handbook of Language Contact*

Evangelia Adamou and Yaron Matras

Rationale

With *The Routledge Handbook of Language Contact*, our goal is to provide an overview of the state of the art of current research into contact linguistics. We aim in the first instance to present contact linguistics as an established field of investigation in its own right, rather than (just) the application of general linguistic methods to data sets that represent contact. This, to our knowledge, is a first, and constitutes one of the volume's key features.

Our distinctive approach is manifested firstly in the way we bring together historical-typological approaches to contact linguistics with approaches that are often considered to be sociolinguistic or discourse based. This is exceptional in this field of study, where there is a tendency to find methodological, theoretical, and therefore also empirical separation between the study of bilingualism/multilingualism, and the study of historical contact-induced structural change and the cumulative effect of contact. This reflects our personal interest as editors in such a combined approach, as applied in our own research (Matras, 2009 outlining a theory of contact; Adamou, 2016 on a corpus-based approach to language contact). The Handbook thus brings together themes that are of interest to language typology and historical linguistics as well as to sociolinguistics and conversation and discourse analysis.

The volume also offers a novel form of organization, with separate areas of focus devoted to methods and theory (Part 1), the processes that shape contact and its different dimensions (Part 2), the outcomes of contact (Part 3), and the study of linguistic areas (Part 4). This follows directly from the overall rationale of the book, which is to flag contact linguistics as a field that has, since its inception, developed methods and theoretical models of its own. This contrasts with previous approaches that tended to justify the relevance of contact in relation to other (external) fields of investigation in linguistics (concentrating for instance on contact and historical linguistics, contact and grammatical theory, and so on). Through integrated parts on processes and outcomes, we highlight how similar circumstances give rise both to language mixing at the individual level and to the reshaping of language through contact, and how in turn the various outcomes of contact derive from a shared set of factors; this contributes further to presenting the wide array of contact phenomena as an integrated package, and from an epistemological point of view it represents the need to take a variety of factors into

consideration when assessing all different outcomes of contact. The final part on linguistic areas flags the fact that manifestations of contact are ubiquitous and that they are as relevant in defining relations among languages as phylogenetic and typological frameworks.

The volume is guided by the principle of combining theory with empirical approaches. Rather than list case studies and run the risk of a somewhat random selection, we have invited contributors to base their discussion on empirical evidence, and to make reference to published empirical case studies that further support the arguments made in the chapters.

Our list of contributors includes some of the most influential figures who have shaped the field of contact linguistics over the past decades. We invited them to outline their own models and position them in the context of other approaches, and to reflect on challenges to some of the principles and notions that have guided their own work which may have arisen as a result of recent developments in the field. We have also included researchers and teams of researchers who have in recent years contributed to redefining the methodological and theoretical boundaries of the field of contact linguistics and whose work promises to make an even more substantial contribution in the future. In this respect, the volume can be seen as a comprehensive representation of the field's established past, present, and future.

How to read the Handbook

The book is divided into four parts, which deal with *Methods and Theoretical Approaches*, *Processes and Dimensions* (referring to the processes that condition and shape language contact), *Outcomes* (outlining the impact of language contact on individuals, language change, language repertoires, and language formation), and *Linguistic Areas* (outlining the geographical spread of structural features through language contact).

In line with our ambition to present the field's 'past, present, and future,' the selection of contributions to the various parts reflects on the one hand essential areas that users would expect to be covered, and on the other hand newly emerging sub-fields of investigation. The former include major theoretical approaches in the field such as the variationist approach, the 4-M model, the constraint-based and constraint-free approaches, as well as the usage-based approach. Essential topics include the discussion of social, pragmatic, cognitive, and typological factors, and of the specificities of bilingual acquisition, as well as outcomes of language contact such as borrowing, code-switching, convergence, pidgins, creoles, and mixed languages. Newly emerging sub-fields of investigation include work with corpora and natural language processing methods, experimental approaches in the lab and in the field, urban multilingualism, multi-ethnolects, and first-language attrition. In its final part, the Handbook introduces 'classic' linguistic areas, such as the Balkans, as well as some lesser-known cases of areal convergence, such as Eastern Polynesia and Linguistic Melanesia.

Under the four parts, we present 26 chapters. In order to help the reader navigate the Handbook and ensure that the discussion is as exhaustive as possible, each chapter includes some of the following sections:

- 1 Introduction/definitions

This section conveys a descriptive introduction of the topic with definitions that are deemed essential for an understanding of the main topic of the chapter.

- 2 Historical overview

This section offers a historical overview of the topic under discussion. The goal is not just to offer a view of the history of research on the topic, but to offer readers an understanding of the field's progress and analytical issues involved in its study.

- 3 Critical issues and topics
This section provides an overview of critical issues as they arise in the research literature. It includes topics that may have been the locus of controversies by offering, as much as possible, arguments on all sides.
- 4 Current contributions and research
This section offers the opportunity to read an overview of recent studies, their focus, and their results. It is accompanied by a critical discussion that allows to demonstrate how research has been evolving in recent years, as well as the main questions and conclusions that have been put forward by various authors.
- 5 Main research methods
Where appropriate, research methods are discussed, including a consideration of their merits and pitfalls.
- 6 Future directions
In this section, authors address new fields of investigation that have been receiving growing interest, and introduce what, in their opinion, remains to be done in the future.
- 7 Further reading
This section provides a list of key further readings, consisting of bibliographic entries with a short description of their contents.
- 8 Related topics
This section offers some keywords for topics connected to the main topic of the chapter. These keywords refer specifically to the other chapters in the Handbook that readers might find useful.
- 9 References
A full list of all the bibliographical references cited in the text.

In conclusion, we hope that *The Routledge Handbook of Language Contact* succeeds in reflecting the state of the art and that it offers comprehensive coverage of key issues in the study of language contact.

Finally, we wish to thank the contributors to this Handbook, Routledge publishers for offering us the opportunity to position the book in their Handbook series, and lastly Beth Lamarra for support with the production process.

Paris and Manchester, December 2019

References

- Adamou, E. (2016). *A corpus-driven approach to language contact: Endangered languages in a comparative perspective*. Boston and Berlin: De Gruyter Mouton.
- Matras, Y. (2009). *Language contact*. Cambridge: Cambridge University Press.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Part 1

Methods and theoretical approaches



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Processing multilingual data

*Barbara E. Bullock, Almeida Jacqueline Toribio,
Jacqueline Serigos, and Gualberto Guzmán*

1. Introduction and definitions

Patterns and typologies of language mixing are a central theme in contact linguistics (cf., Thomason and Kaufman, 1988; Myers-Scotton, 1993; Muysken, 2000; Winford, 2003). Scholars use various techniques to study the influence of the linguistic, social, and individual factors on language mixing in natural interactions and in a range of other linguistic genres, including medieval Macaronic sermons and poetry (e.g., Schulz and Keller, 2016; Demo, 2018) and multilingual hashtags on the Internet (e.g., Jurgens, Dimitrov and Ruths, 2014). In linguistic analyses of the morphosyntactic implications of multilingual practices, patterns of language mixing have been described for a specific language or across languages using selected examples with accompanying qualitative descriptions. However, in an age of ‘big data,’ new techniques for analyzing larger, and more diverse, language samples are called for. Irrespective of the size of the corpus analyzed, especially valuable for testing general proposals regarding language mixing would be the ability to characterize and compare mixed corpora in a systematic way, thereby affording information into the range of diversity of code-switching across languages (see Adamou, 2016).

Methods for the quantification of language mixing are essential, as noted by several linguistic scholars of language contact (cf., Barnett et al., 2000; Bullock, Guzmán and Toribio, 2019; Donnelly and Deuchar, 2011; Adamou, 2016; Guzmán et al., 2016, 2017; Myslin and Levy, 2015). Quantitative comparisons using standardized metrics are applicable to any type of data and the predictive information that they yield may facilitate the development of language technologies (dictionaries, taggers, parsers) that can accommodate multilingual speech. The development of tools for the automatic processing of mixed speech data will broaden digital access and representation for multilinguals, including speakers of endangered languages (Trosterud, 2006; Partanen et al., 2018). Towards achieving these aims, the present chapter outlines a computational approach to multilingual data that permits the quantification and modelling of language mixing in a unified way. Importantly, the tools and techniques discussed here are *language-independent*, i.e., they can be applied to data from any language combination irrespective of the typologies they are drawn from.

Why do we seek to quantify code-switching? Ideally, linguists want to be able to gauge the frequency and degree to which languages in contact interact at the level of their constituent grammars. For this purpose, it is necessary to record the language of each token in a sequence so that the ratio of languages and the probability and regularity of switching between them can be determined. Although there are points of ambiguity in discerning which language a token may belong to – for example, cognates, borrowings, proper names, numbers, and mixed lexemes (in which stems and affixes are drawn from different languages) – in most cases, humans who are trained to manually annotate texts are able to confidently label a token with its etymological source even in contexts of unmarked mixing such as contemporary Hindi–English (Diab and Kamboj, 2011). A prerequisite for processing multilingual data for linguistic research is the means to automatically identify the language of a token in an accurate, consistent, and cost-effective fashion. Automatic language annotation of tokens is a first step to a series of other language processing tasks, like part-of-speech (POS) annotation and syntactic parsing that are necessary for the exploration of the interaction of two or more grammars in multilingual speech. Language annotation is also essential for modelling the complexity and the time-course or intermittency of language mixing. Making progress in building tools and techniques to process multilingual data will translate directly into improved insights into language contact phenomena. In turn, linguistic insights into the statistical regularities of code mixing will improve these same tools and techniques, in particular, methods that are built on machine learning models that are trained to accurately annotate a text with little human intervention.

The tools and techniques to automatically process multilingual data are being continually refined, and even so, it is unlikely that mixed language data will ever be annotated with the speed or accuracy of non-mixed language data from major languages like English, Russian, or Chinese. An obstacle to the pursuit of adequate tools for the processing of multilingual data is the scarcity of texts from contact languages. Vast repositories of text data are required to train the statistical models that underlie the computational tools for automatic linguistic processing (e.g., tokenizers, dictionaries, POS taggers, parsers). Even if the languages in contact under study are widely spoken ones for which large data resources are available, such as Hindi and English, their mixed language versions are considered *low-resource varieties* (Sitaram and Black, 2016), because they have ‘fewer technologies and especially data sets’ (Cieri et al., 2016, p. 3). While small data sets may not be sufficient to accurately train statistical models, they do offer important information for linguists. They reveal patterns that can be quantified and compared to other data sets, yielding new insights regarding the potential limits to language mixing.

Scholars pursue questions of multilingualism and language contact from different frames of reference. As a consequence, the data that they gather and analyze are diverse and, in some subfields, comprise non-linguistic observations such as survey responses or measures of reaction times and brain responses (e.g., eye movements and event-related potentials). The disciplinary perspective that guides this chapter is Natural Language Processing (NLP). NLP models can be informed by the non-linguistic results of experiments and surveys conducted with human participants, but they are primarily trained on raw data in the form of naturally produced texts. The specifics of processing raw multilingual data and the benefits that this would yield for advancing novel findings in language contact is the focus of this contribution.

The chapter is organized as follows. Section 2 presents the problems posed by a traditional qualitative approach to contact data. The critical issues of linguistic data processing – data

resources, annotations schemes, and automated techniques – applied to multilingual data are reviewed in Section 3. Section 4 lays out a language-independent method for investigating language mixing across corpora and demonstrates how it generates new findings and opens potential avenues of enquiry. The chapter concludes in Section 5 with a discussion of a future line of research that couples language contact with computational and data sciences.

2. Historical overview

Much of the research in contact linguistics to date consists of qualitative analyses of data that is not publicly available. Traditionally, a few examples from a text or corpus are selected to illustrate particular patterns of language mixing that are presumed to be representative of overarching trends. Yet, information regarding the numeric frequency of such patterns within the data sets analyzed is rarely provided. Without this information, it is not possible to ascertain the degree to which observations of language mixing throughout or across data sets are alike or different or whether the observations are sporadic or regular. The result is that theoretical proposals in the field can be only weakly supported. For instance, Muysken's (2000) typology of mixing is intended to make generalizations about structural patterns and the socio-historical contexts that give rise to them. He posits two main categories: *insertion*, where a guest language sequence is embedded into the grammar of a matrix language, and *alternation*, where the grammars of each language frame the structure of the utterance. While providing a useful shorthand for language mixing phenomena, the classification of mixing types often rests on subjective evaluations of data rather than on theoretically neutral classifications. In practice, it is not always clear whether a given utterance illustrates insertion, alternation, or something else (e.g., congruent lexicalization, lexical borrowing, tagging, or flagging). Similarly, it is often difficult to definitively determine which of the languages in interaction serves as the matrix language (Myers-Scotton, 1993; Winford, 2003; Auer and Muhamedova, 2005; Liu, 2008; Bhat, Choudhury and Bali, 2016). Using only qualitative classification schemes, typological comparison remains elusive. And since the totality of utterances in a data set are rarely studied, the predominance of any one pattern and the regularity with which diverse mixing patterns occur within a data set remain unknown.

The culture of producing qualitative research in language contact extends to the levels of morphosyntactic and sub-lexical analysis, with scholars largely regarding structural models of language mixing as absolute rather than as probabilistic. The Equivalence Constraint (Poplack, 1980), the Matrix Language Frame Hypothesis (Myers-Scotton, 1993), the Functional Head Constraint (Belazi, Rubin and Toribio, 1994), and other models of grammatical interaction have been variously bolstered and refuted on the basis of a few hand-selected examples. The result is that the issue of their efficacy as pertinent probabilistic generalizations about language mixing remains undecided even decades later (but see recent corpus-based, quantitative studies by Herring et al., 2010; Bhat, Choudhury and Bali, 2016; Parafita Couto and Gullberg, 2017; Bullock et al., 2018).

3. Critical issues and topics

Sources of data

One impediment to the pursuit of a quantitative approach to language contact is the absence of linguistically annotated data. Big data sets permit the development of training data, which

are required for developing accurate statistical language models that assign probabilities to linguistic sequences and permit the extrapolation of specific linguistic features particular to groups of speakers. In order to scale up to rigorously explore the linguistic and social factors that contribute to patterns of language mixing at the macro (group) and micro (individual) levels, we need data resources across a range of languages, in formats that make them amenable to quantitative analyses, and potentially of a size that will allow for the establishment of effective training data.

Large corpora are constructed to be representative of a language or language variety in initiatives that are often funded by government agencies with an interest in projecting a standardized variety. Elements of non-normative variation, like sequences of text in a foreign language that would be useful for language contact analyses, are commonly excluded in these corpora. While there have been numerous efforts to collect samples of mixed language speech corpora among linguists, these efforts are typically restricted to small numbers of participants because of the cost and time required for manual transcription and annotation. For example, the BillingBank, which forms part of the TalkBank repository (MacWhinney, 2007), currently houses around a dozen bilingual corpora, most comprised of interviews ranging from just 10 to 100 interviewees, the exception being the Køge Corpus, which boasts 291 interviews. Archives of endangered languages, like the Pangloss Collection (Michailovsky et al., 2014), can be rich in language mixing as are those for heritage languages, such as the Heritage Language Variation and Change in Toronto project (Nagy, 2011) or the Corpus of American Nordic Speech (Johannessen, 2015), and learner corpora, such as the International Corpus of Learner English (Granger et al., 2009). However crucial these resources are, they are still small in size relative to the resources available for monolingual varieties. For example, the South East Asian – Mandarin English corpus (Lyu et al., 2010, 2015) is likely to be the one of larger code-switching corpora available. Currently at 192 hours of recorded conversation between 156 speakers, according to its catalogue description in the Linguistic Data Consortium (LDC), it is still under 2 million words. In contrast, the British National Corpus (BNC), which is a monolingual corpus of English, exceeds 100 million words.

Outside of the linguistic research community, there has been little imperative to mine mixed language data. To date, the technical and scientific resources of commercial technology companies have been dedicated to developing tools for the languages that have the largest web presence, i.e., English, Chinese, Spanish, and Arabic (Streiter, Scannell and Steuffesser, 2006), though some speech scientists have begun to turn attention to developing technologies for multilingual language use as evidenced in the theme of the Interspeech 2018 Conference held in Hyderabad, India, *Speech Research for Emerging Markets in Multilingual Societies*. Still, it falls largely to linguists and NLP scholars to pursue alternative, but potentially larger, sources of naturally produced linguistic data for the documentation of low-resource varieties. The challenges and benefits of doing so are reviewed briefly here (and see Çetinoğlu, Schulz and Vu, 2016 for a discussion specific to NLP).

One common source of data for the exploration of variation in natural language within the NLP literature is the Internet, including sites for the harvesting of subtitles and song lyrics, discussion forums like Reddit, video-sharing sites like YouTube, and postings on social media such as Facebook (Bali et al., 2014) and, especially, Twitter (Lignos and Marcus, 2013; Jamatia, Gambäck and Das, 2015; Solorio et al., 2014; Vilares, Alonso and Gómez-Rodríguez, 2016). On the one hand, these sources of data can provide large quantities of text in various language combinations within relatively short amounts of time and at a minimal cost. However, the ease of access to these types of online data can vary greatly. Twitter data, for

example, used to be readily accessible, as it allowed third parties to share data through user-friendly interfaces. Twitter has ended this practice, and users must now access data through an API, requiring programming knowledge for those wishing to collect data from this platform. Additionally, there are limitations as to how many tweets can be collected and how they can be shared. Many online sources require a researcher to have programming knowledge in order to scrape and process data. Scholars are sometimes granted permission to publicly share the resulting repositories so that the one-time effort to gather the data yields a permanent resource, as is the case of the published ACTIV-ES corpus of Spanish subtitles (Francom, Hulden and Ussishkin, 2014) and the language-independent OpenSubtitle corpora (Lison and Tiedemann, 2016).

While acquiring the skills or securing the assistance necessary to scrape data from the Internet is one challenge, finding mixed language data in the first place is not an easy task. Presently, there are no means to generically search internet platforms for all posts, tweets, or utterances containing two or more targeted languages, so researchers need to use their knowledge about language creatively. For instance, the *codeswitchador* of Lignos and Marcus (2013) uses a model to probabilistically identify a word within an existing Twitter corpus as Spanish or English and then classifies tweets as code-switched if they contain a threshold of both languages. And Solorio et al. (2014) created three data sets for the First Shared Task on Language Identification in Code-Switched Data by sampling social media posts from different populations known for code-switching. The Spanish-English data set was generated by gathering tweets from geographical areas with high concentrations of Spanish-English bilinguals (California, Texas, Miami, and New York), the Nepali – English data set was developed by manually identifying users who switched frequently and sampling the tweets of the users that appear in their mentions, and the Hindi-English corpus was compiled from posts and comments from the Facebook pages of prominent public figures in India to which bilingual users are known to frequently contribute.

Another issue that often confronts linguists attempting to work with internet data is that of quality (Dorleijn and Nortier, 2009). Rife with hashtags, stylistic irregularities (e.g., *sooo excited*), incorrect graphematic realizations (e.g., *there friend* instead of *their friend*), and out-of-vocabulary words like acronyms and abbreviations (e.g., *FOMO* (*fear of missing out*), French *biz* (*bisoux* ‘kisses’)), internet data often requires extensive post-processing to normalize the text, a crucial step in the NLP pipeline prior to higher-level annotations, such as POS tagging or syntactic parsing. For some language pairings, ad hoc and variable vernacular romanizations are used in lieu of the native scripts (Ball and Garrette, 2018). Even languages employing the same alphabet, such as English and Spanish, can present problems when their character encoding contrasts (e.g., UTF-8 vs ASCII). In addition to the quality of the data, the content itself may differ greatly from the type of data that we have come to expect from traditional linguistic corpora in terms of style, length, or even discourse. Twitter users, for example, address an unspecified audience of known and unknown individuals, a context that distinguishes this communication medium from face-to-face interaction (Marwick and Boyd, 2011).

Mixed-language data can also be artificially generated. This type of data is created by computational linguists in need of large amounts of training or test data for building language models (Solorio and Liu, 2008a, 2008b; Bhat, Choudhury and Bali, 2016; Zhang et al., 2018). Consider the mixed variants of the same sentence shown in example 1, machine generated by Solorio and Liu (2008a) using different statistical classification models. The first two were produced by two separate statistical algorithms for determining the most likely sequence of

switches, the third chooses switches at random, and the last is provided as comparison to human performance.

1. a. Naive Bayes classifier
Pero siendo this a new year, es tiempo de empezar de nuevo que no?
- b. Value Feature Interval classifier
But this being a new year, *es tiempo de empezar* over isn't it ?
- c. Random
But this being a new *año*, it's *tiempo* to start over isn't it?
- d. Human
Pero this being a new year, it's a time to start over *que no?*
'But this being a new year, it's a time to start over, isn't it?'

This data presents limited vocabulary because the sentences are all generated from the same base text and can contain switching that may be judged as inauthentic. Solorio and Liu subjected the sentences to evaluation by bilingual speakers and found that those generated by Naive Bayes were rated most similar to those naturally produced, and much higher than those generated by the Random classifier. In other words, a statistical classification model trained on very little data can begin to approximate human performance.

Annotation of data

Whether data is collected from field interviews or online sites, or is generated with NLP models, there is a need for it to be uniformly annotated to draw robust comparisons between data sets. The number of annotation levels for multilingual data differs according to the needs of the analyst, but two crucial levels for considering language mixing are the language and POS tags, such that every token in a data set bears these labels. Since languages vary in their morphosyntactic typologies, there is a movement among NLP scholars with interests in language mixing to adopt the POS tagset used in the Universal Dependencies (UD) project (McDonald et al., 2013). The UD project envisions a consistent treebank annotation for all languages and includes a universal set of 17 POS tags that are superordinate to a wider set of language-particular tags (Nivre et al., 2016). This project is rapidly gaining traction for multilingual and language-mixed data and is continuing to develop more language-particular tagsets. The universal POS tags are a necessary first step in creating another layer of annotation, a dependency parse, which reflects the syntactic relations between tokens. Partanen et al. (2018) have recently conducted a successful test of multilingual dependency parsing on Russian mixed with the endangered Komi-Zyrian language. They used an existing Finnish dependency-parsed training data set to model Komi grammar. The fact that the Finnish parser provides a reasonable parse for the endangered language that has, as yet, no dependency parser, demonstrates the utility of a tagset that is designed for broad coverage.

Annotated data can be rendered most useful to the field when data files conform to standard conventions. Standard formats are easier for others to work with and encourage the sharing of data. As there are numerous ways to format data files, we will highlight one specific standard that has been developed by linguists and describe the general conventions it employs. The modified ConLL-U annotation standard (<http://universaldependencies.org/format.html>) offers a simple and flexible format that allows for comparison across multilingual texts (Buchholz and Marsi, 2006). The current ConLL-U scheme assumes that the data to be annotated is drawn from a single language; however, it can be easily adapted to include tags for languages.

For example, in annotating Komi-Russian data, Partanen et al. (2018) augment the CoNLL-U file with the tag `Lang`, for which the ISO language values are defined for Komi-Zyrian (`kpv`), Russian (`rus`), and Mixed, as shown in Table 1.1.

As for the data files themselves, plain text files that can be generated using any built-in text editor offer the advantage of being both human- and machine-readable and can be structured using a markup language, such as XLM, or delimited to create rows and columns using a csv (comma separated values) or tsv (tab separated values) format. While PDFs and Word documents are user-friendly and familiar, they are not easily handled by machines and therefore are not recommended for storing linguistic data. Another beneficial practice is to specify the character encoding at the start of each data file. Character encoding is a convention that ensures that data is displayed properly; it's especially important for handling scripts that include diacritics and non-Roman characters. Knowing the text encoding assists those working with the data to avoid common display problems that often arise when working with multiple scripts. Currently the standard character encoding is UTF-8 because it can encode virtually any character, and therefore writing system, of interest. In formatting a text file for processing, it is standard practice to place each linguistic token, including punctuation, on a separate row. Each token is then tagged for every field of annotation, using an unambiguous coding scheme, in a tab-separated, column-based format, as shown in Table 1.1. Ideally, the language annotation field should use the proper ISO language code.

When POS tags are used, it is important to note the treebank system that researchers adopt for each language and, if the coding has been done by hand, it is equally important to publish the protocols for selecting the proper annotation of ambiguous tokens and for the tokenization process. In the Komi-Russian annotation in Table 1.1, for instance, the Russian tag is used for words that contain any Russian morphology where other analysts may have chosen to employ a Mixed tag. If the annotation protocol is novel, it needs to be documented in enough detail to allow for it to be replicated or mapped to other systems. Additionally, the level or levels of annotation used should be specified as metadata: morpheme, word, intonational unit, sentence, or conversational turn. Finally, any special tag, such as Named Entity or Lemma, that the annotators use can be defined and placed in its own field.

Ideally, for replicable comparisons, corpora need to be annotated identically. For instance, it proves difficult to reconcile data sets if clitics are tokenized independently of their hosts in

Table 1.1 Komi-Russian annotations

| <i>Index</i> | <i>Token</i> | <i>POS</i> | <i>Dependency Parse</i> | <i>Language</i> |
|--------------|--------------|------------|-------------------------|--------------------------|
| 1 | Рӧдитчи | VERB | Root | Lang=Mixed |
| 2 | ме | PRON | Nominal subject | – |
| 3 | шейсят | NUM | Numeric modifier | Lang=Rus |
| 4 | четвѣртэй | DET | Numeric modifier | Lang=Rus |
| 5 | годын | NOUN | Oblique nominal | Lang=Mixed |
| 6 | октяб | NOUN | Nominal modifier | Lang=Mixed |
| 7 | тӧлысе | NOUN | Oblique nominal | – |
| 8 | тундраын | NOUN | Oblique nominal | Lang=Mixed SpaceAfter=No |
| 9 | . | PUNCT | Punctuation | – |

Source: GitHub, Niko Partanen, <https://github.com/nikopartanen>.

one data set and as part of the host in another. And hand annotations, no matter how carefully realized, can be subject to human error and bias, which vary from one data set to the next. The best way to ensure cross-corpus comparability is to automatically process the raw texts using the same tools. To be sure, there are errors in automatically annotated data, but the procedures can be refined to bring these errors in line with the range of accuracy of human annotators.

Automated annotation methods

Automated methods such as tokenizers, POS taggers, and dependency parsers dramatically accelerate the process of linguistic annotation. They were designed for large monolingual corpora, where potentially billions of words may be available for training, and they can show impressive rates of accuracy on annotation tasks. Scaling down to smaller, multilingual corpora presents challenges because the model must tag tokens that it has likely not encountered in the training data, leading to incorrect, or noisy, annotations. Small data also fail to provide sufficient information to learn transition probabilities between bigrams (i.e., how likely consecutive pairs of words are), which are advantageous in constructing language models that assign probabilities to phrases and utterances. Furthermore, training language models on small data sets renders them vulnerable to ‘memorizing’ the training data instead of generalizing from it since there are few linguistic sequences to learn from. The risk is that the model will be overfit and will fail to accurately process other corpora that is unlike the training data.

Of course, the aforementioned challenges apply to small monolingual and multilingual data sets alike. Fortunately, scholars have demonstrated that models can learn from monolingual data in spite of its sparsity (Garrette and Baldrige, 2013) and some of these methods could be extended to noisy, mixed language data (Ball and Garrette, 2018). Automation has become a viable option for some multilingual data processing tasks, especially word-level language identification, and research is ongoing for others, including language modelling (Adel et al., 2013, 2014), Named Entity (proper noun) classification, POS tagging, syntactic parsing, machine translation, and automatic speech recognition.

Word-level language identification has grown out of NLP research on document-level language identification. At the level of the document, there are numerous features for language identification models to consider, such as character encoding, characteristic letter sequences, correlation between word and part of speech, and closed grammatical classes. However, at the level of the word, the features are reduced to the character sequence of a given token and information about the preceding n tokens in an n -gram model. As such, most basic word-level language identification systems use a combination of character n -gram models, which learn the statistical probabilities of character sequences within a given language, and Hidden Markov Models, which learn the probabilities of consecutive pairs of words. Recently, Zhang et al. (2018) have introduced a language identification model that is explicitly designed to rapidly and accurately annotate the language of word tokens in multilingual documents with a more complicated approach that utilizes neural networks. Testing their system on computer-generated code-mixed data the researchers achieved accuracy rates ranging from 87.4% on mixed language/mixed script data (Devanagari Hindi/Romanized Hindi/English) to 98.4% accuracy on Devanagari Hindi–English and 92.4% on English–Spanish code-switched data. We note that accuracy rates above 90% are within the range of the reliability of human annotators.

The issue of POS tagging for multilingual text is still considered to be a work-in-progress for the NLP community as accuracy rates for existing multilingual taggers hover well below

the rates of monolingual taggers, which are considered to be highly effective (Baldwin and Lui, 2010; Lui and Baldwin, 2012). A major issue in developing POS taggers is that of tagsets. POS tagsets can often be highly specialized, such as the Penn TreeBank Tagset with 36 tags (Marcus, Santorini and Marcinkiewicz, 1993) and the BNC Basic Tagset for English with 61 tags (Leech, Garside and Bryant, 1994). While a coarse tagset makes the automation of a POS tagger much more feasible and accurate, it may compromise the information needed by researchers, since distinct tags in a robust monolingual tagset may be collapsed in a less fine-grained set. For example, the Penn TreeBank tagset differentiates four categories of nouns – singular or mass nouns (NN), plural nouns (NNS), proper singular noun (NNP), and proper plural nouns (NNPS) – whereas the basic UD tagset uniformly classifies these as NOUN. One solution for researchers wishing to preserve distinctions lost in the UD tagset is to map to the coarse UD tagset from a field of annotation of more fine-grained tags.

In spite of these computational advances, a major challenge for researchers of language contact remains that of access and usability of extant tools. Many of the multilingual tools developed remain inaccessible to contact linguists due to their proprietary nature, such as those developed by large tech companies, or due to their steep learning curve as most tools are designed for developers or coders. As tools are created and adapted for linguistic research, it is paramount for researchers to push for greater accessibility, via data and code-sharing platforms with clear instructions designed for those with limited coding background.

4. Current contributions and research

In agreement with Adamou (2016) we submit that quantitative measures offer researchers the only objective means to compare language mixing across corpora and subcorpora. Descriptors such as ‘heavily code-switched’ or typological characterizations of mixing as ‘insertional’ or ‘alternational’ need to be augmented with numerical reference points that quantify degrees of mixing in more precise ways. Quantification also invites visualization, permitting rapid observations into mixing patterns. In this section, we focus on the various ways in which we can visualize code-switching and quantitatively characterize language mixing in a data set given nothing more than a sequence of language tags.

To date, much of linguistic research into code-switching has been concerned with modelling intrasentential mixing, identifying the grammatical sites of switching within a single clause. While such a fine-grained view has its merits in providing insights about discrete aspects of morphosyntax, a broader lens can reveal a more complete picture of code-switching patterns. In examining an entire conversation rather than isolated sentences, it is equally informative, if not more so, to understand *how* switching occurs as it is to know *where* it occurs. Consider the Spanish–English mixed samples in (3) and (4), each containing the same number of Spanish and English words.

3. *Anyway* al taxista *right away* le noté un acentito, *not too specific*
Anyway, on the taxi driver . . . I noted a slight accent . . . (*Killer Crónicas*)
4. Sí, ¿y lo otro no lo es? *Scratch the knob and I'll Kill you*
Yes, and the other one isn't? . . . (*Yo-Yo Boing!*)

These examples can be reduced to a sequence of language tokens that can be grouped into monolingual spans, or sequences of language tokens in exactly one language, as in (5) and

(6), where *E* stands for a word of English and underlined *S*, a word of Spanish. Note that (6) presents two spans, one in Spanish followed by one in English, both seven words in length.

5. ESSEESSSEEEE

6. SSSSSSEEEEEEEE

Depicted in this way, we can make several observations that are independent of the syntactic composition of the utterances. Despite the fact that each contains the same ratio of languages (1:1) and roughly the same number of words, there are more language spans in (5) than in (6), and the spans in (5) are of different lengths whereas those in (6) are of equal length. In other words, the probability of switching in (5) or (6) is decidedly different. Relevant questions then become whether multilingual speakers who switch frequently in a pattern such as (5) do so regularly or only sporadically and to what degree they might do so relative to other possible choices, including producing entirely monolingual utterances. Providing answers to these questions requires methods that consider data beyond the sentence level alone.

Once we no longer limit our empirical observations about mixing patterns to questions of morphosyntax, we become free to explore language contact in novel ways, particularly with regard to variation in what we might call code-switching *signatures*, the extent to which switching patterns vary statistically across users, communities, conversational turns, and the like. For instance, Myslin and Levy (2015) observe that in a small sample of heritage Czech bilinguals in California, speakers frequently switch to English at the end of an Intonational Unit (IU), a location of high information load. The data, bearing only language tags but no lexical tokens, was shared with us by the first author of that study. A snippet of conversation between the fluent speakers ‘H’ and ‘J’ is represented in Table 1.2, with Czech words tagged as Language ‘C’ and English words as ‘E.’ The relatively long spans of Czech, followed by short insertions of English, seem to suggest that the Czech-English mixture is of a different nature than that of the Spanish-English utterances in (5), where the spans of both languages are all relatively short, and in (6), where they are all of the same length. We visually capture the sequential distribution of mixing of the Czech-English conversation excerpted in Table 1.2 by plotting the length of the spans of Czech and of English as they occur within the conversation in a mirrored bar plot (Figure 1.1).

The mirrored plot of the spans in Figure 1.1 provides visual cues to conceptualize the quantification of code-switching as seen over the time course of a conversation. There are roughly 60 monolingual spans represented in this data file; the inequality of the distribution, or the statistical dispersion of Czech relative to English, is represented as deviating in different directions from a horizontal axis, with the English spans falling below the 0 line and the Czech

Table 1.2 Czech-English data

| <i>IU</i> | <i>Speaker code</i> | <i>Language</i> |
|-----------|---------------------|-----------------|
| 33040 | H | C C E |
| 33041 | J | C C C C C |
| 33042 | J | C C C |
| 33043 | J | C C C C |
| 33044 | J | C C C E |

Source: Mark Myslin (personal communication)

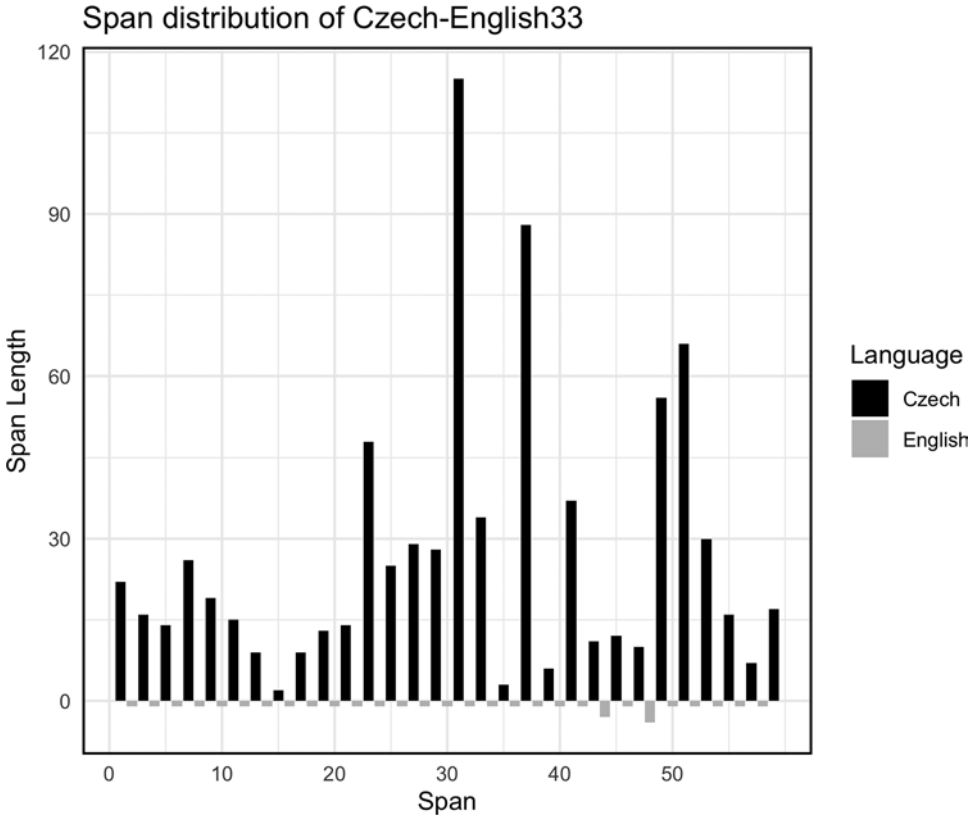


Figure 1.1 Ordered distribution of spans in Czech-English conversation

spans falling above it. This inequality is ordered in a particular way in this corpus: multiple short English spans punctuate the longer Czech ones. One could imagine other configurations composed of the same counts of each language; for instance, there could be a single long English span rather than the 30 short ones shown here. The nature of the alternation between the languages is revealed by the span lengths in Figure 1.1, where it can be noted that the English spans are noticeably brief compared to the Czech language spans. The mirrored plot also shows that the speakers do not switch to English randomly, rather they repeatedly alternate patterns of longer spans of Czech ($\mu \cong 25$ words) followed by short spans of English ($\mu \cong 1$ word). Bullock et al. (2018) model the length of spans in interaction with POS information to demonstrate that switches to English in four different US Spanish-English corpora are significantly shorter after the category of determiners that they are after any other grammatical category where switching occurs. In this way, they are able to provide quantitative support for the notion that switches after determiners are likely to indicate borrowings rather than grammatical alternations.

5. Main research methods

To model variation in language mixing quantitatively, we advocate for a characterization of language mixing along several dimensions that serve to numerically capture what we observe

in Figure 1.1, including measures to gauge (1) the inequality of distribution of languages in a text, (2) the probability of switching between them given this inequality, (3) the randomness of the switching, and (4) the order of the bursts of switching in time. The motivation behind a multi-dimensional characterization of code-switching is that it will allow us to examine the nature of mixing holistically, within an individual unit (utterance, sentence, tweet, IU, etc.) and, more importantly, across the aggregate of units that comprise a corpus. It is only by examining the totality of patterns observed across utterances that we can begin to understand, and be able to compare, trends in language mixing worldwide.

There are several ways to quantify the inequality of distribution in the count of words in each language. Adamou and Granqvist (2015) and Adamou (2016) do so via a word count method, describing the percentage of the data represented by each language (e.g., 65% Romani, 35% Finnish). Another option is to measure language inequality via an *entropy* score: the more evenly distributed the number of tokens are between the languages represented in a corpus, the higher the entropy (Guzmán et al., 2017). The Multilingual index (Barnett et al., 2000) is one measure of the entropy of a distribution; it reflects the ratio of words in each language within a multilingual text on a scale from 0 to 1. An M-index of 0 represents a completely monolingual text, where there is only one language, whereas a text that is completely balanced in terms of languages (e.g., one-third Romani, one-third Greek, one-third Turkish) will result in an M-index of 1.

These measures capture the language makeup of a text but do not reflect in any way the degree to which those languages are interspersed within a text. A parallel corpus, such as Europarl, which consists of European parliamentary transcripts in 21 European languages, is likely to contain a fairly equal percentage of each language, achieving an M-index close to 1, even though it contains little to no mixing. Measures of the balance of languages in a corpus, then, need to be complemented by measures of how the languages are mixed. There are several existing metrics that aim to capture integration (or mixing) complexity. One such metric is the code-mixing index, CMI, introduced by Gambäck and Das (2014). It requires the determination of a matrix language and considers the fraction of total words within an utterance that do not belong to the matrix language. Building upon the CMI, the Complexity Factor (Ghosh, Ghosh and Das, 2018) accounts for three aspects of a text: the fraction of non-matrix language present in a text, the relative order of words, and the number of languages present. Similar in spirit to the CMI is the Integration Index or I-index (Guzmán et al., 2016, 2017), which calculates the probability of switching languages between any two tokens within a corpus. The I-Index does not require the determination of a matrix language for an utterance. It returns a floating point number bounded between 0 and 1; the higher the index, the greater the probability of switching after any token.

The ratio of languages in mixed speech when coupled with a measure of their integration provides insights into whether a language sample tends toward the monolingual or multilingual end of a continuum and locates the samples along this continuum according to the complexity of their mixing. Metrics are also needed to describe how randomly distributed the lengths of language spans between switch points are; for this purpose, measures of *burstiness* (Goh and Barabási, 2008) have been used (Guzmán et al., 2017; Bullock et al., 2018). Alternatives to burstiness include measuring span length entropy (Guzmán et al., 2017) and *surprisal*, which is used to measure the informativeness of an event, such that the less predictable, the more informative, or surprising, the event is assumed to be (Myslín and Levy, 2015). A measure of the temporal distribution of language mixing, or more specifically, the ordering of the language spans could be measured by *memory* (Goh and Barabási, 2008); two corpora may be identical in terms of burstiness values but have very different ordering in their switch points.

Memory quantifies the extent to which the length of a language span is auto-correlated with the length of the span preceding it. In other words, it measures the degree of similarity between the lengths of consecutive pairs of language spans. A measure such as this may eventually be useful for comparing the intermittency of code-switching between corpora should large enough mixed language corpora be available for statistical analysis.

As a caveat, we note that all of the potential metrics presented in this section need to be tested against large quantities of data. Only further data and statistical analysis will allow us to ascertain what are the ‘normal’ ranges of language inequality for multilingual texts, what integration, burstiness, and time-course of switching are predicted, and how much deviation from normal along each of these dimensions might be considered significant. For instance, do certain language pairings show more language integration relative to others? Do some individuals in a given sample show burstier signals in their language mixing than others? What are the limits of language integration?

Applying metrics, we can compare patterns as well as the rate, or frequency, of language mixing across different corpora in a unified way. To illustrate the efficacy of quantitative metrics in showing measurable similarities/differences between corpora in replicable ways, we have compiled a series of data sets in which each token bears a language label. The corpora are listed in (8):

8. Sources and types of corpora
 - A. Type: Pangloss Collection (Michailovsky et al., 2014)
 - a. HSB = Colloquial Upper Sorbian + German (6,100 words)
 - b. RMN = Thrace Romani + Turkish + Greek (2,600 words)
 - c. SVM = Molise Slavic + Italian (27,200 words)
 - d. HRV = Burgenland Croatian + German (2,400 words)
 - e. MKD = Balkan Slavic + Greek (8,000 words)
 - B. Type: Mark Myslín (personal communication)
 - a. Czech 30, Czech 32, Czech 33 = recorded conversations between fluent Czech + English speakers (622 words)
 - C. Type: US Spanish + English fiction
 - a. KC = *Killer Crónicas* by Suzanna Chávez-Silverman (2004) (40,469 words)
 - b. YYB = *Yo-yo Boing!* by Giannina Braschi (1998) (64,848 words)
 - D. Type: U.S. Spanish + English transcripts of interviews
 - a. SpinTX = Spanish in Texas Corpus (Bullock and Toribio, 2013) (416,000 words)
 - b. S7 = interview transcript from Solorio and Liu (2008a) (7,000 words)
 - c. Maria40 = maria40 conversation from the Miami Corpus (Canoflan ESRC Research Centre for Bilingualism, 2015) (7,638 words)
 - E. Type: Quebec French + English film transcript
 - a. BCBC = film transcript of *Bon Cop Bad Cop* (13,502 words)

The documents from the Pangloss Collection (8A) comprise interviews with native speakers of endangered language varieties in Europe. Each token within the data sets analyzed here have been annotated with a language field. The Czech-English interview data (8B) is of the format shown in Table 1.2; it consists only of language annotations. We compiled a Spanish-English data set from multiple available and open collections. The SpinTX data set comprises transcripts from interviews with over 90 Spanish-speaking individuals in Texas; S7 is a transcript of a conversation between three Spanish-English bilinguals, and Maria40 is a file from the Miami Corpus selected because it is known to contain code-switching (Deuchar

et al., 2014). To overcome the issue of scarcity of mixed language corpora of natural speech, we include two works of bilingual fiction recognized as works of ‘Spanglish’ (8C) and a film transcript of the popular Quebec film, *Bon Cop Bad Cop* (8E) that was pieced together using the French and English subtitled transcripts found on the multilingual database, Opensubtitles.org. We have annotated SpinTX, KC, YYB, and BCBC using the automatic methods outlined in Guzmán et al. (2016). S7, the Czech-English data, and Maria40 were hand annotated by their compilers.

Each corpus was processed according to the same conventions, using python scripts (freely available on <https://github.com/Bilingual-Annotation-Task-Force>). First, a count of the number of words in each language returned data on the inequality of distribution of the languages as an M-Index. Transitions between bigrams were marked for whether they encoded a change in language or not. From these transitions, the probability of switching languages was codified as an I-Index. Each corpus was split into a series of monolingual spans from which a data frame of the length and language of each span was created and a measure of corpus burstiness extracted following Guzmán et al. (2017). The results of these analyses, in which the M-Index and the I-Index are log-scaled, can be visualized as in the 3D plot in Figure 1.2.

Although each of the data sets plotted in Figure 1.2 is multilingual, the switching signatures are far from uniform and the data points could be clustered in various ways, depending on the dimension of variation. As a point of reference, KC stands out with the highest values on the planes of multilingualism (M-Index) and probability of switching (I-Index) while also showing the lowest value for burstiness, indicating that switching between Spanish and English occurs regularly throughout the KC. Conversely, switch events in SpinTX occur in a highly sporadic fashion and the data manifests the lowest values for multilingualism and for integration of all the data sets plotted. While not monolingual, it is much more so than other corpora under analysis. The most multilingual corpora – KC, BCBC, and YYB – share high values for the M-Index, but KC falls to a different extreme from BCBC and YYB along the dimensions of integration and burstiness, demonstrating the partial independence of one metric from the next.

The Pangloss data, plotted in Figure 1.2 with the open circle points, appear to split into two clusters, which track those that Adamou (2016) identified on the basis of the word count ratio. But, in Figure 1.2, the clusters are also distinguished by the relative periodicity of the switching in the Molise Slavic (SVM) and Thrace Romani (RMN) data compared to the burstiness of switching in the remaining Pangloss corpora (HRV, MKD, HSB), which group together along every dimension. Thus, it is not only that the number of contact words relative to endangered language is higher but also switching occurs with greater regularity in SVM and RMN than in the other Pangloss data. While differences in burstiness (vertical axis) are not easily apparent from this orientation of the three-dimensional plot, the data show a separation between data sets such that intervals of switching occur more regularly in KC, Czech30, Czech32, SVM, and RMN than in the other data sets. Finally, the Spanish-English conversation data in S7 and Miami40 show a greater balance of languages than all the other naturally elicited conversational data sets but these, too, show that the language switching within the data sets occurs in bursts, entailing more frequent and longer monolingual spans than in the Czech, Molise Slavic, or Thrace Romani data.

These results, preliminary though they are, suggest that there is no single typology of language mixing; instead, we might categorize corpora along different dimensions and in a gradient fashion. Only more data and statistical analyses will tell us whether there are anticipated ‘typical’ values for language switching from which utterances or texts may deviate more significantly than others. *A priori*, we cannot know whether such ‘constants’ of mixing exist at

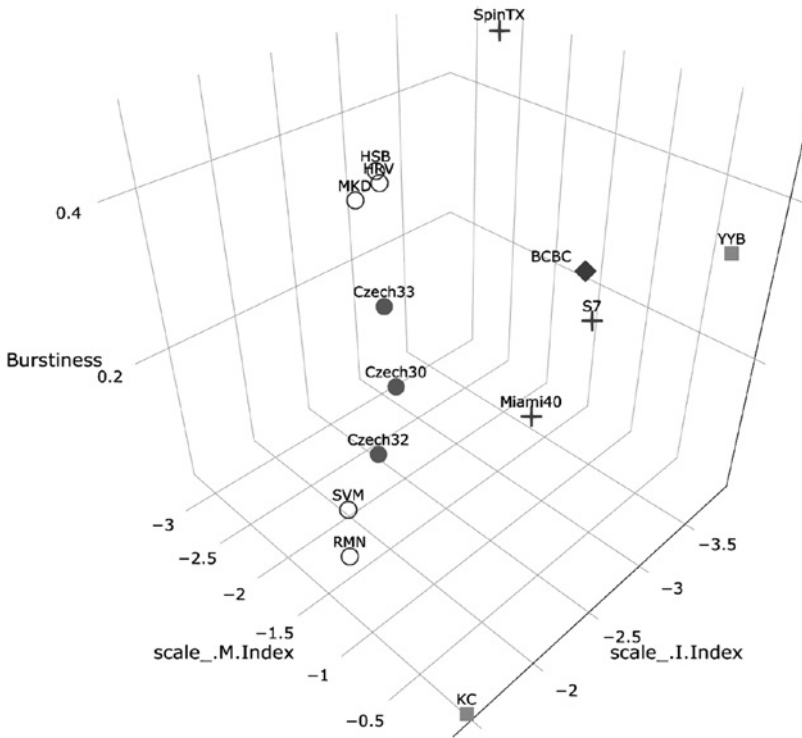


Figure 1.2 Three-dimensional plot of corpora by burstiness * M-Index * I-Index

the universal level or at a level particular to a specific language pairing because we do not have the data to conduct the necessary experiments. Nonetheless, we have a method by which the switching profiles of corpora can be measured and systematically compared, and these produce continuous values that can easily translate to variables for statistical models of code-switching (as for example in Bullock et al., 2018, 2019).

6. Future directions

The methods and approaches to processing multilingual data that we have outlined here are under continuous evaluation, testing, and refinement. However, metrics like these are agnostic with respect to language; they can be applied to any combination of languages. They present a benefit over qualitative approaches in being anonymous and in modelling data at the level of language annotation rather than at the level of the lexical token. This holds the potential to lift the blockade on data sharing that is imposed by concerns for the protection of participants' privacy. In our view, pooling data from various sources, as we have done here, may be the way forward in advancing research on contact languages in the immediate future. Given the ability to quantify language mixing, we can ask new questions whose answers until now could only have been hypothesized. As automated POS tagging improves for code-switching data, we may eventually be able to ascertain whether quantitative differences in switching between

two corpora correlate with qualitative differences (Bullock et al., 2019). In the future, it is likely that we will find it desirable to model utterances at the lexical level to further study the weight of semantics and information structure (Myslín and Levy, 2015) on language mixing. In the meantime, there is much that we can learn about patterns of language data with minimal annotation.

7. Further reading

Adamou, E. (2016). *A corpus-driven approach to language contact: Endangered languages in a comparative perspective*. Boston and Berlin: De Gruyter Mouton.

This comparative analysis of spoken corpora is informative of the manner and extent to which languages may become intertwined. The quantification of mixing in corpora of unrelated contact languages – Balkan Slavic and Thrace Romani in Europe and Ixcatec in Mexico – allows for the data to be situated along a mixing continuum by number of contact tokens.

Bullock, B. and Toribio, A. J., eds. (2009). *The Cambridge handbook of linguistic code-switching*, 1st ed. Cambridge: Cambridge University Press.

This volume surveys the principle theoretical and empirical contributions to the study of code-switching over several decades of scholarship. Eighteen chapters by leading figures address conceptual and methodological considerations in code-switching research, social aspects of code-switching, structural implications of code-switching, psycholinguistics and code-switching, and formal models of code-switching.

Çetinoğlu, O., Schulz, S. and Vu, N. T. (2016). Challenges of computational processing of code-switching. In: *Proceedings of the Second Workshop on Computational Approaches to Code Switching*. Austin, Texas: Association for Computational Linguistics, pp. 1–11.

This publication addresses the challenges that multilingual data present for Natural Language Processing (NLP) tasks such as normalization, language identification, language modelling, part-of-speech tagging, and dependency parsing, some of which are cyclically dependent in pipeline approaches.

Myslín, M. and Levy, R. (2015). Code-switching and predictability of meaning in discourse. *Language*, 91, pp. 871–905.

This study presents an account of code-switching based on a rigorous statistical analysis of a corpus of Czech-English conversational data. Evidence shows that information content, operationalized as meaning predictability, is a determinant in switching: high information content is encoded in one language, and more predictable content is encoded in another.

8. Related topics

Mixed languages, code-switching, typological factors

Abbreviations

| | |
|------|-------------------------|
| BCBC | Bon Cop Bad Cop |
| BNC | British National Corpus |
| C | Czech |
| CMI | code-mixing index |
| csv | comma separated values |
| DET | determiner |
| E | English |
| HRV | Croatian |

| | |
|---------|--|
| HSB | Colloquial Upper Sorbian |
| I-index | Integration index |
| ISO | International Organization for Standardization codes |
| IU | intonational unit |
| KC | Killer Crónicas |
| kpv | Komi-Zyrian |
| LDC | Linguistic Data Consortium |
| M-index | Multilingual index |
| MKD | Macedonian |
| NLP | Natural Language Processing |
| NN | singular or mass nouns |
| NNP | proper singular nouns |
| NNPS | proper plural nouns |
| NNS | plural nouns |
| NUM | number |
| POS | part-of speech |
| PRON | pronoun |
| PUNCT | punctuation |
| RMN | Thrace Romani |
| rus | Russian |
| S | Spanish |
| SpinTX | Spanish in Texas Corpus |
| SVM | Molise Slavic |
| tsv | tab separated values |
| UD | universal dependencies |
| YYB | Yo-yo Boing! |

References

- Adamou, E. (2016). *A corpus-driven approach to language contact: Endangered languages in a comparative perspective*. Boston and Berlin: De Gruyter Mouton.
- Adamou, E. and Granqvist, K. (2015). Unevenly mixed Romani languages. *International Journal of Bilingualism*, 19(5), pp. 525–547.
- Adel, H., Kirchoff, K., Telaar, D., Vu, N.T., Schlippe, T. and Schultz, T. (2014). Features for factored language models for code-switching speech. In: *Proceedings of spoken language technologies for under-resourced languages*. St. Petersburg, Russia: ISCA, pp. 32–38.
- Adel, H., Vu, N.T., Kraus, F., Schlippe, T., Li, H. and Schultz, T. (2013). Recurrent neural network language modeling for code switching conversational speech. In: *The 38th international conference on acoustics, speech and signal processing (ICASSP)*, Vancouver, Canada: IEEE, pp. 8411–8415.
- Auer, P. and Muhamedova, R. (2005). ‘Embedded language’ and ‘matrix language’ in insertional language mixing: Some problematic cases. *Rivista Di Linguistica*, 17(1), pp. 35–54.
- Baldwin, T. and Lui, M. (2010). Language identification: The long and the short of the matter. In: *Human language technologies: The 2010 annual conference of the North American chapter of the association for computational linguistics*. Los Angeles, CA: Association for Computational Linguistics, pp. 229–237.
- Bali, K., Sharma, J., Choudhury, M. and Vyas, Y. (2014). “I am borrowing ya mixing?” An analysis of English-Hindi code mixing in Facebook. In: *Proceedings of the first workshop on computational approaches to code switching*. Doha, Qatar: Association for Computational Linguistics, pp. 116–126.

- Ball, K. and Garrette, D. (2018). Part-of-Speech tagging for code-switched, transliterated texts without explicit language identification. In: *Proceedings of the 2018 conference on empirical methods in natural language processing*. Brussels: Association for Computational Linguistics, pp. 3084–3089.
- Barnett, R., Codó, E., Eppler, E., Forcadell, M., Gardner-Chloros, P., van Hout, R., . . . Sebba, M. (2000). The LIDES coding manual: A document for preparing and analyzing language interaction data version 1.1 – July 1999. *International Journal of Bilingualism*, 4(2), pp. 131–271.
- Belazi, H.M., Rubin, E.J. and Toribio, A.J. (1994). Code switching and X-bar theory: The functional head constraint. *Linguistic Inquiry*, 25(2), pp. 221–237.
- Bhat, G., Choudhury, M. and Bali, K. (2016). Grammatical constraints on intra-sentential code-switching: From theories to working models. *ArXiv Preprint ArXiv:1612.04538*.
- Bon Cop Bad Cop*. (2006). [DVD] Montreal: Alliance Atlantic Vivafilm.
- Braschi, G. (1998). *Yo-Yo Boing!* Pittsburgh, PA: Latin America Literary Review Press.
- Buchholz, S. and Marsi, E. (2006). CoNLL-X shared task on multilingual dependency parsing. In: *Proceedings of the tenth conference on computational natural language learning*. New York: Association for Computational Linguistics, pp. 149–164.
- Bullock, B.E., Guzmán, G.A., Serigos, J. and Toribio, A.J. (2018). Should code-switching models be asymmetric? In: *Interspeech 2018*. Hyderabad, India: International Speech Communication Association, pp. 2534–2538.
- Bullock, B.E., Guzmán, W. and Toribio, A.J. (2019). The limits of Spanglish? In: B. Alex, S. Degaetano-Ortlieb, A. Kazantseva, N. Reiter, and S. Szpakowicz, eds., *Proceedings of the 3rd Joint SIGHUM workshop on computational linguistics for cultural heritage, social sciences, humanities and literature*. Minneapolis: Association for Computational Linguistics, pp. 115–121.
- Bullock, B.E. and Toribio, A.J. (2009). *The Cambridge handbook of linguistic code-switching*. Cambridge and New York: Cambridge University Press.
- Bullock, B.E. and Toribio, A.J. (2013). The Spanish in Texas corpus project. [online] *Center for open education resources and language learning (COERLL)*. Available at: www.spanishintexas.org
- Canolfan ESRC Centre for Research on Bilingualism. Corpus: Miami. Available at: <http://bangortalk.org.uk/speakers.php?c=miami> [Accessed 23 Jan. 2015].
- Çetinoglu, O., Schulz, S. and Vu, N.T. (2016). Challenges of computational processing of code-switching. In: *EMNLP 2016, Proceedings of the second workshop on computational approaches to code-switching*. Austin, TX Association for Computational Linguistics, pp. 1–11.
- Chávez-Silverman, S. (2004). *Killer Crónicas: Bilingual memories*. Madison, WI: University of Wisconsin Press.
- Cieri, C., Maxwell, M., Strassel, S. and Tracey, J. (2016). Selection criteria for low resource language programs. In: *Proceedings of the tenth international conference on language resources and evaluation (LREC, 2016)*. Portoroz, Slovenia: European Language Resources Association.
- Demo, Š. (2018). Mining macaronics. In: P. Pahta, J. Skaffari, and L. Wright, eds., *Multilingual practices in language history: English and beyond*, 1st ed. Berlin: De Gruyter Mouton, pp. 199–221.
- Deuchar, M., Davies, P., Herring, J., Couto, M.C.P. and Carter, D. (2014). Building bilingual corpora. In: E.M. Thomas and I. Mennen, eds., *Advances in the study of bilingualism*, 1st ed. Bristol: Multilingual Matters, pp. 93–111.
- Diab, M. and Kamboj, A. (2011). Feasibility of leveraging crowd sourcing for the creation of a large scale annotated resource for Hindi English code switched data: A pilot annotation. In: *Proceedings of the 9th workshop on Asian language resources*. Chiang Mai, Thailand: Asian Federation of Natural Language Processing.
- Donnelly, K. and Deuchar, M. (2011). The Bangor autoglosser: A multilingual tagger for conversational text. Presented at: *The fourth international conference on internet technologies and applications (ITA II)*. Wrexham, Wales.
- Dorleijn, M. and Nortier, J. (2009). Code-switching and the internet. In: B.E. Bullock and A.J. Toribio eds., *The Cambridge handbook of linguistic code-switching*, 1st ed. Cambridge: Cambridge University Press, pp. 127–141.

- Francom, J., Hulden, M. and Ussishkin, A. (2014). ACTIV-ES: A comparable, cross-dialect corpus of ‘everyday’ Spanish from Argentina, Mexico, and Spain. In: *Proceedings of the ninth annual conference on language resources and evaluation (LREC, 2014)*. Reykjavik, Iceland: European Language Resources Association, pp. 1733–1737.
- Gambäck, B. and Das, A. (2014). On measuring the complexity of code-mixing. In: *Proceedings of the 11th international conference on natural language processing*. Goa, India: NLP Association of India, pp. 1–7.
- Garrette, D. and Baldrige, J. (2013). Learning a part-of-speech tagger from two hours of annotation. *Proceedings of the 2013 conference of the North American chapter of the association for computational linguistics: Human language technologies*. Atlanta, GA: Association for Computational Linguistics, pp. 138–147.
- Ghosh, S., Ghosh, S. and Das, D. (2018). Complexity metric for code-mixed social media text. *ArXiv Preprint ArXiv:1707.01183*.
- Goh, K.-I. and Barabási, A.-L. (2008). Burstiness and memory in complex systems. *EPL (Europhysics Letters)*, 81(4), p. 48002.
- Granger, S., Dagneaux, E., Meunier, F. and Paquot, M. (2009). *International corpus of learner English*. Louvain-la-Neuve: Presses Universitaires de Louvain.
- Guzmán, G., Ricard, J., Serigos, J., Bullock, B.E. and Toribio, A.J. (2017). Metrics for modeling code-switching across corpora. In: *Interspeech 2017*. Stockholm, Sweden: International Speech Communication Association, pp. 67–71.
- Guzmán, G.A., Serigos, J., Bullock, B.E. and Toribio, A.J. (2016). Simple tools for exploring variation in code-switching for linguists. In: *Proceedings of the second workshop on computational approaches to code-switching*. Austin, TX: Association for Computational Linguistics, pp. 2–20.
- Herring, J.R., Deuchar, M., Couto, M.C.P. and Quintanilla, M.M. (2010). ‘I saw the Madre’: Evaluating predictions about codeswitched determiner-noun sequences using Spanish – English and Welsh – English data. *International Journal of Bilingual Education and Bilingualism*, 13(5), pp. 553–573.
- Jamatia, A., Gambäck, B. and Das, A. (2015). Part-of-speech tagging for code-mixed English-Hindi Twitter and Facebook chat messages. In: *Proceedings of the international conference recent advances in natural language processing*. Hissar, Bulgaria: INCOMA Ltd., p. 239.
- Johannessen, J.B. (2015). The corpus of American Norwegian speech (CANS). In: *Proceedings of the 20th Nordic conference of computational linguistics (NODALIDA 2015)*. Vilnius, Lithuania: Linköping University Electronic Press, pp. 297–300.
- Jurgens, D., Dimitrov, S. and Ruths, D. (2014). Twitter users #codeswitch hashtags! #moltoimportante #wow. In: *EMNLP 2014, Proceedings of the first workshop on computational approaches to code switching*. Doha, Qatar: Association for Computational Linguistics, pp. 51–61.
- Leech, G., Garside, R. and Bryant, M. (1994). CLAWS4: The tagging of the British National corpus. In: *Proceedings of the 15th conference on computational linguistics*. Stroudsburg, PA: Association for Computational Linguistics, pp. 622–626.
- Lignos, C. and Marcus, M. (2013). Toward web-scale analysis of codeswitching. In: *Proceedings of annual meeting of the linguistic society of America*. Boston, MA.
- Lison, P. and Tiedemann, J. (2016). Opensubtitles 2016: Extracting large parallel corpora from movie and tv subtitles. In: *Proceedings of the 10th international conference on language resources and evaluation (LREC2016)*. Portorož, Slovenia: European Language Resources Association, pp. 923–929.
- Liu, Y. (2008). Evaluation of the matrix language hypothesis: Evidence from Chinese-English code-switching phenomena in blogs. *Journal of Chinese Language and Computing*, 18(2), pp. 75–92.
- Lui, M. and Baldwin, T. (2012). langid.py: An off-the-shelf language identification tool. In: *Proceedings of the ACL 2012 system demonstrations*. Jeju Island, Korea: Association for Computational Linguistics, pp. 25–30.
- Lyu, D.-C., Tan, T.-P., Chng, E.-S. and Li, H. (2010). An analysis of a Mandarin-English code-switching speech corpus: SEAME. In: *Interspeech 2010*. Makuhari, Japan: ISCA, pp. 25–28.
- Lyu, D.-C., Tan, T.-P., Chng, E.-S. and Li, H. (2015). Mandarin – English code-switching speech corpus in South-East Asia: SEAME. *Language Resources and Evaluation*, 49(3), pp. 581–600.

- MacWhinney, B. (2007). The talkbank project. In: J.C. Beal, K.P. Corrigan, and H.L. Moisl, eds., *Creating and digitizing language corpora*, 1st ed. London: Palgrave Macmillan, pp. 163–180.
- Marcus, M., Santorini, B. and Marcinkiewicz, M.A. (1993). *Building a large annotated corpus of English: The Penn Treebank*. University of Pennsylvania.
- Marwick, A.E. and Boyd, D. (2011). I tweet honestly, I tweet passionately: Twitter users, context collapse, and the imagined audience. *New Media & Society*, 13(1), pp. 114–133.
- McDonald, R., Nivre, J., Quirnbach-Brundage, Y., Goldberg, Y., Das, D., Ganchev, K., . . . Täckström, O. (2013). Universal dependency annotation for multilingual parsing. In: *Proceedings of the 51st annual meeting of the association for computational linguistics (Volume 2: Short papers)*. Sofia, Bulgaria: Association for Computational Linguistics, pp. 92–97.
- Michailovsky, B., Mazaudon, M., Michaud, A., Guillaume, S., François, A. and Adamou, E. (2014). Documenting and researching endangered languages: The Pangloss collection. *Language Documentation and Conservation*, 8, pp. 119–135.
- Muysken, P. (2000). *Bilingual speech: A typology of code-mixing*. Cambridge: Cambridge University Press.
- Myers-Scotton, C. (1993). *Duelling languages: Grammatical structure in codeswitching*. Oxford: Oxford University Press (Clarendon Press).
- Myslin, M. and Levy, R. (2015). Code-switching and predictability of meaning in discourse. *Language*, 91(4), pp. 871–905.
- Nagy, N. (2011). A multilingual corpus to explore variation in language contact situations. *Rassegna Italiana Di Linguistica Applicata*, 43(1–2), pp. 65–84.
- Nivre, J., De Marneffe, M.-C., Ginter, F., Goldberg, Y., Hajic, J., Manning, C.D., . . . Silveira, N. (2016). Universal dependencies v1: A multilingual treebank collection. In: *Proceedings of the tenth international conference on language resources and evaluation (LREC, 2016)*. Portoroz, Slovenia: European Language Resources Association.
- Parafita Couto, M.C. and Gullberg, M. (2017). Code-switching within the noun phrase: Evidence from three corpora. *International Journal of Bilingualism*, 3(2), pp. 695–714.
- Partanen, N., Lim, K., Rießler, M. and Poibeau, T. (2018). Dependency parsing of code-switching data with cross-lingual feature representations. In: *Proceedings of the fourth international workshop on computational linguistics of Uralic languages*. Helsinki, Finland: Association for Computational Linguistics, pp. 1–17.
- Poplack, S. (1980). Sometimes I'll start a sentence in Spanish y termino en español: Toward a typology of code-switching. *Linguistics*, 18(7–8), pp. 581–618.
- Schulz, S. and Keller, M. (2016). Code-switching ubique est-language identification and part-of-speech tagging for historical mixed text. In: *Proceedings of the 10th SIGHUM workshop on language technology for cultural heritage, social sciences, and humanities*. Berlin, Germany: Association for Computational Linguistics, pp. 43–51.
- Sitaram, S. and Black, A.W. (2016). Speech synthesis of code mixed text. In: *Proceedings of the tenth international conference on language resources and evaluation (LREC, 2016)*. Portoroz, Slovenia: European Language Resources Association, pp. 3422–3428.
- Solorio, T., Blair, E., Maharjan, S., Bethard, S., Diab, M., Gohneim, M., . . . others. (2014). Overview for the first shared task on language identification in code-switched data. In: *Proceedings of the first workshop on computational approaches to code switching*. Doha, Qatar: Association for Computational Linguistics, pp. 62–72.
- Solorio, T. and Liu, Y. (2008a). Learning to predict code-switching points. In: *Proceedings of the conference on empirical methods in natural language processing*. Seattle, WA: Association for Computational Linguistics, pp. 973–981.
- Solorio, T. and Liu, Y. (2008b). Part-of-speech tagging for English-Spanish code-switched text. In: *Proceedings of the conference on empirical methods in natural language processing*. Seattle, WA: Association for Computational Linguistics, pp. 1051–1060.

- Streiter, O., Scannell, K.P. and Stuffesser, M. (2006). Implementing NLP projects for noncentral languages: Instructions for funding bodies, strategies for developers. *Machine Translation*, 20(4), pp. 267–289.
- Thomason, S.G. and Kaufman, T. (1988). *Language contact, creolization, and genetic linguistics*. Berkeley: University of California Press.
- Trosterud, T. (2006). Grammatically based language technology for minority languages. In: A. Saxena and L. Borin, eds., *Lesser-known languages of South Asia: Status and policies, case studies and applications of information technology*. Berlin and New York: De Gruyter Mouton, pp. 293–316.
- Vilares, D., Alonso, M. and Gómez-Rodríguez, C. (2016). EN-ES-EC: An English-Spanish code-switching Twitter corpus for multilingual sentiment analysis. In: *Proceedings of the tenth international conference on language resources and evaluation (LREC 2016)*. Portoroz, Slovenia: European Language Resources Association, pp. 4149–4153.
- Winford, D. (2003). *An introduction to contact linguistics*. Oxford: Blackwell.
- Zhang, Y., Riesa, J., Gillick, D., Bakalov, A., Baldridge, J. and Weiss, D. (2018). A fast, compact, accurate model for language identification of codemixed text. *ArXiv:1810.04142 [Cs]*.

Language contact in the lab

*Paola E. Dussias, Judith F. Kroll,
Melinda Fricke, and Michael A. Johns*

1. Introduction

A great deal of research has been conducted to examine language contact phenomena that takes the bilingual speech community as its angle of vision. This approach to language contact is anchored in sociolinguistics (e.g., Labov, 2012) and is concerned with understanding the linguistic outcomes of language contact and the role of internal linguistic factors in this process. But language contact is also an individual enterprise; the two languages of a bilingual speaker interact in the bilingual mind, and these interactions have consequences for language processing, for cognition, and for the brain over the life span. Historically, the popular view has been that becoming bilingual during early infancy produces confusion and language delay, that acquiring a second language as an adult is a feat accomplished only by those with a gift for language learning, and that individuals who freely code-switch between their two languages do not speak either language proficiently or are mentally lazy (for a discussion, see Kroll and Dussias, 2017). However, recent evidence from lab-based research has shown, quite compellingly, that these assumptions and attitudes are, in fact, myths (Kroll et al., 2015). Bilingualism is not a special case or circumstance, and individuals who learn a second language well past early childhood are not doomed to fail. Quite to the contrary, recent lab-based behavioural and neuroscience evidence demonstrates that bilinguals are able to acquire subtle aspects of a second language and can sometimes process linguistic input in the second language in a manner that is indistinguishable from monolingual speakers. Infants are not confused by hearing two or more languages; instead, they develop an intricate ability to distinguish the languages they hear. And regularly switching from one language to another, far from being a sign of random interference, is a display of exquisite knowledge and control of two languages (e.g., Green and Wei, 2014). Bilingualism provides language scientists with the potential to understand how humans negotiate the boundaries of two languages within a single mind. For cognitive psychologists, the bilingual and multilingual speaker has become a model for understanding the way that language experience shapes the mind and the brain (e.g., Kroll et al., 2012).

In the past several decades, lab-based research with bilingual and multilingual speakers has led to a number of important discoveries about what it means to be bilingual. Being bilingual is not only about acquiring and using a second language; it is also about the ways in which the

native language changes in response to a new second language. There is fundamental permeability across the two languages that makes early and late bilinguals more, rather than less, similar to one another. This permeability also makes cross-language interactions bidirectional, changing the way that bilinguals process each of their two languages. One exciting discovery from lab-based research on bilingualism is that the use of two or more languages across the lifespan has consequences that extend beyond language processing to domain-general cognitive functions. These consequences are particularly beneficial to individuals later in life, when typical cognitive decline places greater demands on the cognitive system.

In this chapter, our goal is to discuss the various experimental methodologies that have been employed in lab-based research to study the consequence of having two languages in a single mind. The focus will be mostly on bilinguals who have acquired a second language during adulthood – once a native language has been firmly established. We will begin by reviewing central findings in bilingual language processing. We will then describe some of the most common behavioural methods that have been used to uncover the different levels of language at which cross-language interactions have been observed in bilingual speakers, as well as how research on second language learning and bilingualism has benefited from the introduction of tools for examining neurocognitive processes. We conclude with some suggestions that we see as critical in advancing the study of bilingualism in lab-based contexts.

2. Critical issues and topics: language contact in the mind and brain of bilingual speakers

Models of language processing are in general agreement with respect to the basic levels of linguistic representation that are required for language production and comprehension. While some debate persists in terms of the precise organization and content of linguistic representations and on the way in which information flows through the network, all major models agree that speakers must have representations that capture semantic, syntactic, distributional, and phonological/word form information (cf. Dell, 1986 for (monolingual) language production; McClelland and Elman, 1986 for (monolingual) spoken word recognition). Similarly, while they may disagree on the finer points, psycholinguists who study bilingualism generally agree that the languages of a bilingual speaker interact to some degree during language processing. This portion of the chapter first briefly describes some evidence for the foundational idea that cross-language activation is pervasive in bilingual language processing. It then presents some more recent findings that add nuance to the overall picture, addressing questions such as *how variable are cross-language activation patterns in different speakers? how proficient in a second language does one need to be for the first language to be affected? and how can we relate these laboratory studies to bilinguals' everyday experiences?*

The languages of bilingual and multilingual speakers are intertwined

For essentially all aspects of language processing that have been studied, there is evidence that the languages of a bilingual speaker interact. This is true for the semantic (e.g., Morford et al., 2011), syntactic (e.g., Loebell and Bock, 2003), and phonological (e.g., Goldrick, Runnqvist and Costa, 2014) levels of representation; it is true for language production (Kroll and Gollan, 2014) as well as comprehension (Kroll and Dussias, 2012); and interestingly, not only does the first or more dominant language (L1) affect the second or less dominant language (L2), recent work demonstrates that the L2 in many cases can affect the L1. For all of these reasons, Grosjean (1989) famously argued that ‘The Bilingual Is not Two Monolinguals in One Person.’ In

other words, because the languages of a multilingual speaker appear to be fundamentally and reciprocally intertwined, it would be misguided to think of either L1 or L2 processing as taking place independently of the other language.

The central role of cross-language activation in bilingual language processing has been widely agreed on for some time. A more recent development in bilingualism research, and in psycholinguistics more generally, has been a shift in the focus from so-called group level analyses to what are often referred to as ‘individual differences’ in language processing (for a review see Fricke et al., 2019). Broadly speaking, new experimental and statistical methods increasingly allow psycholinguists to model variation in language processing behaviour in more sophisticated, and more principled, ways. Driven partially by these new techniques, and partially by the concomitant acknowledgment that no matter what their language background, no population of speakers is homogenous in their language processing behaviour, in recent years it has become less common to conduct experiments comparing a single group of bilingual speakers to a supposedly ‘matched’ group of monolinguals, and more common to compare different types of bilingual speakers whose language experience or cognitive profiles vary along some dimension(s) of theoretical interest (e.g., Luk and Bialystok, 2013). To take just two recent examples, Beatty-Martínez and Dussias (2017) found that processing of a particular structure unique to code-switched speech differed when comparing highly proficient bilinguals with and without code-switching experience, and Zirnstein, van Hell and Kroll (2018) found that performance on a non-linguistic cognitive inhibition task predicted the neural correlates of language processing behaviour in both monolinguals and bilinguals. Such studies signal a profound shift in the way psycholinguists think about the status of the bilingual speaker, i.e., not as a special ‘type’ of language user, but rather as an encapsulation of many issues crucial to the field’s understanding of how language processing works, such as how language experience shapes speakers’ and listeners’ mental representations, and how language processing interacts with cognition more generally.

Another shift in psycholinguists’ thinking about bilingualism can be seen in recent work concerning second language learning. While it has long been acknowledged that the age at which speakers begin acquiring a second language tends to be negatively correlated with their ultimate attainment in a variety of linguistic domains (Birdsong, 2018), many researchers have begun to focus less on when such ‘age of acquisition’ effects emerge, and more on when they are *not* in evidence. Recent research on L2 sentence processing in particular has demonstrated that under certain circumstances, the parsing strategies of proficient L2 learners are equivalent to those of native speakers (e.g., Steinhauer, 2014). Thus, while L1-to-L2 transfer effects are undoubtedly an important part of L2 processing, an elaborated account of L2 learning must also account for the fact that native-like processing of the target language no longer seems as elusive as it once did. To reconcile these facts, many researchers have begun to see L2 learning as fundamentally dynamic, and to focus their efforts on probing the limits of this dynamic process.

In line with the view that L2 learning is fundamentally dynamic, recent studies have also affirmed the fact that, counter to what one might assume, *the L1 is similarly dynamic*, and this appears to be the case for all levels of linguistic representation. Bice and Kroll (2015), for example, presented evidence that access to L1 lexical representations is affected by the existence of the L2 even at early stages of L2 acquisition; work by Chang (2012) has similarly demonstrated that L1 phonetic categories are susceptible to L2 influence as early as the onset of L2 acquisition; and work by Dussias and Sagarra (2007) has shown that prolonged immersion in a second language environment can ultimately result in transfer of syntactic parsing preferences from the L2 to the native language. It is not yet clear how the influence of the L2 on the L1 may differ depending on the type of linguistic structure or process being

examined, whether or how proficiency plays a role in mediating this influence, or what factors may encourage versus discourage such transfer effects. For example, the discussion in Bice and Kroll (2015) focuses on learners' ability to inhibit the more dominant language, while Chang (2013) proposes that it is the novelty of recently experienced L2 exemplars that may play a crucial role. These proposals are not necessarily mutually exclusive, and much work remains to be done in order to understand the relative contribution of the relevant factors. Taken together with recent findings demonstrating the possibility of native-like processing of the L2 even for late learners, though, the emerging picture is one of a highly integrated and plastic linguistic-cognitive system that in many cases remains capable of accommodating and supporting new language learning.

Why psycholinguists are excited about studying code-switching

A recurring theme in the preceding sections is the idea that psycholinguistics is in a period of reorientation toward the ways in which language processing is variable, dynamic, and continually susceptible to the influence of new linguistic experience. Along with this reorientation, psycholinguists have recently begun to turn their attention to bilingual code-switching. For the purposes of this discussion, we define code-switching as any multilingual speaker's use of two or more languages within a single conversation. In some sense, of course, psycholinguists could be seen as a bit 'late to the party' on this topic, since linguists have been studying code-switching for years (e.g., see chapters in this volume and references therein). But given the current state of the field, there is good reason for psycholinguists to turn their attention to code-switching now. For one thing, one consequence of the parallel activation of the bilinguals' two languages is the ability of highly proficient bilinguals to code-switch. The ubiquity with which certain bilingual communities engage in code-switching challenges the strong unilingual perspective prevalent in psycholinguistics research. Code-switching requires the seamless and successful integration of two grammars at multiple linguistic levels, i.e., phonology, morphology, syntax, and discourse; consequently, how bilinguals systematically engage and disengage their languages in real time becomes a new and important avenue of inquiry for understanding bilingual sentence processing and language control. There is yet another reason for studying code-switching in the lab. Code-switching research in linguistics has focused to a lesser extent on the comprehension of code-switched sentences. This gap is significant given that code-switching is ubiquitous in some bilingual communities. Research has instead focused almost exclusively on production in both theoretical and sociolinguistic frameworks. For bilingual listeners or readers, however, switches can be unexpected and thus potentially more difficult to process than within-language sentences. This observation has given rise to two main threads of inquiry in the study of code-switching. One is whether integrating code-switched speech is costlier than unilingual sentence processing (see for example, Adamou and Shen, 2019; Gullifer, Kroll and Dussias, 2013); the other one is how bilinguals adapt their parsing strategies to better anticipate upcoming code-switches. The first approach typically pairs code-switched stimuli with unilingual stimuli and tests whether integrating code-switched text or speech is costly relative to non-switched or unilingual speech/text. The second approach, which is emerging in the neuro- and psycho-linguistic study of code-switching, is how bilinguals adapt to rapidly integrating code-switched speech. Both approaches are novel in that they assume that regardless of whether code-switching results in switch costs, bilinguals engage in it. Therefore, bilingual code-switchers must adapt their parsing strategies in order to accommodate to this linguistic behaviour. This observation has created a perfect storm for advancing the study of code-switching.

Much as we argued that the bilingual speaker is an encapsulation of many issues critical to the understanding of language processing more generally, we would also argue that code-switching is not *only* an interesting phenomenon in and of itself: code-switched language provides a perfect venue for asking questions concerning the ways in which speakers' day-to-day experience with their language impacts their language processing behaviour.

The immediate consequences of language contact in the bilingual mind

In the preceding section, we considered how traditional claims about dual language processing have been modified by discoveries about the plasticity of the bilingual's two languages and resulting variation. The dynamics of cross-language exchange have profound consequences for the way in which each language is processed as learners first acquire an L2 and once individuals are proficient bilinguals. But the consequences extend beyond language processing because language processing engages cognitive and neural resources that reshape not only the language system but also the domain general cognitive networks that support it.

A key discovery in the studies of bilingual brain structure and function is that the two languages are largely supported by the same neural tissue (e.g., Perani and Abutalebi, 2005; but see Xu et al., 2017). As we noted earlier, the evidence on language processing in bilinguals suggests that there are dynamic interactions across a bilingual's two languages at every level of processing, from lexicon to grammar to phonology. The questions that have framed the research at each of these levels, and that are alluded to in the preceding section, have focused on how each language reflects the contact and interaction with the other language. Critically, bilinguals also have to develop the ability to negotiate the interactions and resulting competition that result from these cross-language exchanges. In other reviews, we have characterized bilinguals as model mental 'jugglers' who can negotiate these interactions with ease, enabling proficient performance in each language alone but also switching languages when the circumstances make it required or appropriate (e.g., Kroll et al., 2012; Kroll and Navarro-Torres, 2018). What are the cognitive and neural resources that support bilingual performance? A series of recent studies has demonstrated that domain general resources are recruited to enable monolinguals and bilinguals alike to resolve competition that arises from ambiguities at both the syntactic and lexical levels (e.g., Hsu and Novick, 2016). These studies typically involve dual task designs in which a conflict resolution task, like the Stroop task, is combined with a language processing task, like resolving a syntactically ambiguous sentence. The dual task paradigm, in its various forms, makes it possible to observe the recruitment of cognitive resources on the fly, as language processing is in progress. What we see is that demands on language processing that introduce conflict also increase demands on cognitive resources. For bilinguals, the interactions between the two languages create a continual situation of having to navigate the joint activation of the two languages and decisions about when each language should be spoken and with whom. Models of bilingual control assume that these factors determine how cognitive resources are recruited to effectively change bilingual minds and brains relative to those of their monolingual counterparts (e.g., Green and Wei, 2014). However, the resulting dynamics are complex, because bilinguals use the two languages in different contexts and for different communicative purposes.

The long-term consequences of language contact in the bilingual mind

Many studies have now asked whether and how the engagement of cognitive resources when two languages are in frequent contact might create special circumstances for bilinguals across

their lives. Although there has been controversy about the consequences of bilingualism for cognition and for the neural mechanisms that support it (e.g., García-Pentón et al., 2016), there is overwhelming evidence that suggests that the use of two or more languages tunes those aspects of cognitive function that support proficient language use. Notably, much of the research that has been critical of the claim that bilingualism confers cognitive advantages has focused on a narrow range of executive function tasks that ask the simple question of whether bilinguals, regardless of who they are and how they use the two languages, outperform monolinguals in behavioural measures of executive function. Finding that we sometimes see differences between bilinguals and monolinguals and we sometimes do not seem surprised given the research on variation in language processing that we have mentioned earlier. Some bilinguals code-switch frequently with others and other bilinguals use their two languages in different contexts. Some bilinguals live immersed in the second language environment, whereas others acquire and use a second language in the context of their native language. Some bilinguals acquired the two languages from birth and others begin to learn a second language as adults. The more recent studies suggest that it is not a matter of a simple bilingual advantage or not because individuals come to be bilingual in many different ways and use the two languages in ways that differ in the demands that they make on cognitive resources (e.g., Pot, Keijzer and de Bot, 2018).

What we see is that within the first year of life, exposure to more than one language begins to change the processing of speech in ways that open the language system more broadly (e.g., Werker, 2012) and that begin to affect the development of brain functions for cognitive control (e.g., Ferjan Ramírez et al., 2017). Although little is known about the trajectory of these changes from infancy through adulthood, by the time of old age, bilingualism appears to confer a set of remarkable protections against cognitive decline in healthy older adults (e.g., Bialystok, Craik and Freedman, 2007; Alladi et al., 2013).

If the evidence is compelling that language contact in bilinguals has extended consequences for the mind and brain, then why has there been controversy in the literature? We and others have argued (e.g., Fricke et al., 2019) that a remarkable feature of the body of work on the consequences of bilingualism is that it largely ignores the nature of *language contact* itself. Most studies that assess the consequences of bilingualism have compared bilinguals and monolinguals without consideration of language use beyond self-assessed measures of language proficiency. The result is an extensive body of correlational research that fails to identify those aspects of language use that may determine the consequences of bilingualism. To have a causal account of how language contact may create the circumstances that give rise to changes in the mind and brains of speakers, we need to have a full account that specifies how the mechanisms of language processing relate to cognition under different circumstances of language use.

3. Research methods to study language contact in lab

Lab-based studies of bilingualism have employed a range of behavioural and neuroscience methods to examine the way in which bilingual readers, listeners, and speakers negotiate the presence of two languages in a single mind. As mentioned earlier, because the language systems of bilinguals are not independent, there is a critical question about how bilingualism affects basic aspects of language processing. In this section, we will discuss a number of research methods that have been employed to reveal the type of cross-language interactions that are a hallmark of bilingual language processing. Before we move to describe these methods, it is important to note that our discussion will necessarily be restricted to research on bilingual language comprehension. However, many recent findings concerning the way in

which bilingual speakers produce words and sentences in each of their languages are largely consistent with results demonstrating the parallel activation of the two languages of bilingual speakers. Our discussion will focus on the most frequently employed behavioural paradigms and neuroscience methods: reaction time methods, self-paced reading, eye-tracking during reading, eye-tracking during spoken language comprehension, and the recording of brain activity. Our aim is to highlight the type of methodological cross-pollination in bilingual lab-based research that has led to discoveries about what it means to be bilingual. Our approach will be to provide a brief overview of each method and to discuss some of the recent findings in the bilingualism literature that highlight how the method has helped researchers inform key issues in bilingual language processing. We will focus on three such issues, which were discussed earlier: cross-language activation and language non-selectivity, the malleability of the native language system, and the role of variation in the processing of code-switched language. We note that each of the methods described here have been used to investigate major areas of bilingual research activity in experimental psycholinguistics: how bilinguals recognize words when they are spoken or read in each language; how the sounds of each of the bilingual's two languages are processed when they are heard or spoken; how the grammatical structures and preferences associated with each of the bilingual's languages are affected by the presence of both languages. Within each of these topics our review will necessarily be brief, but we hope to illustrate the logic of the experimental approach in a way that will provide a useful guide to the primary literature.

Response-time techniques

Response-time techniques measure the amount of time that elapses from the moment participants perceive a stimulus to when they give a response in some way or another (typically through a button or keyboard press). In general, tasks that require a simple response (such as *lexical decision*, *masked priming*, *picture-word Stroop*, *semantic relatedness judgments*), a choice (e.g., press the blue button if you hear a unilingual sentence and press the red button if you hear a code-switched sentence) or a selection (e.g., press a button if you hear a unilingual sentence, otherwise, don't do anything) are all instances of response-time techniques. The dependent measure in a response-time experiment is reaction time (RT), which is known to be modulated by at least three factors: perception of the stimulus, processing of the stimulus, and the response (itself influenced by the motor agility to respond to the stimulus).

Response-time paradigms such as the ones just mentioned have been used extensively to examine visual word recognition, spoken word recognition, and spoken word production in bilingual and multilingual speakers (e.g., Schwartz, Kroll and Diaz, 2007). And tasks as simple as lexical decision have provided the context in which a set of factors can be manipulated to determine whether only one or both languages are active when a string of letters is presented. In lexical decision, a string of letters (e.g., piano) is presented on a computer screen, and participants are asked to indicate whether the string forms a real word or not via a button press. Speed and accuracy are recorded. When the string is a real word, the speed and accuracy with which participants respond is influenced by the lexical characteristics of the word (e.g., part of speech, word length, lexical frequency, number of syllables, number of phonemes, number of orthographic neighbours, number of phonological neighbours, imageability of the word, etc.). When the letter string does not form a real word, it is typically a *legal nonword* in that it is pronounceable and follows the spelling rules of the language (e.g., fendle). The inclusion of legal nonwords is important to ensure that the mental lexicon itself is accessed to determine whether the word is known.

Interestingly, one of the most well-established findings in the bilingual literature – that each of the bilingual’s two languages activates the other even when only one of the languages is in use – comes from studies that have used response-time techniques. The logic in these studies has been to exploit similarities that exist across languages in orthography or phonology. For example, Spanish and English have many words that are identical or very similar in their spelling patterns and carry the same meaning. Such *cognate* words provide an ingenious way to test whether bilinguals are able to function monolingually in performing a task such as lexical decision. In Spanish and English, the words ‘piano’ and ‘banana’ are cognates because they have the same spelling and meaning in both languages. Other cognates, such as ‘blusa’ in Spanish and ‘blouse’ in English are similar, but not identical, in the two languages. If bilinguals function as monolinguals in the sense that they access a word in one language without activating the other language, then lexical decision performance for cognates should be no different than lexical decision for words that do not share characteristics between the two languages. In other words, a Spanish-English bilingual performing a lexical decision task on English words should not be influenced by the fact that the same cognates also exist in Spanish. Moreover, bilinguals should not perform any differently on cognate words when compared to monolinguals. However, many experiments (e.g., van Hell and Dijkstra, 2002) show that bilinguals are in fact faster to decide that a string of letters is a cognate – a finding typically referred to as the *cognate facilitation effect*.

Although many of the bilingual results obtained using response-time techniques are still broadly accepted, some of the findings have also been questioned on the basis of distortions that are introduced by the conscious process of decision-making. There is, however, recent evidence from lab-based studies using methods such as the recording of brain responses to investigate the neural basis of bilingual word recognition, that provide support for the observation that there is parallel activation of the bilingual’s two languages (e.g., Thierry and Wu, 2007). There is also compelling evidence from word recognition tasks such as word naming, lexical decision, and reading while eye movements are tracked, demonstrating that even when language ambiguous words, such as cognates, are not presented in isolation but are instead embedded in sentences, the information that the sentence is only in one language is not sufficient to create language-specific conditions that functionally allow bilinguals to read or listen to speech as if they were monolingual. The point of these examples is that response-time techniques, despite being less intricate than other more sophisticated methods such as brain imaging and electrophysiological techniques, have produced a number of key findings about what it means to be bilingual which have stood the test of time.

Self-paced reading

The self-paced reading task is, in essence, a type of response-time task. In a typical self-paced reading experiment, a stimulus sentence is presented on a computer screen, segmented into words or phrases that are displayed one at a time. Participants initiate the experiment by pressing a *trigger* (e.g., key on a button box or on a computer keyboard). This action brings up the first segment (or display). Participants read the segment, press the trigger to request the next display, and continue performing the same action until they reach the end of the sentence. The measure of interest is the time that participants spend reading a critical word or phrase compared to a control condition. In one variation of the self-paced reading task known as the *reading moving-window* (Just, Carpenter and Woolley, 1982), the display on the computer screen moves from left to right in tandem with each trigger press to allow the words of the sentence to occupy the same position on the screen that would surface if the sentence were displayed

as a whole. Each letter, except from the letters of the word in current view, is replaced with a dash (or similar ‘place holder’). In the reading moving window paradigm, the text can be presented *cumulatively* (previously read words remain on the screen as new ones are added) or *non-cumulatively* (only a single word or segment is displayed at a time, with all preceding and successive words hidden from the participant’s view). Because the cumulative version has the disadvantage that participants may press the trigger to display all the words in a sentence, and only later initiate the actual reading task, researchers typically favour non-cumulative displays over cumulative ones.

Although one advantage of the moving window task is that it allows for the collection of word-level reading times, thereby allowing researchers to identify the specific loci of processing difficulty, a criticism levelled against the self-paced reading task is the likelihood that participants’ reading strategies are influenced by the type of segmentation employed by the experimenter (Gilboy and Sopena, 1996). A second objection is that it relies on a secondary task (a button or a key press) to produce the dependent measure. These, and other factors, have led researchers to favour methods that provide a richer body of data than the single latency that results from self-paced reading. In the next section, we discuss a few of the measures that have allowed researchers to determine with more precision the existence, locus, and time course of processing difficulty in bilingual speakers.

Eye-tracking during reading

Among the techniques that psycholinguists employ to study the interaction of two languages in the bilingual mind, the recording of eye movements is a popular method. This is in part because several decades of eye-movement research during reading have generated very detailed information about how visual information is processed while our eyes move across a line of text (Rayner, 1983). For example, when text becomes more complex or contains uncommon words, eye fixation duration increases and saccade length (i.e., small jumps made by the eye to move through text) decreases. Unexpected words also have immediate effects on fixation duration. We know that readers tend to look longer at unpredictable words than predictable words, and they skip over predictable words more frequently than unpredictable words. Similarly, high frequency words decrease fixation duration (e.g., Rayner and Pollatsek, 1987) compared to lower frequency words, even when both types of words are matched for length, number of syllables, meaning, and sentence frame. What is critical for researchers is that the sensitivity of eye movements to the characteristics of the text can be captured in the *gaze duration* of readers (i.e., the initial amount of time a reader spends in a region from first entering it until the eyes move to another word). Eye movements are likewise influenced by variations in the content of the text. To illustrate further, in sentences that are syntactically ambiguous, when disambiguating information is inconsistent with the syntactic interpretation assigned by a reader, there is considerable disruption in eye movement (e.g., more eye regressions to the difficult material in the text). The fact that inconsistencies associated with the structural analysis of a particular word or words are noticed by readers as soon as they arise provides support for the *immediacy assumption* – the assumption that readers do not wait to interpret text until a number of words have been encountered, but rather interpret each word of a text as soon as it is encountered. Another major advantage of the eye-movement recording technique is that it allows researchers to obtain evidence about sentence comprehension moment by moment, as processing unfolds, without significantly altering the normal characteristics of either the task or the presentation of the stimuli. In addition, to obtain the dependent

measure, participants are not required to perform a secondary task (such as a button or keyboard press, name a word or a picture) that might disrupt the normal comprehension process.

Many recent studies using eye-tracking methods during reading have contributed to a nuanced understanding of how the languages of bilingual and multilingual speakers are processed in an integrated language system in which there is extensive interaction. In Dussias and Sagarra (2007), for example, monolingual Spanish speakers and Spanish-English bilinguals with limited and extended immersion experience in their L2 English read syntactically ambiguous Spanish sentences containing a relative clause that was preceded by a noun phrase (NP) with two potential attachment sites (e.g., *Arrestaron a la hermana del hombre que estaba enferma*/Someone arrested the sister of the man who was ill_{FEM}). For these structures, past research has shown that Spanish speakers attach the ambiguous relative clause (e.g., *que estaba enferma*/who was ill_{FEM}) to the first noun (*hermana* ‘sister’; NP1). Conversely, English speakers resolve the ambiguity in favour of the second noun (man; NP2). Dussias and Sagarra found that Spanish monolinguals and bilinguals with little immersion in the L2 environment attached the relative clause to the first noun, a finding that replicated prior research (Carreiras and Clifton, 1999). Critically, the bilinguals who had been living in an environment in which English was predominant, initially attached the relative clause to the second noun, and revised their interpretation once they realized that it was incorrect. In other words, for these speakers, exposure to a preponderance of English constructions that favours NP2 resolution rendered this interpretation more available, resulting in an NP2 preference when reading in their first language.

The observation that exposure to L2 parsing strategies may come to affect L1 parsing highlights the dynamic nature of the linguistic system and provides support for experience-based models of sentence processing (e.g., MacDonald, 2013 and references therein), given the assumption within these models that frequency-based exposure is crucial to parsing. If the parser’s configuration is related to intense language experience, bilinguals’ parsing preferences are expected to change as a function of the frequency with which the relevant structure appears in the environment. Subsequent studies have demonstrated that this type of shifting is in fact a natural consequence of the inherent flexibility of the language, but importantly for the purposes of this illustration, the Spanish-English speakers showed a robust preference which emerged in eye-tracking records of the participants.

Eye-tracking during spoken language comprehension

Although reading processes have provided important insights into the mechanisms involved during bilingual language processing, one may wonder whether the processing characteristics uncovered to date are specific to reading or can be generalized to spoken-language comprehension. One experimental methodology that has had great success in research on auditory language processing is the *visual world paradigm* (e.g., Allopenna, Magnuson and Tanenhaus, 1998). The visual world combines experimental designs typically employed in eye tracking with spoken language comprehension studies, and has successfully been used to answer research questions related to virtually any area of spoken language comprehension (for a review, see Huettig, Olivers and Hartsuiker, 2011). Researchers have also been able to extend the use of the visual world paradigm into the realm of language production and message generation to answer questions about temporal links between eye movements and speech planning. The paradigm has also been used in research with children (e.g., Trueswell et al., 1999) as well as in studies involving bilingual speakers (e.g., Hopp, 2012).

In a typical visual world experiment, auditory material is concurrently presented with a related visual scene containing pictured objects that are displayed on a computer screen. The auditory material plays spoken instructions related to the objects (e.g., *click on the candy*) which participants are asked to follow. During the experiment, participants' eye movements to the objects are monitored as the object name mentioned in the instruction unfolds over time. Researchers using the visual world paradigm look for the presence of *competitor* and *anticipatory* effects relative to a neutral baseline. Competitor effects are taken to reflect delayed processing, whereas anticipatory effects are interpreted as indexing facilitated processing.

The critical manipulation in bilingual studies examining competitor effects is the presence of objects whose spoken name in one language is phonologically similar to the name of an unrelated object in the other language. The goal of these studies has been to ask whether bilingual speakers activate words from their two languages in parallel when they hear words in one language alone. To illustrate, in Spivey and Marian (1999), L1 Russian speakers proficient in English heard the instruction 'Put the marker below the cross' in the presence of a visual display that contained four objects: a marker, a stamp (whose translation in Russian *marka* shares initial phonetic features with the English *marker*), and two other objects whose English and Russian names had no phonetic similarity to the target word. Findings showed that when Russian-English speakers heard the word *marker* in English, they were likely to also look at the Russian between-language competitor *marka*. This result, replicated with L2 speakers of other language backgrounds, suggests that bilingual listeners do not appear to be able to deactivate the irrelevant mental lexicon.

Facilitatory effects in visual world studies with bilingual speakers have been used to examine the type of L1 'reconfiguration' that has served as evidence in support of the plasticity of the native language. For example, Valdés Kroff et al. (2017) considered the social context in which code-mixing occurs to ask whether intense contact with Spanish-English code-switched speech had consequences for the processing of grammatical gender in the participants' L1 Spanish. The processing of grammatical gender was chosen because researchers have consistently documented that in some Spanish-English bilingual communities, the Spanish masculine article *el* surface with English nouns whose Spanish translations are masculine or feminine (e.g., *el juice*/Spanish *jugo*_{MASC}; *el cookie*/Spanish *galleta*_{FEM}). In contrast, mixed NPs with the Spanish feminine article *la* are less frequent, and only surface with English nouns whose Spanish translations are feminine (e.g., *la cookie*_{FEM} but never *la juice*_{MASC}). These production asymmetries, to which many Spanish-English bilinguals are exposed, stand in marked contrast to monolingual Spanish, where grammatical gender is obligatorily encoded and not interchangeable, and the distribution between masculine and feminine nouns in Spanish is roughly half. Given the asymmetry observed in production data, it seemed plausible that the gender-marking of articles would facilitate to a lesser extent the processing of code-switched speech.

To investigate this, the eye movements of Spanish-English bilinguals were recorded while they heard Spanish-only sentences. Participants saw pairs of objects displayed on a computer screen in non-informative contexts (i.e., when the gender of the two pictures was the same, and thus participants must wait to hear the onset of the target word to follow the spoken instructions), and in informative contexts (i.e., when the gender of the two pictures was different, and so the target noun could be anticipated by paying attention to the grammatical gender encoded in the determiner). Pictures were presented while participants listened to Spanish sentences [e.g., *Encuentra el libro*_{MASC} 'Find the book' in the presence of a book and a bed *cama*_{FEM}]. Results showed that in informative contexts, the gender information present in the determiner was used anticipatorily only when the article was feminine. When it was masculine, participants did not launch anticipatory looks but rather waited to hear the noun. The fact that

masculine trials showed no anticipatory effects suggests that immersion in a code-switching environment impacted the strategies that bilinguals use to employ gender anticipatorily in Spanish, their native language.

With these examples, our goal has been to illustrate that the use of eye-tracking during reading and spoken language processing has raised the prospect of addressing increasingly subtle questions in bilingual lab research: as the nature of the research questions have become more refined, the need for more sophisticated on-line behavioural measures, such as the recording of eye-movements, has become central.

Electroencephalography (EEG)

Behavioural techniques, including eye-movement records and reaction time methods, despite providing important insights about the temporal flow of information, can leave questions unanswered about the relationship between the dependent measures and the cognitive processes underlying syntactic and semantic processing. That is, slower reading in reaction-time studies or longer fixation durations in eye-tracking studies do not unambiguously provide information about the time course of particular component processes (Carreiras and Clifton, 2004). In addition, longer eye fixations and increased regressive eye movements have been shown to occur in the vicinity of syntactic or semantic anomalies, making it challenging to differentiate whether a particular response is due to the detection of a syntactic or a semantic irregularity.

One method that has been used to examine whether bilingual language processing is fundamentally similar to or different from native language processing is electroencephalography (EEG). Time-averaged event-related potentials (ERPs) derived from EEG are useful because the brain's electrophysiological responses have been found to be sensitive to language-related events and also to the linguistic levels (i.e., semantic vs. syntactic) of anomaly of these events. For example, past studies have shown that neurophysiological responses related to the processing of semantic information, including plausibility information, elicit a negative-going waveform peaking at around 400 ms after the onset of the target word – the so-called *N400*. The *N400* is part of the typical electrical brain activity that arises when semantic expectancy is violated, such as when readers encounter a stimulus that is semantically anomalous or that is semantically congruent but has a low cloze probability. The *N400* has been reported when readers encounter words that do not fit the semantic context of a sentence (for a review, see Kutas and Federmeier, 2011), as would be expected, for example, when an optionally transitive verb is followed by an implausible direct object (e.g., The scuba diver discovered the headache was caused by changes in. . .). Grammatical and syntactic violations, on the other hand, sometimes give rise to a positive deflection that peaks at around 600 ms post-stimulus onset, and that is evident over the midline and posterior regions of the scalp (*P600*; Osterhout and Holcomb, 1992).

An important caveat of the ERP research with bilingual (and monolingual) speakers is that the interpretation of the ERP effects rests on the assumption of homogeneity across individuals. However, recent evidence has shown that ERP responses are sensitive to individual differences both in the L1 (Pakulak and Neville, 2010) and in the L2 (Tanner et al., 2013). Given this, a central aspect of the study of bilingual language processing in the lab should be the role of the speakers' linguistic experience and of variation in the input that they are exposed to (see discussion in Boland et al., 2016). To illustrate this in the context of code-switching, recent evidence suggests that bilinguals who code-switch are more sensitive to code-switching structures that are consistent with attested distributional patterns and therefore demonstrate facilitated processing as compared to unattested code-switches (e.g., Adamou

and Shen, 2019). Conversely, bilinguals who do not frequently code-switch should not show differential processing to attested vs. unattested code-switches, as these code-switches are virtually all unattested to bilingual non-code-switchers. Indeed, recent studies have shown that distributional regularities involving attested code-switching patterns act as cues heightening the probability of upcoming switches (e.g., Fricke, Kroll and Dussias, 2016). Moreover, these observations also predict variability among speakers who engage and do not engage in code-switching. To illustrate this approach, Beatty-Martínez and Dussias (2017) conducted a study using two groups of bilinguals who differed in code-switching experience. The goal of the study was to examine the consequences of adaptation to language processing across different communities of speakers. The first experiment analyzed ERPs to compare the processing of code-switches that were either rarely attested or commonly attested in bilingual corpora from a habitual code-switching community. For code-switchers, rarely attested code-switches evoked an N400 effect in comparison to common code-switches, suggesting greater difficulty with lexical integration. Non-code-switchers, on the other hand, processed these two types of code-switches similarly. These findings underscore how the processing of code-switched language largely depends on bilinguals' language experience, in other words, on the type of code-switching strategies available in the speakers' discourse environment.

4. Future directions

We have shown that lab-based studies of bilingual language processing have revealed much about how the two languages of a bilingual interact with one another in the bilingual mind. There is ample evidence of shared syntactic representations in bilinguals and multilinguals alike (Hartsuiker, Pickering and Veltkamp, 2004; Kantola and van Gompel, 2011), an ever-growing literature on the interaction between bilingualism and cognition and the effects the former may have on the latter (e.g., Kroll and Bialystok, 2013), and a large number of studies investigating language contact at its perhaps most intimate: code-switching (Adamou and Shen, 2019; Beatty-Martínez and Dussias, 2017; Guzzardo Tamargo, Valdés Kroff and Dussias, 2016; Johns, Valdés Kroff and Dussias, 2019). Using a variety of methodologies, from lexical decision, to EEG, to eye-tracking, psycholinguists have gained a better understanding of the neural and cognitive mechanisms underlying bilingual language use.

A feature that has been notably lacking from lab-based studies is the consideration of the social context in which bilinguals use their two languages. While some influential cognitive models have attempted to capture these different contexts (e.g., Green and Wei, 2014), the incorporation of sociolinguistic insights into the lab-based study of bilingualism is recent. Green and Wei (2014), in their Adaptive Control Model, have taken into consideration the social contexts in which languages are used, to propose that the environment in which bilinguals finds themselves plays a large role in shaping the cognitive control mechanisms that oversee language selection. While bilinguals who tend to speak only one language at home and another language outside the home may engage in a more 'competitive' control mode (where only one language is selected at a time), bilinguals who engage in regular code-switching may be more adept at using an 'open' control mode (with both languages active).

Aside from the studies described here, most work on bilingual language processing and language production in the lab has largely eschewed the social contexts in which bilinguals actually use their two languages. Given that language is an inherently social act, and bilingual language use is certainly no exception, the lab-based study of bilingualism stands to benefit greatly from the incorporation of sociolinguistic insights into both the interpretation and design of experiments. This may be accomplished, we propose, by turning to

perhaps one of the most widely used approaches to language contact *outside* of the lab: the variationist approach.

Used for decades to study both monolingual speech patterns as well as determine the locus of language contact phenomena, the variationist approach uses conditioned rates of usage of particular variants (be them phonetic, morphological, or syntactic) to determine which variants are preferred (or dispreferred) in particular contexts (be them phonetic, morphological, syntactic, *or social*). Importantly, this approach relies on *spontaneous* speech rather than elicited speech to more accurately reflect the ways in which individuals use their language(s). Variationist approaches to language contact have been used to study, among many other topics, subject expression in contact varieties of Spanish (e.g., Otheguy and Zentella, 2011), the use of lone English-origin words in otherwise Finnish speech (Poplack, Wheeler and Westwood, 1989), and the utterance-final intonation contours of Catalan-Spanish bilinguals (Simonet, 2011). Of particular importance to variationist approaches to language is the role of linguistic experience in language use: namely, the ways in which language is represented and used is based in large part on the experiences of the individual and the practice of the speech community in which individuals find themselves. The findings just summarized show that sociolinguistic data can explain the results of lab-based studies of bilingual language processing (e.g., Johns, Valdés Kroff and Dussias, 2019) and highlight the importance of understanding production patterns in bilingual speech; specifically, when experimental design deviates significantly from what bilinguals encounter in spontaneous productions, the generalizability of results is affected.

It is important to note that the study of variation in language processing is not necessarily a new phenomenon; instead, its application to bilingualism and language contact in the lab has only recently begun to emerge. Monolingual studies of the effects of linguistic experience on processing have been conducted since the early and mid-2000s. For example, previous research has found that object relatives in English (e.g., *The boy that the girl hit. . .*) are more difficult to process in general than subject relatives (e.g., *The girl that hit the boy. . .*). Wells et al. (2009), however, found that even short-term changes in exposure to object relatives can facilitate their processing. Similarly, Fraundorf and Jaeger (2016) found that readers unfamiliar with the *needs* + past participle structure (e.g., *The car needs washed*) initially exhibit difficulty processing the unfamiliar structure, but over the course of a single experiment adapt to the structure and show similar processing strategies as individuals from communities where the *needs* + past participle structure is attested.

The results of both monolingual and multilingual studies of language processing discussed here, as well as those of many other studies, show that the incorporation of sociolinguistic data, reflective of the natural variation found in language, can be used to not only interpret the findings of lab-based studies of language processing but also to design experiments and experimental stimuli. We suggest that future work must consider the importance of evaluating real language usage and experimental evidence within a single, well-defined bilingual community. Ultimately, as new questions emerge and methodologies are developed, our unifying aim should be to stimulate an ecologically valid, holistic approach to linguistic study.

5. Further reading

Kroll, J. F. and De Groot, A. M. B. (2005). *Handbook of bilingualism: Psycholinguistic approaches*. New York: Oxford University Press.

This edited volume discusses developments within psycholinguistic research on bilingualism, including cognitive neuroscience approaches.

Dussias, P. E., Valdés Kroff, J. R., Beatty-Martínez, A. L. and Johns, M. (in press). What language experience tells us about cognition: variable input, and interactional contexts affect bilingual sentence processing. In: J. W. Schwieter, ed., *The Handbook of the neuroscience of multilingualism*, 1st ed. Hoboken, NJ: John Wiley & Sons Ltd.

In this article, the authors argue that linguistic experience – the input that bilinguals are exposed to and the interactional contexts in which bilinguals find themselves – serve an influential role in bilingual language processing and thus must be integrated in lab-based studies of bilingualism.

Wei, L. and Moyer, E. (2009). *The Blackwell guide to research methods in bilingualism and multilingualism*. Oxford: Blackwell.

This edited volume discusses the strong interdisciplinary nature that characterizes research on bilingualism, and offers a practical guide to the procedures and tools for collecting and analyzing data. The volume presents the most common procedures, methods and tools to study bilingualism and multilingualism inside and outside the lab, discussing core concepts and approaches.

6. Related topics

A variationist perspective on language contact, the 4-M model: different routes in production for different morphemes, usage-based approaches, code-switching

Abbreviations

| | |
|------|--------------------------|
| EEG | Electroencephalography |
| ERP | event-related potentials |
| FEM | feminine |
| L1 | first language |
| L2 | second language |
| MASC | masculine |
| NP | noun phrase |
| RT | reaction time |

References

- Adamou, E. and Shen, X.R. (2019). There are no language switching costs when codeswitching is frequent. *International Journal of Bilingualism*, 23, pp. 53–70.
- Alladi, S., Bak, T.H., Duggirala, V., Surampudi, B., Shailaja, M., Shukla, A.K., Chaudhuri, J.R. and Kaul, S. (2013). Bilingualism delays age at onset of dementia, independent of education and immigration status. *Neurology*, 81, pp. 1938–1944.
- Allopenna, P.D., Magnuson, J.S. and Tanenhaus, M.K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, 38, pp. 419–439.
- Beatty-Martínez, A.L. and Dussias, P.E. (2017). Bilingual experience shapes language processing: Evidence from codeswitching. *Journal of Memory and Language*, 95, pp. 173–189.
- Bialystok, E., Craik, F. I. M. and Freedman, M. (2007). Bilingualism as a protection against the onset of symptoms of dementia. *Neuropsychologia*, 45, pp. 459–464.
- Bice, K. and Kroll, J.F. (2015). Native language change during early stages of second language learning. *NeuroReport*, 26, pp. 966–971.
- Birdsong, D. (2018). Plasticity, variability and age in second language acquisition and bilingualism. *Frontiers in Psychology*, 9, p. 81.
- Boland, J.E., Kaan, E., Valdés Kroff, J.R. and Wulff, S. (2016). Psycholinguistics and variation in language processing. *Linguistic Vanguard*, 2, pp. 3–12.

- Carreiras, M. and Clifton, C. Jr. (1999). Another word on parsing relative clauses: Eyetracking evidence from Spanish and English. *Memory & Cognition*, 27, pp. 826–833.
- Carreiras, M. and Clifton, C. Jr. (2004). *The on-line study of sentence comprehension: Eyetracking, ERPs and beyond*. New York: Psychology Press.
- Chang, C.B. (2012). Rapid and multifaceted effects of second-language learning on first-language speech production. *Journal of Phonetics*, 40, pp. 249–268.
- Chang, C.B. (2013). A novelty effect in phonetic drift of the native language. *Journal of Phonetics*, 41, pp. 520–533.
- Dell, G.S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, 93, pp. 283–331.
- Dussias, P.E. and Sagarra, N. (2007). The effect of exposure on syntactic parsing in Spanish – English bilinguals. *Bilingualism: Language and Cognition*, 10, pp. 101–116.
- Ferjan Ramírez, N., Ramírez, R.R., Clarke, M., Taulu, S. and Kuhl, P.K. (2017). Speech discrimination in 11 month old bilingual and monolingual infants: A magnetoencephalography study. *Developmental Science*, 20, e12427.
- Fraundorf, S.H. and Jaeger, T.F. (2016). Readers generalize adaptation to newly-encountered dialectal structures to other unfamiliar structures. *Journal of Memory and Language*, 91, pp. 28–58.
- Fricke, M., Kroll, J.F. and Dussias, P.E. (2016). Phonetic variation in bilingual speech: A lens for studying the production – comprehension link. *Journal of Memory and Language*, 89, pp. 110–137.
- Fricke, M., Zirnstein, M., Navarro-Torres, C. and Kroll, J.F. (2019). Bilingualism reveals fundamental variation in language processing. *Bilingualism: Language and Cognition*, 22, pp. 200–207.
- García-Pentón, L., Fernández García, Y., Costello, B., Duñabeitia, J.A. and Carreiras, M. (2016). The neuroanatomy of bilingualism: How to turn a hazy view into the full picture. *Language, Cognition and Neuroscience*, 31, pp. 303–327.
- Gilboy, E. and Sopena, J.M. (1996). Segmentation effects in the processing of complex NPs with relative clauses. In: M. Carreiras, J.E. García-Albea, and N. Sebastián-Gallés, eds., *Language processing in Spanish*. Mahwah, NJ: Lawrence Erlbaum Associates, pp. 191–206.
- Goldrick, M., Runqvist, E. and Costa, A. (2014). Language switching makes pronunciation less native-like. *Psychological Science*, 25, pp. 1031–1036.
- Green, D.W. and Wei, L. (2014). A control process model of codeswitching. *Language, Cognition and Neuroscience*, 29, pp. 499–511.
- Grosjean, F. (1989). Neurolinguists, beware! The bilingual is not two monolinguals in one person. *Brain and Language*, 36, pp. 3–15.
- Gullifer, J.W., Kroll, J.F. and Dussias, P.E. (2013). When language switching has no apparent cost: Lexical access in sentence context. *Frontiers in Psychology*, 4, 278.
- Guzzardo Tamargo, R.E., Valdés Kroff, J.R. and Dussias, P.E. (2016). Examining the relationship between comprehension and production processes in codeswitched language. *Journal of Memory and Language*, 89, pp. 138–161.
- Hartsuiker, R.J., Pickering, M.J. and Veltkamp, E. (2004). Is syntax separate or shared between languages? Cross-linguistic syntactic priming in Spanish-English bilinguals. *Psychological Science*, 15, pp. 409–414.
- Hopp, H. (2012). The on-line integration of inflection in L2 processing: Predictive processing of German Gender. In: A.K. Biller, E.Y. Chung and A.E. Kimball, eds., *BUCLD 36: Proceedings of the 36th annual Boston University conference on language development*. Boston: Cascadilla Press, pp. 226–245.
- Hsu, N.S. and Novick, J.M. (2016). Dynamic engagement of cognitive control modulates recovery from misinterpretation during real-time language processing. *Psychological Science*, 27, pp. 572–582.
- Huetig, F., Olivers, C.N.L. and Hartsuiker, R.J. (2011). Looking, language, and memory: Bridging research from the visual world and visual search paradigms. *Acta Psychologica*, 137, pp. 138–150.
- Johns, M.A., Valdés Kroff, J.R. and Dussias, P.E. (2019). Mixing things up: How blocking and mixing affect the processing of codemixed sentences. *International Journal of Bilingualism*, 23, pp. 584–611.

- Just, M.A., Carpenter, P.A. and Woolley, J.D. (1982). Paradigms and processes in reading comprehension. *Journal of Experimental Psychology: General*, 111, pp. 228–238.
- Kantola, L. and van Gompel, R.P. (2011). Between-and within-language priming is the same: Evidence for shared bilingual syntactic representations. *Memory & Cognition*, 39, pp. 276–290.
- Kroll, J.F. and Bialystok, E. (2013). Understanding the consequences of bilingualism for language processing and cognition. *Journal of Cognitive Psychology*, 25, pp. 497–514.
- Kroll, J.F. and Dussias, P.E. (2012). The comprehension of words and sentences in two languages. In: B.T. Ritchie, ed., *The handbook of bilingualism and multilingualism*, 2nd ed. Malden, MA: Wiley-Blackwell, pp. 216–243.
- Kroll, J.F. and Dussias, P.E. (2017). The benefits of multilingualism to the personal and professional development of residents of the US. *Foreign Language Annals*, 50, pp. 248–259.
- Kroll, J.F., Dussias, P.E., Bogulski, C.A. and Valdés Kroff, J. (2012). Juggling two languages in one mind: What bilinguals tell us about language processing and its consequences for cognition. In: B. Ross, ed., *The psychology of learning and motivation*, vol. 56, 1st ed. Amsterdam: Elsevier, pp. 229–262.
- Kroll, J.F., Dussias, P.E., Bice, K. and Perrotti, L. (2015). Bilingualism, mind, and brain. *Annual Review of Linguistics*, 1, pp. 377–394.
- Kroll, J.F. and Gollan, T.H. (2014). Speech planning in two languages: What bilinguals tell us about language production. In: M. Goldrick, V. Ferreira, and M. Miozzo, eds., *The Oxford handbook of language production*, 1st ed. Oxford: Oxford University Press, pp. 165–181.
- Kroll, J.F. and Navarro-Torres, C. (2018). Bilingualism. In: J. Wixted, ed., *The Stevens' handbook of experimental psychology and cognitive neuroscience*, Section IV. (S. Thompson, Schill, ed., *Language and thought*). Hoboken, NJ: Wiley-Blackwell, pp. 245–274.
- Kutas, M. and Federmeier, K.D. (2011). Thirty years and counting: Finding meaning in the N400 component of the event related brain potential (ERP). *Annual Review of Psychology*, 62, pp. 621–647.
- Labov, W. (2012). What is to be learned: The community as the focus of social cognition. *Review of Cognitive Linguistics*, 10, pp. 265–293.
- Loebell, H. and Bock, K. (2003). Structural priming across languages. *Linguistics*, 41, pp. 791–824.
- Luk, G. and Bialystok, E. (2013). Bilingualism is not a categorical variable: Interaction between language proficiency and usage. *Journal of Cognitive Psychology*, 25, pp. 605–621.
- MacDonald, M.C. (2013). How language production shapes language form and comprehension. *Frontiers in Psychology*, 4, 226.
- McClelland, J.L. and Elman, J.L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, 18, pp. 1–86.
- Morford, J.P., Wilkinson, E., Villwock, A., Piñar, P. and Kroll, J.F. (2011). When deaf signers read English: Do written words activate their sign translations? *Cognition*, 118, pp. 286–292.
- Osterhout, L. and Holcomb, P.J. (1992). Event-related brain potentials elicited by syntactic anomaly. *Journal of Memory and Language*, 31, pp. 785–806.
- Otheguy, R. and Zentella, A.C. (2011). *Spanish in New York: Language contact, dialectal leveling, and structural continuity*. Oxford: Oxford University Press.
- Pakulak, E. and Neville, H.J. (2010). Proficiency differences in syntactic processing of monolingual native speakers indexed by event-related potentials. *Journal of Cognitive Neuroscience*, 22(12), pp. 2728–2744.
- Perani, D. and Abutalebi, J. (2005). The neural basis of first and second language processing. *Current Opinion in Neurobiology*, 15, pp. 202–206.
- Poplack, S., Wheeler, S. and Westwood, A. (1989). Distinguishing language contact phenomena: Evidence from Finnish-English bilingualism. *World Englishes*, 8, pp. 389–406.
- Pot, A., Keijzer, M. and de Bot, K. (2018). Intensity of multilingual language use predicts cognitive performance in some multilingual older adults. *Brain Sciences*, 8, 92.
- Rayner, K. (1983). The perceptual span and eye movement control during reading. In: R. Rayner, ed., *Eye movements in reading: Perceptual and language processes*, 1st ed. New York: Academic Press, pp. 97–139.

- Rayner, K. and Pollatsek, A. (1987). Eye movements in reading: A tutorial review. In: M. Coltheart, ed., *Attention and performance XII: The psychology of reading*. London: Lawrence Erlbaum Associates, pp. 327–362.
- Schwartz, A.I., Kroll, J.F. and Diaz, M. (2007). Reading words in Spanish and English: Mapping orthography to phonology in two languages. *Language and Cognitive Processes*, 22(1), pp. 106–129.
- Simonet, M. (2011). Intonational convergence in language contact: Utterance-final F0 contours in Catalan – Spanish early bilinguals. *Journal of the International Phonetic Association*, 41(2), pp. 157–184.
- Spivey, M. and Marian, V. (1999). Crosstalk between native and second languages: Partial activation of an irrelevant lexicon. *Psychological Science*, 10(3), pp. 281–284.
- Steinhauer, K. (2014). Event-related potentials (ERPs) in second language research: A brief introduction to the technique, a selected review, and an invitation to reconsider critical periods in L2. *Applied Linguistics*, 35(4), pp. 393–417.
- Tanner, D., McLaughlin, J., Herschensohn, J. and Osterhout, L. (2013). Individual differences reveal stages of L2 grammatical acquisition: ERP evidence. *Bilingualism: Language and Cognition*, 16, pp. 367–382. <https://doi.org/10.1017/S1366728912000302>
- Thierry, G. and Wu, Y.J. (2007). Brain potentials reveal unconscious translation during foreign-language comprehension. *Proceedings of the National Academy of Sciences*, 104(30), pp. 12530–12535.
- Trueswell, J., Sekerina, I., Hill, N. and Logrip, M. (1999). The kindergarten-path effect: Studying on-line sentence processing in young children. *Cognition*, 73, pp. 89–134.
- Valdés Kroff, J.R., Dussias, P.E., Gerfen, C., Perrotti, L. and Bajo, M.T. (2017). Experience with codeswitching modulates the use of grammatical gender during sentence processing. *Linguistic Approaches to Bilingualism*, 7, pp. 163–198.
- van Hell, J.G. and Dijkstra, T. (2002). Foreign language knowledge can influence native language performance in exclusively native contexts. *Psychonomic Bulletin & Review*, 9(4), pp. 780–789.
- Wells, J.B., Christiansen, M.H., Race, D.S., Acheson, D.J. and MacDonald, M.C. (2009). Experience and sentence processing: Statistical learning and relative clause comprehension. *Cognitive Psychology*, 58, pp. 250–271.
- Werker, J. (2012). Perceptual foundations of bilingual acquisition in infancy. *Annals of the New York Academy of Sciences*, 1251, pp. 50–61.
- Xu, M., Baldauf, D., Chang, C.Q., Desimone, R. and Tan, L.H. (2017). Distinct distributed patterns of neural activity are associated with two languages in the bilingual brain. *Science Advances*, 3, e1603309.
- Zirnsstein, M., van Hell, J.G. and Kroll, J.F. (2018). Cognitive control ability mediates prediction costs in monolinguals and bilinguals. *Cognition*, 176, pp. 87–106.

Acknowledgements

This work was supported in part by NSF grants BCS-1535124, OISE-0968369, OISE-1545900; NIH grant HD082796 to PD and JK, HD071758 to PD.

A variationist perspective on language contact*

Shana Poplack

1. Introduction and definitions

Variability is a hallmark of speech, and bilingual speech is no exception. The variationist perspective capitalizes on this fact, arguing that the outcomes of language contact cannot be fully understood without considering the fine details of inherent variability – not only in the presumed recipient of any contact effects, but in each of the languages involved. This chapter reviews the methodological and analytical tenets associated with variationist sociolinguistics, focusing on the type of data that is targeted for analysis and the quantitative reasoning and standards of proof that underlie its claims. Key among methodological imperatives are the validation of speakers in the context of their community and the primacy of actual bilingual performance data, as instantiated by the compilation of bilingual corpora constructed on scientific principles. Fundamental analytical requirements include systematic analysis of the usage patterns that emerge from such data, contextualization of these patterns across speakers, mixing strategies and contact varieties, and application of quantitative reasoning to their interpretation. Results of independent studies that adopt the comprehensive empirical method outlined here confirm its ability to address and resolve controversies around language contact that have plagued scholars for decades.

1.1 *The variationist paradigm*

As it relates to language contact, the variationist paradigm is perhaps best characterized by the twin notions of variability and accountability. Its core construct, the linguistic variable, captures the phenomenon whereby a speaker employs different linguistic means to express the same referential value or grammatical function (Labov, 1966/2006, 1969). Establishing ‘sameness’ is crucial; it involves locating the specific context(s) in which different variants may alternate with no change in referential meaning, even if elsewhere they are imbued with different senses or functions. This is known as *circumscribing the variable context*. Underlying this procedure is the *Principle of Accountability* (Labov, 1966/2006, 1972), which requires that each alternative way of ‘saying the same thing’ be situated with respect to every context where it *could* have occurred in the relevant environment, even if it did not. A foundational working

principle is that the structure – grammatical and social – that underlies the heterogeneity can be inferred from the distribution and conditioning of these competing variants. Application of the *Principle of Accountability* also transports the analyst from the realm of raw numbers into that of proportions, thereby *contextualizing* the form of interest in discourse, rather than relying on categorical perception or intuitions about its frequency and/or distribution. Such contextualization is a key component of the variationist framework.

1.2 Applications to language contact

In what follows we illustrate the application of this paradigm to the three major linguistic manifestations of language contact: lexical borrowing, multiword code-switching, and grammatical convergence. For the purposes of this discussion we construe language contact as a situation in which members of a speech community access two or more languages *regularly* in the course of their normal interactions. We define lexical borrowing as the act of incorporating words originating from one language, to which we refer as the *donor language* (L_D), into the discourse of a *recipient language* (L_R). This characterization distinguishes the process of borrowing from retrieval of already established loanwords, an operation which need not implicate L_D at all, and which in fact basically mirrors lexical access of native items. Variationist methodology has revealed that the main mechanism of borrowing as defined here is the conversion of other-language material into L_R grammatical structure. The prototypical code-switch, on the other hand, involves juxtaposition of a (usually, but not necessarily, multiword) stretch of one language with a stretch of another. The same methodology reveals that neither language acts as recipient in code-switching; instead, each retains the grammatical structure of its respective lexifier (Poplack, 2018; Poplack and Dion, 2012; Poplack and Meechan, 1998a). The identity of the grammar giving rise to these two types of language mixing (L_R in the case of borrowing; L_D in the case of code-switching) is thus their single greatest distinguishing feature. In many cases, however, overt evidence of the operative grammar is lacking, often because the inherent variability that characterizes the grammar in one or both languages involves *null* elements (e.g., Budzhak-Jones, 1998; Eze, 1997, 1998; Sankoff, Poplack and Vanniarajan, 1990). To complicate matters, though the quintessential borrowing consists of a single word, and the quintessential code-switch of several, neither of these mixing types is limited to these sizes. Single-word code-switches are also theoretically possible. Differentiating them from single-word borrowings requires the kind of systematic quantitative comparison associated with variationist sociolinguistics (Poplack, 2018; Poplack and Dion, 2012; Poplack and Meechan, 1998a). The proper classification of language contact phenomena (and the ensuing discovery that single-word code-switching is generally avoided in favour of borrowing) ranks among the major contributions of this approach to the study of bilingual behaviour.

The third major outcome of language contact, contact-induced change, or *convergence*, is notoriously difficult to identify (e.g., Poplack, 1997; Poplack and Levey, 2010; Poplack, Zentz and Dion, 2012). An elementary characterization involves the attestation in a contact variety of a feature that differs, either by its presence or absence, from its counterpart in a pre- or non-contact stage of that variety, and is thought to have arisen due to contact with a presumed source variety. Two key challenges in establishing convergence involve demonstrating (1) that a bona fide change (as opposed to garden-variety variability) has in fact taken place, and if so, (2) that it was actually contact induced. Because most changes are gradual and incremental, this requires a methodological apparatus capable of detecting such alterations, making the variationist framework especially appropriate for this sort of endeavour.

2. Historical overview

The variationist approach has dominated sociolinguistics for well over half a century, but until quite recently, its application to the study of language contact has remained sparse. Most early work in this area, which until the late 1970s basically focused on lexical borrowing or contact-induced grammatical change, was diachronic in nature, often undertaken by historical linguists, and accordingly framed within the qualitative and diachronic methods associated with that field. Another research tradition produced taxonomies of borrowed forms in a variety of languages, but with little regard for whether, when or how they had entered the target lexicon, or how diffused they were at the time of writing (if at all). The work of Weinreich (1953/1968), Haugen (1950a, 1950b, 1969) and Hasselmo (1970) in the mid-twentieth century represented a real turning point in this regard, thanks to their increased focus on the synchronic effects of language contact (often via actual participant observation in bilingual communities), copious illustrations of bilingual behaviours (mainly lexical borrowing and calquing, with some attention to grammatical convergence), and, especially, explicit attempts to categorize or otherwise organize their detailed and extremely rich (not to mention entertaining) descriptions of the different ways in which bilinguals combine their languages. But even this work remained largely anecdotal, with no real indication of whether the strategies described were idiosyncratic to the speaker(s) or communities observed, or more widespread. Nor were all of them dealt with: intrasentential code-switching, the darling of contemporary contact research, was conspicuously absent. As Weinreich (1953/1968, p. 73) famously asserted, this type of language mixing was not the province of the ‘ideal’ bilingual, who might switch from one language to another according to changes in the speech situation, but ‘certainly not within a single sentence.’ His influential characterization may explain why this area remained understudied (if not unstudied) until decades later. Interestingly, it was the analysis of intrasentential code-switching, in particular the work of Pfaff (1976, 1979) and Poplack (1980, 1981) that introduced the quantitative paradigm into the realm of language contact research. Despite the undeniable presence of numbers, however, that work cannot accurately be qualified as variationist *strictu sensu*. One reason is because ‘the crucial task in applying accountable reporting is knowing what to count’ (Pfaff, 1979, p. 295), and in those early days researchers with an empirical bent were still struggling with that issue. Another is that the core element of the variationist paradigm, *quantitative reasoning*, was not meaningfully invoked until much later (see Section 4.2).

3. Critical issues and topics

Recognizing that many linguistic manifestations of language contact are difficult to identify *a priori* and out of context, variationists have sought empirical answers to a number of questions. Notable among them are the following:

- Are all types of language mixture instantiations of the same process, or do they result from different processes with distinct outcomes?
- Can the behaviour of lone L_D -origin elements and multiword fragments both be accommodated under a single theory?
- Under what conditions are L_D -origin items integrated into an L_R , and why do some of them appear to resist integration?
- Does code-switching give rise to (nonce) borrowing and thence to established loanwords?
- What is the role of the community in conventionalising language mixing strategies, and how do these conventions interact with linguistic constraints on language mixing?

- Is change an inevitable outcome of language contact?
- Can contact be said to *accelerate* change, and if so, how can this be established?
- Which areas of the grammar (if any) are most resistant to change?
- How can contact-induced change be recognised in bilingual discourse and how can it be distinguished from garden-variety variability?

4. Main research methods

In this section we review the methodological and analytical tenets associated with the variationist framework that are most relevant to issues in language contact. On the data side, these include the principled selection of participants and their situation in social context, and the pre-eminence of spontaneous speech reflecting their actual usage; on the analytical side, the imperative to circumscribe the object of study, and the importance of contextualization, comparison and quantitative reasoning in elucidating language contact phenomena.

4.1 Data

4.1.1 Community

The starting point for variationist analysis, of monolingual as well as bilingual discourse, is the *speech community*, loosely defined as a group of individuals who share a set of norms regarding the use of language (Labov, 1966/2006). In a speech community where more than one language is spoken, these norms extend to all of them. Many may be characterized as *communities of practice* (Eckert and McConnell-Ginet, 1992; Holmes and Meyerhoff, 1999), whose members are privy to a shared repertoire of linguistic practices, even if they do not all engage in each of them. Members alternate between the languages regularly, whether actively or passively, not only to accommodate different interlocutors and interactions, but even with the same interlocutors, in the same interactions and within the same conversational topic. Taking the speech community as a reference point enables researchers to consider individuals of varying proficiencies, and distinguish idiosyncratic utterances from the systematic patterns that characterize regular exchanges among members. Communities studied from this perspective have been found to display striking commonalities in the bilingual strategies they draw on – the major one being lexical borrowing (Poplack, 2018), with a good deal less multiword code-switching (Poplack and Dion, 2012) or grammatical change (Poplack and Levey, 2010). The detailed *patterning* of these uses may differ from one community to the next, however, even those featuring the same languages in comparable contact situations (Poplack, 1980, 1981, 1985; Poplack et al., 2015). Such differences as may arise are not fully predictable from either the linguistic configuration of the languages in contact or the extra-linguistic characteristics of the contact situation (Poplack, 1985). This fact underlines the power of *community norms* in elucidating language contact phenomena (Poplack, 2018, p. 214). Adherence to community norms may result in conventionalized use of constructions considered ungrammatical under certain theories, while violation of community norms may lead to rejection of otherwise grammatical mixed utterances. This confounds linguistic inquiry and highlights the danger of analyzing material out of context.

The nature and makeup of the community may also be marshalled to test hypotheses about extra-linguistic factors often claimed to be explanatory of contact-induced phenomena; for instance, the longer and more intense the contact, and the fewer and more marginal the speakers, the greater the likelihood of language mixing and structural interference (e.g., Romaine,

1995; Thomason, 2001; Winford, 2003; among many others). Independent variables like intensity and length of contact, status of the languages in the community (minority or majority), and size of speaker populations can be built right into the selection of communities and speakers (e.g., Poplack, 1989; Poplack, Walker and Malcolmson, 2006), allowing the analyst to assess empirically which (if any) may in fact be operating. Most importantly, change, perhaps the most widely invoked result of contact, can only be established on the basis of diffusion, and for diffusion to proceed, a community is required.

4.1.2 *Speakers*

Recognizing that communities are constituted of individuals of different social classes, educational levels, genders, etc., and that membership in these cohorts may affect linguistic production profoundly, variationists seek to represent the extra-linguistic axes relevant to the community (or the research question) by targeting specific subgroups of participants for study. In a bilingual community, sampling protocols can be guided by hypotheses about the factors that may affect the various contact phenomena, and speakers can be specifically selected to represent individual dimensions of the contact axis. According to one such hypothesis, individual level of bilingualism is inversely correlated with mastery of the minority-language grammar, such that the greater the proficiency in the other language, the more likely the possibility of simplification or loss of minority-language linguistic structure. Another hypothesis proposes that the incorporation of majority-language material (via code-switching or borrowing) into otherwise minority-language discourse may bring with it associated grammatical properties, and these in turn may lead to structural convergence (e.g., Backus, 2005; Thomason, 2001; Winford, 2003). A speaker's propensity to code-switch would thus be a predictor of contact-induced change. Individual attitudes toward, and relative prestige of, each of the languages could also affect the extent and direction of change (e.g., Appel and Muysken, 1987, p. 158; Romaine, 1995, p. 66; Thomason, 2001). These and other hypotheses can be tested on linguistic production data provided by speakers who represent different combinations of the relevant conditions, thus enabling us to ascertain which are most explanatory of contact-induced phenomena, and which community members are most likely to lead them. Whatever the hypothesis, speaker selection should always be informed by the recognition that bilingual communities are typically composed of individuals with widely varying competencies in the languages involved; yet as a rule, they nonetheless develop strategies to function – to some extent – in the bilingual mode. The importance of individual bilingual ability for the correct interpretation of bilingual behaviour cannot be overemphasized; it must be controlled for each participant (either by stratifying the speaker sample accordingly, or otherwise identifying and accounting for individual differences at the analysis stage), even if no particular proficiency-based conditions for sample inclusion are imposed.¹

4.1.3 *Speech*

Once a set of speakers has been selected, variationists seek to record as much of their high-quality bilingual discourse as possible. Good data is paramount, since whatever material is available necessarily informs the researcher's understanding of the extent, structure and even existence of the phenomena under study. A specific goal is to tap into the *vernacular* that characterizes casual or intimate situations: this is taken to reflect the most systematic form of the language (Labov, 1966/2006). Access to the vernacular is also of particular practical importance in the bilingual context, since in more formal or closely monitored speech styles

speakers may avoid the bilingual phenomena of interest altogether (Poplack, 1981). It has already been amply observed, for instance, that bilinguals generally do not mix their languages out of *need*; they do so when they deem it appropriate for the situation. An additional motivation for analysis of the vernacular is to provide a benchmark for the structure of (the speaker's) L_R . Since bilinguals do not integrate L_D items into an L_R they do not speak (such as the idealized standard language that is often implicitly appealed to when comparisons are effected), familiarity with the vernacular L_R benchmark enhances our understanding of how speakers handle other-language material. Reference to L_R turns out to be one of our most powerful tools in resolving such refractory issues in language contact research as how words are borrowed, whether contact-induced change has occurred and the extent to which other-language structures at all levels of the grammar have been integrated, among many others.

4.1.4 Corpus

The most fertile means of exploiting data, especially bilingual data, is via a *corpus*. The corpus is typically a finely transcribed representation of the recorded speech of the individuals preselected for study. To the extent that corpus builders attend to issues of accessibility and searchability, interrogating the corpus can be a comparatively quick and easy way to locate the often elusive and infrequent contact phenomena of interest within a sea of unmixed discourse. Another advantage of well-structured corpora is that they can be marshalled to target a wide variety of research questions, provided that the initial criteria for sample constitution are principled. Any linguistic phenomenon can be used to advantage for this purpose, whether or not it was suspected to be a product of contact at the outset. For example, in the Ottawa-Hull French corpus (Poplack, 1989) or the corpus of Spoken Quebec English (Poplack, Walker and Malcolmson, 2006), a variety of linguistic variables (e.g., pro-drop, preposition stranding, auxiliary alternation, subjunctive expression, etc.) have been successfully deployed to test the effects of extra-linguistic factors, because each speaker instantiates a point on the relevant continua (Leroux and Jarmasz, 2006; Poplack, Zentz and Dion, 2012). Where cohorts representing the different independent variables (e.g., minority vs. majority status) display different patterning of variation, it may be inferred that that variable is operative (in some capacity, depending on its direction of effect; see Section 4) in contact-induced change.

To summarize, the commitment to speakers and their vernacular speech is perhaps most emblematic of the variationist paradigm, and happily, these considerations have been gaining a good deal of traction amongst language contact researchers. This has resulted in the current predilection for bilingual speech corpora, of which there are now a considerable number, covering a wide variety of language pairs in many different locations (e.g., Adamou, 2016; Backus, 1996; Bullock and Toribio, 2013; Deuchar et al., 2014; Gardner-Chloros, 1991; Guzmán et al., 2017; Nortier, 1990; Otheguy, Zentella and Livert, 2007; Torres Cacoullous and Travis, 2018; Treffers-Daller, 1994, among many others). This is good news for description, fact-finding, and observation, even if systematic cross-linguistic and/or cross-corpus comparison and validation of the same language-contact phenomena remain the exception rather than the rule. However, data, corpora, and the methodological desiderata for obtaining them are basically only conduits to the real business of variationist sociolinguistics: analyzing the *structure* of variability. This structure, often expressed in terms of the constellation of factors that together conspire in the choice of one variant over another, can then be rallied to address many key questions about language contact, through quantitative reference to both speakers and data.

4.2 Analysis

4.2.1 Quantitative analysis and quantitative reasoning

Some researchers have in fact generated numbers out of their materials, reporting rates of various contact phenomena and in some cases correlating those rates with the speakers who produced them. But relatively few have applied the *quantitative reasoning* (Labov, 2004) associated with the variationist paradigm to the analysis and interpretation of the data they collected, and even fewer have availed themselves of what many practitioners consider its most valuable appurtenance: the capacity to decide amongst competing theories of bilingual behaviour.

One reason may stem from the fact that strict application of the construct of ‘linguistic variable’ to language contact data is not straightforward. A variable – the centrepiece of the variationist paradigm – is conceptualized as a linguistic entity that can be expressed by two or more competing variants, and defining it requires establishing a closed set consisting of an ‘application value,’ which is construed as alternating with one or more ‘non-applications.’ These alternating variants may both be overt (as with French auxiliary *avoir* vs. auxiliary *être*) or one may be null (e.g., AAVE \emptyset copula vs. contracted and full counterparts). The axioms of probability theory are then applied to determine which variant is preferred in a given context. Such alternations generally occur at the phonetic, phonological, morphological, and (less often) syntactic levels. In bilingual speech, where the elements of interest are often located at the discourse level, determining the *non*-application value of, say, a code-switch or a borrowing is challenging – though not impossible (Sankoff and Poplack, 1981). This is because both the variable context (the *locus* in discourse where the variation is permissible) and the non-applications (e.g., the *non*-code-switched counterparts of the code-switched words) are difficult to pinpoint. In light of this, it is perhaps not surprising that this technical aspect has not yet been widely implemented in most language contact studies.

For many practitioners, however, the crux of the variationist enterprise resides not so much in the linguistic variable per se, but in the capacity its analysis furnishes for the operationalization and testing of hypotheses. This brings empirical evidence to bear on competing theories, through what Labov (2004) terms ‘quantitative reasoning.’ One key type is ‘reasoning from ambiguity’: when the status of an element is unclear and qualitative arguments fail to converge on a solution, the rational quantitative strategy is to treat it as a separate category. A particularly fruitful application of this principle involves the identification and characterization of nonce borrowings, locus of long-standing controversy in the field. Depending on the analyst’s theoretical leanings, these have alternately been identified as loanwords, code-switches, or even dismissed altogether (Gardner-Chloros, 2009; Haspelmath, 2009), with no consensus in sight as to their status, despite decades of debate. Who is right, and how can we tell? Quantitative reasoning enjoins us to rely in the first instance on *unambiguous* instantiations of language mixing (i.e., known code-switches and attested loanwords), and then to systematically compare their properties to each other and to the residual category of ambiguous tokens (in this case, unattested lone other-language items, which are potential candidates for loanword or code-switch status).

4.2.2 Cross-variety comparison

What, specifically, should be compared? Similarities in surface form can be misleading, since they may result from borrowing or transfer, which would justify the inference of change, but they may also be due to interlingual coincidence or to linguistic universals. To ensure diagnosticity

and to rule out alternative explanations, we rely on *conflict sites* (Poplack and Meechan, 1998a, p. 132): functional, structural, and/or quantitative differences between the languages in contact. The conflict site, along with detailed cross-variety comparisons, play a crucial role in identifying contact phenomena, detecting change and identifying their ultimate provenance.

Much of the empirical work in the field of language contact revolves around frequencies. As detailed by Poplack and Tagliamonte (2001, p. 92), these must be used with caution to infer change – contact-induced or otherwise. The key diagnostic in assessing the relationship and provenance of forms is the *constraint hierarchy*: the configuration of environmental factors affecting the probability that a given variant form will be selected, along with the direction of their effects. Constraint hierarchies may be construed as a portion of the grammar underlying the variability. To the extent that they are language-specific, they are powerful indicators of language membership and language change. Departing from the principle that switching to another language involves shifting not only to its lexicon but also to its grammar, while borrowing from another language involves adapting material from that language to the grammar of the L_R , Poplack and Meechan (1998a; see also Poplack and Tagliamonte, 2001; Tagliamonte, 2002) introduced a ‘comparative method’ incorporating systematic scrutiny of the behaviour of unambiguous *benchmarks*. They hypothesized that where rates and (especially) conditioning associated with linguistic variability are language-specific, they can serve as diagnostics of language membership:

If a set of donor-language (L_D) items embedded in otherwise recipient-language (L_R) discourse features the same hierarchy of constraints as their L_R counterparts (while simultaneously differing from those conditioning L_D items in L_D discourse), we may conclude that the grammar constraining the L_D items in otherwise L_R discourse is that of L_R .

(Poplack and Meechan, 1998a, p. 130)

The same reasoning applies to the assessment of convergence:

If the hierarchy of constraints conditioning the variable occurrence of a candidate for change [...] in a contact variety is the same as that of its pre-contact precursor, while differing from that of its presumed source, no structural change has taken place. If it features a constraint hierarchy different from those of both its pre-contact precursor *and* the presumed source, we can infer that change has occurred, but not one that is contact-induced. Only when a candidate for change in a contact variety features a constraint hierarchy *different* from that of its pre-contact precursor, but *parallel* to that of its presumed source, can we conclude in favour of contact-induced change.

(Poplack and Levey, 2010, pp. 400–401)

In what follows we illustrate the utility of comparing the structure of other-language items with that of counterparts in L_R and L_D , as well as with other relevant benchmarks. In cases of presumed change, the element(s) under study may be further contextualized diachronically and synchronically, with respect to a pre-contact stage of the language and a non-contact variety, as well as amongst speakers of varying bilingual profiles.

4.2.2.1 APPLICATION OF THE COMPARATIVE METHOD TO LANGUAGE MIXING

Consider the case of lexical borrowing, where L_D items undergo morphosyntactic *integration* into L_R grammar. When the linguistic behaviour of L_D -origin elements corresponds

categorically to that of L_R counterparts (in displaying case-markers or inflections from that language that are non-existent in L_D , for example), it can straightforwardly be inferred that speakers are applying the grammar of L_R . Thus, for example, English-origin predicate adjectives are preposed to Persian copulas instead of postposed as required by English, while English-origin attributive adjectives enter into the Persian $N + \textit{ezafeh} + \textit{ADJECTIVE}$ construction instead of preceding the noun, as they would in English (Moinzadeh, 1999). However, as has often been observed, such (quasi-)categorical correspondences appear to be the exception rather than the rule, which has led many researchers to despair of establishing constraints on bilingual behaviours like code-switching or borrowing. Applying the variationist comparative method, on the other hand, reveals that much of this apparent vacillation arises from variability inherent to one or both of the contact languages, i.e., alternation amongst competing variants. In many language pairs, for instance, L_D nouns occasionally (or frequently) fail to display the purportedly obligatory case-marking of their L_R counterparts (e.g., Budzhak-Jones, 1998), verbs may go uninflected or adopt apparently idiosyncratic conjugation and inflection strategies (Moinzadeh, 1999; Muysken, 2016; Wichmann and Wohlgemuth, 2008), word order may not hew to the canonical L_R pattern, and so on. In French, for example, a language in which adjectives canonically follow the noun, English-origin nouns are sometimes found with preceding French adjectives. How to analyze these seemingly refractory cases? Systematic comparison with the relevant benchmarks reveals that the answer resides in (often previously unreported) variability in L_R . Thus adjective placement in French-English bilingual discourse mirrors exactly the conditions for preposing and postposing adjectives in French L_R (Turpin, 1995), while ‘lack’ of case markers on some English-origin nouns in languages as diverse as Tamil (Sankoff, Poplack and Vanniarajan, 1990), Ukrainian (Budzhak-Jones, 1998), or Japanese (Yoshizumi and Poplack, 2012) is shown to pattern in parallel with the variable selection of null and overt markers in the unmixed varieties of those languages. Such quantitative parallels in the distribution of overtly and null-marked other-language elements are too detailed to be due to chance. But far more convincing than *rate* parallels are those involving the *conditioning* of variant choice. The study of conditioning addresses the other major question that preoccupies variationists: *why* speakers select one variant over another. As Labov specifies (2004, p. 3), once the variable context is established, the major task is to define the independent variables that are hypothesized to affect variation, themselves dictated by the linguistic or extra-linguistic problem under investigation. These represent the constraints on the variability.

Consider the apparently erratic behaviour of English-origin verbs in Igbo. Most (89%) are inflected with Igbo morphology, albeit (seemingly inexplicably) restricted to only two of the many Igbo verbal categories: participial prefix and affirmative indicative suffix. The others are bare, presumably following an English model. What explains this disparate treatment? Comparison with the unmixed Igbo benchmark reveals striking parallels, not only in rate of inflection (90% in Igbo vs. 89% in English), but in conditioning of verb inflection both across Igbo verbal categories and across serial vs. non-serial (10% vs. 11%) constructions. All of the bare English-origin verbs are headed by the same Igbo light verb *me* ‘make, do,’ which carries the grammatical information. Since both the inflections and more tellingly, their distribution, derive from Igbo grammar and are completely alien to English, the only reasonable explanation of the variable behaviour of English-origin verbs in Igbo is that they are operating under Igbo grammar; i.e., they have been borrowed. Many more examples like these can be found in Poplack (2018), confirming that previously unexplained vacillation can generally be shown to parallel hidden L_R constraints. Even unpredictable strategies, like avoidance of inflecting certain other-language items at all, as documented for French possessive nouns in Tunisian

Arabic, can be shown, by the comparative variationist method, to derive from the variable structure of L_R (Poplack et al., 2015).

4.2.2.2 APPLICATION OF THE COMPARATIVE METHOD TO CONTACT-INDUCED CHANGE

Similar considerations apply to the study of convergence, another area that is fraught with controversy. We defined convergence previously as a change in L_R grammar supposedly due to the influence of L_D . This is widely considered a predictable, even inevitable, outcome of language contact; indeed, it is often claimed that contact accelerates change. Poplack and Levey (2010) review reasons why change is so readily inferred in this context, concluding that the inherent variability characteristic of spontaneous speech is chief among them. When non-standard variants are compared to an idealized (and concomitantly invariant and unchanging) standard variety rather than to relevant pre- and non-contact varieties or to the supposed source of the putative change, the variable occurrence of these non-standard variants is often confused with change. How best to ascertain that change has in fact occurred and if it has, how to establish that it was due to contact? The machinery of variationist sociolinguistics – in particular, cross-variety comparisons of the behaviour of diagnostic features in relevant benchmarks – can again be deployed to identify change and ascertain its source. Given the fact that in many languages, in many areas of the grammar, alternations among variant forms persist for centuries, a first requirement is to determine whether observed variability is involved in change. This necessitates reference to an earlier stage of the language; specifically, one commensurate with the variety hosting the candidate for change. The variability most typically identified with change (i.e., that involving non-standard variants) is generally found in spoken vernaculars. As such, the benchmark should also be a spoken vernacular (or an appropriate surrogate thereof); it should be pre-contact, and so on. Change cannot be inferred simply on the basis of differences from an idealized standard variety. If it has occurred, the innovation should be absent (or incipient or differently conditioned) in the earlier stage. It should not be idiosyncratic to an individual, but should have achieved a certain level of diffusion in the community, where it would be observable in the speech of its relevant members (e.g., the younger cohort, if spearheaded by youth; the highly bilingual, if induced by contact, etc.). The innovation would also be expected to occur in predictable linguistic contexts, as revealed by its position in the wider L_R linguistic system, contextualized both synchronically and diachronically. We must also establish the extent to which it has gained a foothold, its current role in the system and whether it has replaced a native form in one or more functions. Because changes examined synchronically are likely to still be in progress (and thus continue to feature variability), we must again appeal to the *structure* of the variability, as it emerges from the constraints conditioning variant choice. Variationists recognize, in ascending order of importance, change in frequency of one or more competing variant forms, change in statistical significance of one or more factors contributing to variant choice and change in linguistic structure. These elude casual observation, but can be detected with the aid of quantitative analysis and quantitative reasoning. Once it has been established that a change has indeed taken place, it remains to demonstrate that it is contact-induced and not the product of drift, a step that is usually overlooked in the literature. Poplack and Levey (2010) characterize a candidate for contact-induced change as one that is:

present in the presumed source variety and either 1) absent in the pre-contact or non-contact variety, or 2) if present [. . .], is not conditioned in the same way as in the source,

and 3) can also be shown to parallel in some non-trivial way the behaviour of a counter-part feature in the source.

(Poplack and Levey, 2010, p. 398)

This can only be ascertained through systematic quantitative comparisons, of a diagnostic linguistic feature, with an earlier or pre-contact stage (or an appropriate surrogate thereof), with a non-contact variety, and most important, with the presumed source. Such multiway comparisons on linguistic variables that are widely construed as contact-induced changes (e.g., ‘loss’ of the subjunctive (Poplack, 1997) or the advent of preposition stranding (Poplack, Zentz and Dion, 2012) in French; subject expression in Spanish (Torres Cacoullos and Travis, 2018; Otheguy, Zentella and Livert, 2007)) often reveal that ongoing variability owes nothing to change, contact-induced or otherwise.

Summarizing, comparisons are at the heart of inferring change, and therefore figure at least implicitly in all studies of language contact. But many have been carried out without reference to the *Principle of Accountability*, or involve comparison with an ill-suited idealized benchmark. Still others are simply anecdotal. Conspicuous in the variationist program is the rigor of the scientific approach brought to bear on the comparison. The comparative variationist enterprise contributes not only a methodological apparatus – the construction, statistical analysis, and linguistic interpretation of suitable benchmarks for comparison (L_R , L_D , established loanwords and multiword code-switches in language mixing research; pre-contact, non-contact, and post-contact varieties in the study of convergence), but also the *critical capacity to discriminate among competing hypotheses*. The focus is on identifying empirical criteria capable of both detecting change and ascertaining its source, and testing hypotheses empirically to determine goodness of fit with the data of actual usage. This involves principled data collection, enunciation and operationalization of hypotheses and their statistical evaluation in large-scale corpora.

5. Current contributions and research

Application of the methodological and analytical tenets of the variationist paradigm has contributed to the elucidation, if not the resolution, of many long-standing disagreements in language contact research. Many of these involve the identification and characterization of language mixing phenomena. One prominent discovery involves the contentious category of nonce borrowing. Its identity, nature, and even existence have been the subject of enduring controversy in the field. Systematic empirical analysis revealed not only that this is a core component of bilingual discourse, but that it can be readily identified in discourse and distinguished from code-switching. This could only be established through ‘reasoning from ambiguity’: L_D -origin nonce items had to be distinguished operationally from both attested loanwords and the opposing benchmark category, multiword code-switches. Each was first analyzed in its own right and then systematically compared with the others. The hallmark of lexical borrowing was found to be linguistic integration into the morphology and syntax of L_R , including any variability they may feature. Thus if an L_R favours case-marking in dative over accusative contexts, that patterning will be evident in L_D -origin words that have been borrowed. Where more than one avenue of integration is available, the one adopted will generally correspond to the major L_R pattern, barring any specific community-based preference. Lone L_D words, both nonce and established, were found overwhelmingly to undergo such integration. Even where one type of integration is eschewed, as appears to be the case for inflection of borrowed nouns in many communities where Arabic is the recipient language (Heath, 1987, 1989;

Owens, 2002), borrowed items tend to display other unambiguously L_R grammatical characteristics (Poplack et al., 2015). Because their surface linguistic characteristics are shared, there is no reason to distinguish these two types of borrowed word other than on extra-linguistic grounds: they differ only in terms of recurrence and diffusion across the community. Code-switches, on the other hand, contrast both in constitution and in placement with nonce and established borrowings. The internal constituency and positioning in the clause of borrowed words come from L_R , whereas the internal constituency of code-switches is that of L_D , although the placement of the latter tends to respect the word order requirements of both languages involved (Poplack, 1980). The empirical distinction between code-switching and borrowing is perhaps the greatest methodological contribution of this paradigm to the field of contact linguistics, since it allows researchers to construct theories based on the material they claim to be explaining rather than on data muddled by disparate phenomena operating out of different variable contexts and behaving in demonstrably different ways.

Many of these and other unexpected findings on language mixing (see Poplack (2018) for detail) could not have been intuited; they only emerge from systematic analysis of bilingual discourse on the ground. To name but a few: although loanwords are often thought to originate as code-switches, systematic analysis shows that lone L_D items are rarely code-switched; instead, they are borrowed – sometimes only for the nonce. Diachronically, the very first mention of a nonce form already features the full complement of morphosyntactic integration into L_R and, synchronically, it is treated like an attested loanword. Most nonce borrowings are ephemeral, disappearing after their first mention. Vanishingly few achieve the status of attested, or even widespread, loanwords. Phonetic and morphosyntactic integration are independent. The latter is emblematic of borrowing; the former is not pertinent to this process, not only because it remains variable throughout but also because speakers manifestly do not manipulate it strategically or in concert when mixing languages. Only the morphosyntax is a reliable metric for distinguishing language mixing types.

Likewise, when the inference of contact-induced change is pursued scientifically via systematic comparisons across both space and time of appropriate reference varieties, potential candidates often turn out either not to be changes, or not to be contact-induced, but rather cases of ordinary inherent variability firmly rooted in the internal structure of language (Leroux and Jarmasz, 2006; Poplack, 1997; Poplack et al., 2006; Torres Cacoullos and Travis, 2018). Indeed, analysis of a considerable number of presumed and potential changes in different contact situations actually suggests that failure to participate in ongoing changes in tandem with the remainder of the community (i.e., *divergence*) may well be a more common outcome for minority speakers in contact situations than the oft-reported convergence (Poplack, 2008). In any event, the results of variationist analysis enjoin us that contact-induced change cannot simply be assumed, but must be demonstrated.

6. Future directions

Work within the variationist framework has succeeded in overcoming many of the analytical difficulties associated with traditional introspection and anecdotal reporting that characterizes other research paradigms. This is particularly crucial in the study of language contact, where categorical perception tends to inflate the importance of phenomena which may in fact occur only rarely. More important, since the very linguistic manifestations of language contact themselves often defy categorization on casual inspection, a key contribution of this accountable methodology is to help disambiguate many of the *prima facie* ambiguous cases, thereby reducing substantially the amount of uncertainty inherent in any bilingual data set.

An important objective for language mixing research remains to achieve consensus on an empirically verifiable characterization of the rules for combining elements from two or more languages within the confines of the utterance. The study of convergence would likewise be enhanced by greater agreement over what constitutes a change, how to recognize and distinguish one from inherent variability and how to determine whether it is contact induced. To realize these goals, we must improve the fit between theories and data. This would be facilitated by a broader empirical base. To develop an account of the consequences of language contact that is truly accountable to the facts, greater efforts should be made to tease apart the idiosyncratic and/or exceptional uses of individuals from community strategies. This can be accomplished via more reliable analyses of principled collections of bilingual performance, using accepted social science standards of proof, including data-driven reports of rates of occurrence, conditioning of variant choice, and measures of statistical significance.

It is thus to be hoped that existing theoretical, experimental, and anecdotal accounts will be supplemented with many more large-scale empirical studies of actual bilingual behaviour in a much wider variety of well-defined bilingual communities. Apparently, unruly performance facts can be demystified by incorporating more speaker-, context-, and community-based information into our analyses. Consideration of the data of actual bilingual interactions in the context of the speech community in which they were produced would permit researchers to situate bilingual behaviour with respect to the relevant monolingual benchmark vernaculars, account for the disparate mixing strategies that have evolved in different bilingual communities and distinguish among incommensurable manifestations of language contact. Of course, all this will necessarily involve confronting inherent variability rather than abstracting away from it. The versatile methodological machinery of the variationist framework is uniquely positioned to contribute to this endeavour.

7. Further reading

Poplack, S. (2018). *Borrowing: Loanwords in the speech community and in the grammar*. Oxford: Oxford University Press.

This book outlines how bilinguals introduce and adapt foreign words into recipient-language grammatical structure; how these forms diffuse across speakers and communities; how long they persist in real time; and whether they change over the duration. Attacking some of the most contentious issues in language mixing research empirically, it tests hypotheses about established loanwords, nonce borrowings and code-switches on a wealth of unique data sets of typologically similar and distinct language pairs. A major focus is the detailed analysis of *integration*: the principal mechanism underlying the borrowing process.

Torres Cacoullos, R. and Travis, C.E. (2018). *Bilingualism in the community: Code-switching and grammars in contact*. Cambridge: Cambridge University Press.

This book is a model of how to study language contact scientifically. Making use of a new bilingual corpus of English and Spanish spontaneously produced by the same speakers, it highlights variation patterns in the community. Proposing quantitative diagnostics of grammatical similarity, it shows how bilinguals' two languages differ from each other, aligning with their respective monolingual benchmarks. The authors argue that grammatical change through contact is far from a foregone conclusion in bilingual communities.

Poplack, S., Zentz, L. and Dion, N. (2012). Phrase-final prepositions in Quebec French: An empirical study of contact, code-switching and resistance to convergence, keynote article. *Bilingualism: Language and Cognition*, 15(2), pp. 203–225.

This study demonstrates how the comparative variationist framework can be implemented to investigate whether a stereotypical non-standard feature of North American French, preposition stranding,

results from convergence with English, and the role of bilingual code-switchers in its adoption and diffusion. Detailed comparison of the variable constraints on stranding in all the relevant benchmark varieties turned up several lines of evidence militating against the interpretation of convergence. Most compelling are the findings that the conditions giving rise to stranding in French are the same as those operating to produce a similar native strategy, while none of them are operative in the presumed source.

Poplack, S. and Meechan, M. (1998b). How languages fit together in code-mixing. In: S. Poplack and M. Meechan, eds., *Instant loans, easy conditions: The productivity of bilingual borrowing*, Special Issue, *International Journal of Bilingualism*, 2(2), pp. 127–138.

This introductory chapter to a guest-edited issue of *International Journal of Bilingualism* enunciates the first comprehensive empirical method, deriving from variationist sociolinguistics, to distinguish between code-switching and borrowing. Applied to a wide variety of typologically different language pairs, it confirms the ‘nonce borrowing hypothesis’ (Sankoff, Poplack and Vanniarajan, 1990) and demonstrates, in conjunction with the other articles in the volume, that (1) the vast majority of other-language incorporations in the data are borrowings – nonce or established – and (2) their structure differs from that of code-switches.

8. Related topics

Processing multilingual data, usage-based approaches, borrowing, code-switching, convergence

Abbreviations

| | |
|----------------|-------------------------------------|
| L _D | Donor language |
| L _R | Recipient language |
| AAVE | African American Vernacular English |
| N | Noun |

Notes

* The research on which this chapter is based was generously funded by the Social Sciences and Humanities Research Council of Canada in the form of Insight grants and the Canada Research Chair in Linguistics. I am grateful to Nathalie Dion and Suzanne Robillard for their material and moral support in the preparation of the manuscript.

1 Though the ideal would obviously be to stratify the sample according to each independent variable of interest, in practice this would prove onerous if not impossible. This is especially true with respect to bilingual ability, as it would require administering some sort of proficiency measure (in itself far from straightforward) *prior* to collecting any data. We have tended instead to characterize proficiency levels *relatively* and after the fact, through analyses of self-report and actual linguistic performance (e.g., Poplack, 1989; Poplack, Walker and Malcolmson, 2006). This procedure has succeeded in yielding speaker samples featuring a wide range of abilities (as expected and desired in a bilingual community), though it does not result in equal numbers of speakers in each proficiency cohort. The latter is not a problem, since it is in any event unlikely that such a distribution would be obtained in any bilingual community.

References

- Adamou, E. (2016). *A corpus-driven approach to language contact: Endangered languages in a comparative perspective*. Boston and Berlin: De Gruyter Mouton.
- Appel, R. and Muysken, P. (1987). *Language contact and bilingualism*. London: Edward Arnold.

- Backus, A. (1996). *Two in one: Bilingual speech of Turkish immigrants in the Netherlands*. Tillburg: Tillburg University Press.
- Backus, A. (2005). Code-switching and language change: One thing leads to another? *International Journal of Bilingualism*, 9(3–4), pp. 307–340.
- Budzhak-Jones, S. (1998). Against word-internal codeswitching: Evidence from Ukrainian-English bilingualism. In: S. Poplack and M. Meechan, eds., *Instant loans, easy conditions: The productivity of bilingual borrowing*, Special Issue, *International Journal of Bilingualism*, 2(2), pp. 161–182.
- Bullock, B.E. and Toribio, A.J. (2013). *The Spanish in Texas corpus project*. The University of Texas at Austin: Center for Open Education Resources and Language Learning. [online] Available at: www.coerll.utexas.edu/coerll/project/spanish-texas-corpus
- Deuchar, M., Davies, P., Herring, J.R., Parafita Couto, M. and Carter, D. (2014). Building bilingual corpora: Welsh-English, Spanish-English and Spanish-Welsh. In: E.M. Thomas and I. Mennen, eds., *Advances in the study of bilingualism*, 1st ed. Bristol: Multilingual Matters, pp. 93–110.
- Eckert, P. and McConnell-Ginet, S. (1992). Communities of practice. In: K. Hall, M. Bucholtz and B. Moonwomon, eds., *Locating power*. Berkeley: Berkeley Women and Language Group, pp. 89–99.
- Eze, E. (1997). *Aspects of language contact: A variationist perspective on code switching and borrowing in Igbo-English bilingual discourse*. PhD. University of Ottawa.
- Eze, E. (1998). Lending credence to a borrowing analysis: Lone English-origin incorporations in Igbo discourse. In: S. Poplack and M. Meechan, eds., *Instant loans, easy conditions: The productivity of bilingual borrowing*, Special Issue, *International Journal of Bilingualism*, 2(2), pp. 183–202.
- Gardner-Chloros, P. (1991). *Language selection and switching in Strasbourg*. Oxford: Clarendon Press.
- Gardner-Chloros, P. (2009). *Code-switching*. Cambridge: Cambridge University Press.
- Guzmán, G., Ricard, J., Serigos, J., Bullock, B.E. and Toribio, A.J. (2017). Moving code-switching research toward more empirically grounded methods. In: T. Declerck and S. Kübler, eds., *Proceedings of the Workshop on Corpora in the Digital Humanities*. Bloomington, IN, pp. 1–9.
- Haspelmath, M. (2009). Lexical borrowing: Concepts and issues. In: M. Haspelmath and U. Tadmor, eds., *Loanwords in the world's languages: A comparative handbook*, 1st ed. Berlin: De Gruyter Mouton, pp. 35–54.
- Hasselmo, N. (1970). Code-switching and modes of speaking. In: G. Gilbert, ed., *Texas studies in bilingualism*, 1st ed. Berlin: Walter de Gruyter, pp. 179–210.
- Haugen, E. (1950a). The analysis of linguistic borrowing. *Language*, 26(2), pp. 210–231.
- Haugen, E. (1950b). Problems of bilingualism. *Lingua*, 2, pp. 271–290.
- Haugen, E. (1969). *The Norwegian language in America*. Bloomington, IN: University of Indiana Press.
- Heath, J. (1987). *Ablaut and ambiguity: Phonology of a Moroccan Arabic dialect*. Albany: State University of New York Press.
- Heath, J. (1989). *From code-switching to borrowing: Foreign and diglossic mixing in Moroccan Arabic*. London and New York: Kegan Paul International.
- Holmes, J. and Meyerhoff, M. (1999). The community of practice: Theories and methodologies in language and gender research. *Language and Society*, 28(2), pp. 173–183.
- Labov, W. (1966/2006). *The social stratification of English in New York City*. Washington: Center for Applied Linguistics.
- Labov, W. (1969). Contraction, deletion and inherent variability of the English copula. *Language*, 45(4), pp. 715–762.
- Labov, W. (1972). *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.
- Labov, W. (2004). Quantitative reasoning in linguistics. In: U. Ammon, N. Dittmar, K.J. Mattheier and P. Trudgill, eds., *Sociolinguistics/Soziolinguistik: An international handbook of the science of language and society*, vol.1, 2nd ed. Berlin: De Gruyter Mouton, pp. 6–22.
- Leroux, M. and Jarmasz, L. (2006). A study about nothing: Null subjects as a diagnostic of convergence between English and French. *University of Pennsylvania Working Papers in Linguistics: Selected Papers from New Ways of Analyzing Variation* 34, 12(2), pp. 1–14.
- Moinzadeh, A. (1999). Bilingual phenomena: Towards the fateful triangle of language mixture. *Ottawa Papers in Linguistics/Cahiers Linguistiques d'Ottawa (June)*, 27, pp. 31–63.

- Muysken, P. (2016). From Colombo to Athens: Areal and universalist perspectives on bilingual compound verbs. *Languages*, 1(1), pp. 1–18.
- Nortier, J. (1990). *Dutch-Moroccan Arabic in contact: Code-switching among Moroccans in the Netherlands*. Dordrecht, Holland and Providence, RI: Foris Publications.
- Otheguy, R., Zentella, A.C. and Livert, D. (2007). Language and dialect contact in Spanish in New York: Towards the formation of a speech community. *Language*, 83, pp. 1–33.
- Owens, J. (2002). Processing the world piece by piece: Iconicity, lexical insertion, and possessives in Nigerian Arabic codeswitching. *Language Variation and Change*, 14, pp. 173–210.
- Pfaff, C.W. (1976). Functional and structural constraints on syntactic variation in code-switching. In: S.B. Steever, C.A. Walker and S.S. Mufwene, eds., *Papers from the parasession on diachronic syntax*, 1st ed. Chicago: Chicago Linguistic Society, pp. 248–259.
- Pfaff, C.W. (1979). Constraints on language mixing. *Language*, 55(2), pp. 291–318.
- Poplack, S. (1980). ‘Sometimes I’ll start a sentence in Spanish *y termino en español*’: Toward a typology of code-switching. *Linguistics*, 18, pp. 581–618.
- Poplack, S. (1981). Syntactic structure and social function of code-switching. In: R. Duran, ed., *Latino language and communicative behavior*, 1st ed. Norwood, NJ: Ablex Publishing Corp., pp. 169–184.
- Poplack, S. (1985). Contrasting patterns of code-switching in two communities. In: H.J. Warkentyne, ed., *Methods V: Papers from the V International Conference on Methods in Dialectology*. Victoria, BC: University of Victoria, pp. 363–385.
- Poplack, S. (1989). The care and handling of a mega-corpus. In: R.W. Fasold and D. Schiffrin, eds., *Language change and variation*, 1st ed. Amsterdam: John Benjamins, pp. 411–451.
- Poplack, S. (1997). The sociolinguistic dynamics of apparent convergence. In: G. Guy, J. Baugh and D. Schiffrin, eds., *Towards a social science of language*, 1st ed. Amsterdam: John Benjamins, pp. 285–309.
- Poplack, S. (2008). Quebec English. *International Journal of English Studies*, 19(2), pp. 189–200.
- Poplack, S. (2018). *Borrowing: Loanwords in the speech community and in the grammar*. Oxford: Oxford University Press.
- Poplack, S. and Dion, N. (2012). Myths and facts about loanword development. *Language Variation and Change*, 24(3), pp. 279–315.
- Poplack, S., Jones, A., Lealess, A.V., Leroux, M., Smith, C.T., Yoshizumi, Y., Zentz, L. and Dion, N. (2006). Assessing convergence in contact languages. Paper presented at *New ways of analyzing variation 35*. The Ohio State University, Columbus.
- Poplack, S. and Levey, S. (2010). Contact-induced grammatical change. In: P. Auer and J.E. Schmidt, eds., *Language and space – An international handbook of linguistic variation: Volume 1 – Theories and methods*, 1st ed. Berlin: De Gruyter Mouton, pp. 391–419.
- Poplack, S. and Meechan, M. (1998a). How languages fit together in code-mixing. In: S. Poplack and M. Meechan, eds., *Instant loans, easy conditions: The productivity of bilingual borrowing*, Special Issue, *International Journal of Bilingualism*, 2(2), pp. 127–138.
- Poplack, S. and Meechan, M., eds. (1998b). Instant loans, easy conditions: The productivity of bilingual borrowing. Special Issue, *International Journal of Bilingualism*, 2(2).
- Poplack, S., Sayahi, L., Mourad, N. and Dion, N. (2015). An exception to the rule? Lone French nouns in Tunisian Arabic. *Penn Working Papers in Linguistics (Selected Papers from New Ways of Analyzing Variation 43)*, 21(2), pp. 178–186.
- Poplack, S. and Tagliamonte, S. (2001). *African American English in the diaspora*. Oxford: Basil Blackwell.
- Poplack, S., Walker, J. and Malcolmson, R. (2006). An English ‘like no other?’: Language contact and change in Quebec. *Canadian Journal of Linguistics*, 51(2), pp. 185–213.
- Poplack, S., Zentz, L. and Dion, N. (2012). Phrase-final prepositions in Quebec French: An empirical study of contact, code-switching and resistance to convergence. *Bilingualism: Language and Cognition*, 15(2), pp. 203–225.
- Romaine, S. (1995). *Bilingualism*, 2nd ed. Malden, MA: Blackwell.

- Sankoff, D. and Poplack, S. (1981). A formal grammar for code-switching. *Papers in Linguistics*, 14(1), pp. 3–46.
- Sankoff, D., Poplack, S. and Vanniarajan, S. (1990). The case of the nonce loan in Tamil. *Language Variation and Change*, 2(1), pp. 71–101.
- Tagliamonte, S. (2002). Comparative sociolinguistics. In: J. Chambers, P. Trudgill and N. Schilling-Estes, eds., *Handbook of language variation and change*, 1st ed. Malden, MA and Oxford: Blackwell Publishers, pp. 729–763.
- Thomason, S.G. (2001). *Language contact: An introduction*. Edinburgh: Edinburgh University Press.
- Torres Cacoullos, R. and Travis, C.E. (2018). *Bilingualism in the community: Code-switching and grammars in contact*. Cambridge: Cambridge University Press.
- Treffers-Daller, J. (1994). *Mixing two languages: French-Dutch contact in a comparative perspective*. Berlin and New York: De Gruyter Mouton.
- Turpin, D. (1995). *Le français c'est le last frontier. La structure nominale dans le discours bilingue français/anglais*. M.A. University of Ottawa.
- Weinreich, U. (1953/1968). *Languages in contact: Findings and problems*. The Hague: Mouton.
- Wichmann, S. and Wohlgemuth, J. (2008). Loan verbs in a typological perspective. In: T. Stoltz, D. Bakker, and R. Salas Palomo, eds., *Aspects of language contact*. Berlin and New York: De Gruyter Mouton, pp. 89–121.
- Winford, D. (2003). *An introduction to contact linguistics*. Malden, MA: Blackwell.
- Yoshizumi, Y. and Poplack, S. (2012). The status of lone English-origin nouns in otherwise Japanese discourse: A comparative variationist approach to borrowing. Paper presented at *New ways of analyzing variation Asia-Pacific 2*. University of Tokyo, Japan.

The 4-M model

Different routes in production for different morphemes

Janice L. Jake and Carol Myers-Scotton

1. Introduction and definitions

The 4-M model of morpheme classification offers an explanation for critical structures in contact phenomena. The key to such an explanation is how grammatical morphemes are differentially elected at the abstract level/abstract levels of language production. We argue that differences in how morphemes are elected affect the potential for certain morpheme types to occur in contact phenomena. Of course, bilingual speakers are also motivated by psycholinguistic and sociolinguistic factors that are present in any community when they select utterances that result under the influence of one language on another (Myers-Scotton, 1993b). As Matras (2009, p. 123) points out ‘any analysis must take into account the participants’ full range of social networks and social activities.’ However, in this chapter our special interest is how the morphemes that make up contact phenomena are elected. We develop an argument based on the extended 4-M model. As we argued with some felicitous foresight when we first discussed distinctions among morpheme types, ‘the model is less a taxonomy and more a window with a view of how morphemes are elected’ (Myers-Scotton and Jake, 2000, p. 1090).

In this chapter, we demonstrate the utility of the 4-M model in explaining why certain lexical items can cross over from one language to occur in another variety in a given community and other similar elements cannot. The general hypothesis we develop is that only certain morpheme types can cross from one language to another in contact situations. Our specific hypothesis is that the level at which morpheme types are elected in an abstract model of language production affects their ability to become cross-over phenomena. We support this hypothesis with empirical data from contact situations. The phenomena discussed include code-switching, convergence with and without code-switching, mixed languages, and creole development. However, we focus on morphemes in code-switching (hereafter CS) to make our argument. The four types are content morphemes and three types of system morphemes. Early system morphemes are divided from the two types of late system morphemes, bridges and outsiders. In Section 4, Current Contributions and Research, different morpheme types are described and illustrated as they appear in contact data.

2. Historical overview

The 4-M model was introduced as a construct to account for differences in the frequency of different morpheme types in three different types of linguistic data: Broca's aphasia, code-switching, and data in early second language acquisition (Myers-Scotton and Jake, 2000). We show how differences in empirical data can be explained based on their morpheme type. Specifically, we differentiate the distribution of content morphemes (hereafter, CMs) and what we call early system morphemes from late system morphemes (hereafter, SMs) differs.

The central claim developed is that differences in the accuracy and role of morpheme type stems from differences at an abstract level of a general model of the nature of language. That is, the 4-M model is linked to a model of the mental lexicon, suggesting that performance (as different frequencies in data) is linked to competence. Although the 4-M model was first introduced in Myers-Scotton and Jake (2000), the notion that differences in morpheme type underlie different patterns in bilingual data was introduced in the early version of the Matrix Language Frame (MLF) model of code-switching in *Duelling Languages* (Myers-Scotton, 1993a, 1997). This monograph includes several innovations about morpheme type and CS. Two innovations seek to clarify distinctions between CMs and SMs, although this distinction has since been revised and clarified; more accurately characterizing the roles of content versus system morphemes is one motivation for developing the 4-M model.

One innovation in Myers-Scotton (1993a, 1997) is the use of 'system morpheme' as a term for both self-standing morphemes, e.g., determiners and affixes that modify their heads, such as noun plurals or verbal morphology. That is, the terms SM and CM were intended to replace 'open and closed class' elements because closed class includes prepositions and adverbs in addition to some verbal and nominal modifiers that we do not include under SM.

A second innovation in Myers-Scotton (1993a, 1997) characterizes CMs according to their participation in the thematic structure of a clause; this also differentiates them from SMs. CMs receive or assign thematic roles; for example, nouns and pronouns occurring in argument position (but not dummy pronouns, such as weather 'it') receive thematic roles.

What became the cornerstone of the MLF model, the System Morpheme Principle (SMP), is a third innovation in Myers-Scotton (1993a, 1997, p. 83): 'In ML + EL constituents, all system morphemes which have relations external to their head constituent . . . will come from the ML.' Examples of such SMs are morphemes that coindex argument relations within a phrase, such as subject or object agreement in a verb phrase, or case affixes. How the SMP applies to late SMs is clarified in later work (Myers-Scotton, 2002; Myers-Scotton and Jake, 2009). The SMP, along with the Morpheme Order Principle (MOP), identifies the language in a bilingual clause that the MLF model calls the Matrix Language (ML). The other language is called the Embedded Language (EL); it largely contributes CMs to mixed constituents in the bilingual clause.

In CS, the ML is the language that supplies the crucial elements to the clause, notably the late SMs, such as agreement morphology or case, that make more transparent relationships within the clause, especially relationships between arguments and predicates. These are the morphemes that the SMP states must come from the ML in mixed constituents (i.e., constituents with morphemes from both languages). For example, in (1), an example of Swahili-English CS, Swahili affixes illustrate the type of morpheme that must come from the ML, according to the SMP. In this example, the English EL verb *buy* receives a late SM prefix from the ML, Swahili, that co-indexes the subject, *father*. (In all bilingual examples, EL elements appear in **bold**; where possible, a morpheme-by-morpheme gloss has been provided, even when not available in the original source of the example.) The object prefix *-m-* and the suffix

-i- on *buy* indicate that the object is a beneficiary. They are late SMs. The tense and aspect markers are from the ML, Swahili. The consecutive morpheme *-ka-* on the verb of the second clause is in Swahili and establishes relations between the predicates of the two clauses. Only the first clause illustrates CS. The second is entirely in Swahili.

1. **Father** *a-li-m-buy-i-a* *a-ka-potez-a* *vy-ote*
 father 3S-PST-3S-buy-APPL-FV 3S-CONSEC-lost-FV CL.8-all
 ‘Father bought (them) for him and he lost all [of them].’
 (Myers-Scotton, 1993a, 1997, pp. 87–88)

Similarly, in (2), an Italian EL noun *maiale* ‘pig’ occurs with a Swiss German determiner in a clause otherwise in Swiss German, the ML. The Swiss German determiner is in dative case and is a late SM. The ML determiner integrates the NP into the prepositional phrase and the larger clause, in which the EL provides only a CM, a noun.

2. *Hei, verzell ämol was isch mit em maiale*
 hey tell.IMP once what be.3S with DEF.M.DAT pig
 ‘hey, tell [us] again what happened with the pig.’
 (Preziosa-Di Quinzio, 1992, Appendix XX)

The MLF model recognizes another kind of constituent in bilingual clauses, constituents entirely in the EL, EL Islands. For example, in (3) two French EL Islands occur in a clause structured by Moroccan Arabic, the ML.

3. *Un faux geste yqder ydir bezaff d les problems*
 INDEF.M false gesture can.3.IMPERF do.3.IMPERF a.lot of DEF.PL problem-PL
 ‘a false gesture can make a lot of problems’
 (Ziamari, 2008, p. 131)

EL Islands are well-formed constituents in the EL that are integrated into a bilingual clause. In (3), the subject NP is from the EL French, but the ML, Arabic, provides agreement on the auxiliary and main verb. Another NP EL Island occurs as part of a larger quantified NP. Elsewhere Myers-Scotton (2002) and Myers-Scotton and Jake (2017) discuss possible factors contributing to the occurrence of EL Islands in CS.

The MLF model became the starting point for many discussions of the grammatical structure of CS in a variety of language pairs. Myers-Scotton (2002) and Myers-Scotton and Jake (2009) present a more precise way of identifying the scope of the SMP, late outsider SM morphemes.

Because the SMP does not apply to all SMs, it is useful to differentiate among SM types. Unless SMs interact with structure building processes at the functional or positional levels, such as case assignment, the SMP does not apply to them. Under the 4-M model, for example, determiners that are not inflected for case are early SMs, as are most plural morphemes. Plurality, inherent gender, and definiteness realize speaker intentions at the lexical-conceptual level. Such morphemes as these are identified as early SMs in the 4-M model. Later, we discuss early SMs more fully. Briefly, early SMs are referred to as indirectly elected on their route to surface activation (cf. Bock and Levelt, 1994). They depend on their directly elected CM heads for information about their form and meaning. Early SMs add semantic and pragmatic features

that modify the meaning of their heads and realize speaker intentions. These include features of definiteness, plural, and aspect. Like CMs, early SMs are conceptually activated.

In sum, Myers-Scotton (1993a, 1997) is a starting point for understanding the roles that different morpheme types play in a bilingual clause, accounting for their distribution in CS. Later work (e.g., Myers-Scotton, 2002; Myers-Scotton and Jake, 2009, 2017) shows how explicitly connecting a production model to morpheme activation further differentiates types of SMs.

Development of the 4-M model introduces more clarity to discussions of morpheme type in contact phenomena in general, not just CS, such as creole development and convergence features in intense contact zones. First, the model gives a special classification to the morpheme type the SMP was intended to specify as identifying the ML: outsider SMs, discussed later. Second, the 4-M model provides a classification system that identifies morphemes as one of four types. Third, the model goes beyond the SMP in recognizing the roles of all morpheme types in the grammatical structure of bilingual clauses. Fourth, and its most far reaching utility, the model links the surface level distribution of morpheme type to the abstract structures that underlie language production. This explanatory value contributes to our understanding of what elements actually can and do occur in contact phenomena, and has relevance to other linguistic phenomena.

3. Critical issues and topics

Language production and morpheme type

In contrast with many approaches to morpheme classification, the 4-M model focuses the role of morphemes in language production and how they function in constructing the meaning and morphosyntactic structure of a clause. That is, the goals of the 4-M model are not limited to studying how different morphemes participate in word structure or clause building. This model also applies equally to monolingual and bilingual data. The 4-M model leads to hypotheses on how morphemes differ in their election in language processing and provides a language independent means for identifying morpheme type. Importantly, the 4-M model is concerned with the route that morphemes take in language production, as well as the abstract structure of morphemes themselves. See Myers-Scotton and Jake (1995, 2000) for a discussion of the abstract structure of lemmas, entries in the mental lexicon. Predictions the 4-M model makes that are relevant to contact phenomena are based on empirical data, especially the frequency of different morpheme types or constituents containing a specific morpheme type in various types of contact phenomena.

To begin, we review two general principles originally applied to CS data, but relevant to bilingual data in general, the Uniform Structure Principle (USP) and the Asymmetry Principle. Although the need and relevance for the USP is more obvious in bilingual data, it applies to monolingual as well as bilingual data. Myers-Scotton (2002, p. 8) formalizes the USP as

A given constituent type in any language has a uniform abstract structure and the requirements of well-formedness for this constituent type must be observed whenever the constituent appears. . . . In bilingual speech, the structures of the Matrix Language are always preferred.

The USP captures the notion that preferencing the structure of the ML enhances predictability, and predictability is the basis of interpretability.

The Asymmetry Principle states that there is always asymmetry between the participating languages in any bilingual data regarding the ways or the extent that they participate in that

clause. The MLF model is based on this principle: ‘Asymmetry has a dynamic quality; its details depend on changes in relevant factors, but there is always some structural inequality’ (Jake and Myers-Scotton, 2009, p. 209). For example, in classic CS, the language which is the ML may change across clauses in the same corpora, but not within a clause, even though the participating languages remain the same. See Matras (2009, p. 135) for a discussion of this phenomenon. Classic CS contrasts with composite CS, in which the EL influences features of ML late SMs in bilingual clauses as well as provides EL morphemes. See examples of composite CS in Hungarian-English acquisition and attrition in Bolonyai (2002), and in Ewe-English CS in Amuzu (2010). In another, less frequent, kind of language contact, late SMs from both varieties occur in a bilingual clause (excluding EL Islands); such varieties are considered mixed languages. See Bakker (1997) for a discussion of Michif, a Cree-French variety. Adamou and Granqvist (2015) discuss Romani-framed CS with Turkish and Finnish inflected verbs. The 4-M model, together with the assumption that lexemes project abstract lexical structure, enables distinctions to be made across different types of language contact. We discuss how the 4-M model applies to distinguish classic CS, convergence, composite CS, mixed languages, and creole structures in Section 4.

Morpheme types

We illustrate the distinctions among morpheme type by discussing how they are accessed in the production process. First, speaker intentions may directly call morphemes; for example, CMs, such as nouns and verbs, are called this way. Second, some morphemes are indirectly elected by their CM heads; for example, plural and other morphemes realizing the semantic and pragmatic features of speaker intentions are called this way. These are early SMs. Third, the procedures activating some morphemes are called by the structure of a constituent or clause, for example, morphemes realizing the subject-verb agreement. They are structurally assigned late SMs. Myers-Scotton (2002, 2005b) and Myers-Scotton and Jake (2009) introduce and discuss the Differential Access Hypothesis, which suggests that information about late SMs does not become salient until the level of the Formulator.

The 4-M model takes the distinction between CM and SM, first made in Myers-Scotton (1993a) and refines it by adding distinctions among SM types. We use the designations ‘early’ and ‘late’ for SMs for convenience; these terms are more metaphorical than based on psycholinguistic evidence. Early SMs are divided from other SMs, because they have little to do with structure, but instead contribute to the realization of the semantic and pragmatic intentions of CMs. In addition to CMs and early SMs, which are elected at the conceptual level, there are two types of late SMs, both referring to structure, but somewhat differently: outsiders, those identified by the SMP, and bridges. Both types of late SMs are illustrated later in the chapter.

As can be inferred from the preceding discussion, the 4-M model adds a division based on the more abstract notion of conceptually activated or structurally assigned. This opposition emphasizes not the status of morphemes in phrase structure, but implies abstract procedures. Myers-Scotton (2002, p. 74) ‘note[s] how this new opposition enlarges the basis of explanation,’ pointing out how it emphasizes ‘not the status of elements in phrase structure, but their status in the abstract procedures that produce surface statuses.’ This distinction among morpheme types implies that the choices of CMs and early SMs are motivated by satisfying the speaker’s semantic and pragmatic intentions at the conceptual level. In contrast, language-specific grammatical structure becomes salient later in the production process and calls relevant late SMs.

As should be clear, the 4-M model does not classify morphemes according to their status as parts of speech, but rather in terms of how they fall in the dichotomy between conceptually

activated morphemes and those which are structurally assigned. To show the difference this division makes, we include examples of preposition to show how they do not form a uniform class. That is, what is relevant is how a preposition is elected in actual speech. It turns out that preposition is one lexical category that can be realized as an instance of any of the four morpheme types, viewed in terms of the 4-M model. Other lexical categories will also be discussed in regard to how they are classified under the 4-M model. However, prepositions appear to be exceptional because they can be elected by all of the mechanisms underlying morphemes. They can be elected at the lexical-conceptual level and they can be structurally assigned (Jake and Myers-Scotton, 2009).

Content morphemes

Basic communication relies on CMs more than other morpheme types. CMs are directly elected by the speakers' intentions regarding what they wish to communicate; that is, they follow a direct route from abstract levels in production to the surface level (Levelt, 1989). Language production itself is driven by intentions; the semantic and pragmatic content of CMs is what most speakers want to communicate. Notably, nouns are especially easy to transfer across languages in various contact phenomena, not just because they satisfy such intentions, but also because they do not disrupt the hierarchical structure of mapping arguments and predicates in the relevant clause. Nouns participate in thematic role assignment, but only as receivers of thematic roles. In contrast, verbs, prepositions, and in some cases, adjectives assign them.

In CS data, EL nouns occur often and are the most frequent EL elements in a CS clause with an ML frame in the published literature; examples (1) and (2) contain EL nouns framed by another language, the ML. EL adjectives are less frequent than nouns and do not form a unified class. Still, in some corpora, adjectives occur relatively frequently. Example (4) illustrates an English adjective in a Swahili ML frame. In general, English EL adjectives in CS do not receive Swahili agreement markers; that is, they are often bare forms, such as *innocent* in (4). While the EL adjective follows the ML noun *mtu* 'person,' reflecting the word order of the ML, it lacks the agreement marker matching the class 1 prefix of *mtu*.

4. *a-na-onekam-a* *kama* *mtu* *innocent*
 3S-NON.PST-appear-FV as CL1.person innocent
 'he looks like he is innocent'
 (Myers-Scotton, 1993b, p. 77)

Verbs in CS

Even though EL verbs are CMs, inflected EL verbs are very rare in CS corpora or other contact phenomena. Why is this? First, verbal morphology signalling TMA (Tense, Mood, Aspect) or verb class membership raises congruence problems with the ML frame. However, in some cases, EL verb forms are inflected with ML TMA markers. That is, some EL nonfinite verbs (that is, bare verb stems and/or infinitives) appear in the same positions in the verbal assembly as would ML verbs with ML inflections. This was illustrated in example (1). Another similar example is in (5), where the English EL verb *behave* occurs twice in an ML Swahili grammatical frame.

5. *u-na-anz-a* *ku-behave* *kama* *watu* *wa*
 2S-NON.PST-begin-FV INF-behave as CL2.people CL2.ASSOC

huko wa-na-vyo-behave
 there CL2-NON.PST-MANNER-behave
 ‘You will begin to behave as people from there behave.’
 (Myers-Scotton, 1993a, 1997, p. 103)

In the first clause, *behave* occurs with a Swahili infinitive prefix; in the second clause, *behave* has a subject prefix (*wa-* ‘they’), followed by a tense/aspect prefix (*-na-* ‘NON-PAST’) and another prefix indicating manner (*-vyo-*). Such nonfinite EL verbs in ML morphosyntactic frames are discussed at length in Myers-Scotton and Jake (2014, 2015, 2017).

Examples of such inflected nonfinite EL verbs are common when a Bantu language is the ML. (6) is from Congo Shaba Swahili-French CS. The French EL verb *appartenir* comes from the *-ir* conjugation, illustrating how a nonfinite EL verb is inflected with late SMs from the ML, Congo Shaba Swahili.

6. . . . *donc*, (h)ii *richesse* *y-ote* (h)ii
 . . . so CL.9.DEM richness CL.9 – all CL.9.DEM
i-na-tu-appartenir *shi ba-toto* *y-ake*
 CL9-NON.PST-1PL.OBJ-belong.INF us CL.2-child CL.9-his
 ‘So, all this richness, it belongs to us, his children.’
 (de Rooij, 1996, p. 186).

Infinitives are not the only EL nonfinite verbs occurring with ML late SM verbal inflections. In Acholi-English CS in Uganda, for example, the ML is usually Acholi. Acholi verbs are inflected for aspect, and in Acholi-English CS, the nonfinite EL verb form occurring in the ML frame is the English present participle (verb + *ing*) (Myers-Scotton, 2002). For example, in *ka i-boarding taxi* ‘if you board a taxi (if 2S-boarding taxi),’ English *boarding* is inflected with Acholi subject agreement prefix. See example (16) for additional discussion of Acholi-English CS.

Myers-Scotton and Jake (2015, p. 434) have noted a potential psycholinguistic advantage with nonfinite verbs because they do not contribute to or control any elements in the grammatical structure of a bilingual clause: ‘They are less complex because they only meet congruence requirements at the level of the lexical-conceptual structure; that is, the only requirement they must satisfy is meeting the speaker’s intentions regarding semantic/pragmatic content.’ They are CMs, or CMs with early SMs, and do not come with the abstract structural baggage of late SM morphosyntax.

In addition, psycholinguistic studies such as Van Hell, Litcofsky and Ting (2015), suggest that, after a switch into a second language, switching back into the first language appears more costly in response time than was the switch to the second language. These studies are largely limited to single word switches, not CS within a clause. Referring to the notion of cost in CS is not limited to the switch from ML to EL and back. We suggest it also involves referring to the level of production at which morphemes are accessed. EL forms that are accessed at the level of lexical-conceptual structure add less complexity than an inflected verb, which requires information not available until the level of the Formulator, when predicate-argument structure is realized, (Myers-Scotton and Jake, 2009). In sum, nonfinite EL forms within the VP are not complex, although an inflected verb is complex, because ML inflections are late SMs. This may underlie the cost suggested in a switch back to the ML.

Two other strategies frequently integrate EL verbs into an ML frame, the ‘do’ verb construction discussed later and verbalizing affixes. See Sakel (2004) and Matras (2009) for

discussion of verbalizers integrating EL forms into ML grammatical frames. The ‘do’ verb construction is a very common strategy for integrating an EL verb into an ML grammatical frame. In CS, an ML verb meaning ‘do,’ or a related morpheme bleached of specific meaning, is inflected with TMA markers and agreement. The central meaning of the clause comes from an EL nonfinite verb, again often an infinitive, although EL nouns also occur. Backus (1996) has written extensively about the Turkish ‘do’ construction, especially in Turkish-Dutch CS. Typically, the Turkish verb *yap-* is the inflected ML verb. See example (7) with a Dutch EL nonfinite verb, *kijken* (‘watch.INF’).

7. *ja, maar toch, millet kijken yapiyor*
 yeah but still people watch.INF do.PROG.3S
 ‘Yeah, but still, everybody is watching you.’
 (Backus, 1996, p. 238)

In the MLF model, the ‘do’ construction in CS receives an explanation because the ML ‘do’ verb carries all the late SMs required by the SMP. Like other strategies, the ‘do’ verb construction preserves the ML grammatical frame while providing the bilingual with a strategy for realizing meaning with CMs and early SMs from the EL.

Prepositions in CS

Emphasizing election in a production model to identify morpheme type has one important consequence: It explains how surface forms exhibit different distributions depending on the specific composition of the bilingual clause. Consider prepositions, which are not elected through only one procedure; even individual lexemes can be elected via more than one procedure. That is, some prepositions can be elected at the conceptual level, and thus be CMs or early SMs, or structurally assigned late SMs (Jake and Myers-Scotton, 2009).

In some CS corpora, conceptually elected prepositions occur frequently, as in Hebblethwaite’s (2007) corpus of Haitian Creole-English CS; see example (8). Here the EL CM preposition *from* is directly elected by the speaker’s intentions to provide a source as part of the path of the predicate ‘come.’ In (9), an EL preposition *before* introduces an independent temporal phrase, not one that is necessary to complete semantic and pragmatic features of the ML Swahili verb ‘bring.’ Thus, *before* is a directly elected CM.

8. ... *yo djus vini from Ayiti* ...
 ... 3PL just came from Haiti. ...
 ‘... they just came from Haiti. ...’
 (Hebblethwaite, 2007, p. 296)
9. *U-let-e before kesho jioni*
 2S-bring-SUBJUNCT before tomorrow evening
 ‘You should bring [it] before tomorrow evening’
 (Myers-Scotton, 1993a, 1997, p. 124)

In example (10), the English preposition *before* is a subordinator and assigns a temporal thematic role to the subordinate clause, which is entirely in the ML, Xhosa. The fact that some prepositions can be subordinators in many languages means that speakers’

intentions can directly elect EL CM prepositions to introduce a clause otherwise in the ML, as in (10).

10. . . . *u-ya-yazi* *moss* *u-ba* *kwa-kunjami* **before** *ku-fik-e* *inkululeko?*
 . . . 2S-PRES-know exactly INF-be of-how before LOC-arrive-FV freedom
 ‘ . . . do you know exactly how it was before freedom arrived?’
 (Jake and Myers-Scotton, 2009, p. 220)

Jake and Myers-Scotton (2009, p. 221) point out that prepositions that overtly assign case as well as a thematic role often head EL Islands in CS. Another explanation for the frequency of adjunct PPs in the EL is that they are loosely connected to the predicate-argument structure of a clause and do not interfere with the projection of structure of the ML at the level of the Formulator.

Examples of prepositions that are indirectly elected early SMs or one of the two late SMs, bridges and outsiders, are illustrated later.

Early system morphemes

As noted previously, early SMs are indirectly elected on their route to surface activation (cf. Bock and Levelt, 1994). They depend on their directly elected CM heads for information about their form and meaning. The structure of early SMs varies, but they are called by and occur with their CM heads. Examples of early SMs include affixes marking noun class and plural (e.g., *Swahili* Noun Class 5/6: *tundu/ma-tundu* ‘hole/holes’ or Noun Class 1/2: *m-bwa/wa-bwa* ‘dog/dogs’ in *Swahili*). Other EL early SMs include prepositions elected with their CM verbs. In example (11), both the verb and the early SM preposition come from the EL. The rest of the clause is in the ML, Spanish.

11. *Bueno ¿Por que te hicieron beat up ese?*
 ‘Well, why did they beat you up, man?’
 (Pfaff, 1979, p. 297)

In (12), an English EL verb, inflected with Ewe ML TMA late SMs, occurs with its indirectly elected preposition *out*, an early SM. An English adverb occurs as well. The preposition *out* modifies the thematic roles of the subject and the implied object theme.

12. *e-dze* *be* *wó-á-find out* **first** . . .
 3S-be.proper COMP 3.PL-FUT-find out first . . .
 ‘they have to find out first . . .’
 (Amuzu, 2010 p. 160)

In Xhosa-English, EL verbs with prepositional satellites sometimes occur, as in (13), where an English EL verb is inflected with ML late SMs. The EL early SM preposition *over* is indirectly elected by *take*. The directly elected verb *take*, together with the indirectly elected early SM *over*, means ‘become dominant.’

13. . . . *ama-Xhosa* *si-zo-tak-a* **over** . . .
 . . . CL.6-Xhosa we-FUT-take-FV over . . .
 ‘ . . . the Xhosa, we are going to take over. . .’
 (Jake and Myers-Scotton, 2009, p. 215)

Another example of an English EL verb indirectly electing an early SM preposition is in (14), where another Bantu language, Shona, is the ML. Shona provides all of the requisite late SMs in a ‘do’ verb construction. In this example, plural morphology from the ML and EL occur on *lesson*. In the following section we discuss morpheme doubling, a characteristic of some early SMs, such as plural.

14. . . . *va-no-nok-a* *ku-it-a catch up* *mu-ma-lesson-s*
 . . . 3PL-PRES-be.late-FV INF-do-FV catch up LOC-CL.6-lesson-PL
 ‘ . . . they are late to do catch up in [their] lessons.’
 (Myers-Scotton, 2002, p. 136)

Early SMs have two contrasting features that promote their appearance in contact phenomena, including CS: early availability at the conceptual level and limited morphosyntactic roles in clause structure. Because early SMs do not determine morphosyntactic structure, like nonfinite verbs, they seem to add structure at the argument level. However, their contribution to meaning makes it clear that they are conceptually activated, not elected at the level of the Formulator. In some sense, this makes them free to appear in contact phenomena.

Double morphology: early SMs

One illustration of the availability of early SMs is ‘double morphology’ (Myers-Scotton, 2002, pp. 91–93). Double morphology refers to cases of an EL CM occurring with two early SMs both conveying the same meaning, one from the EL and one from the ML. This example from Swahili-English CS illustrates double plurality: *ma-storie-s* has a Swahili class 6 prefix for plural and an English plural suffix. Morpheme doubling was first discussed in Myers-Scotton (1993a, 1997, pp. 132–135) as a kind of mistiming, but the 4-M model later differentiated early SMs from late SMs. Myers-Scotton (2002, p. 92) proposes the Early System Morpheme hypothesis: ‘Only early system morphemes may double in classic codeswitching.’ They have a special relation with their CMs that would promote doubling. Both early SMs and CMs are conceptually activated; early SMs are salient at the same level as their CM heads. Thus, they are ‘available’ if any mistiming is going to occur. Further, early SMs, whether single or double, do not violate the SMP which covers only one type of late SMs, outsiders.

CS corpora also include examples of double infinitive morphology, as in (15), from Congo Swahili-French CS. In the subordinate clause, the ML shifts to Swahili. In this clause, the EL French infinitive is a CM that has indirectly elected the infinitive *-er* with the verb. It is inflected with the ML infinitive prefix *ku-*.

15. *siku (h)ile* *j’etais* *sur le* *point y-a*
 day CL9.DEM 1S’be.PST at DEF.M point CL.9-ASSOC
ku-renvoy-er mon épouse kwa wa-zzazi wa-ke . . .
 INF-return-INF POSS.1S wife LOC CL.2-parent CL.2-POSS
 ‘That day I was on the point of returning my wife to her parents. . .’
 (Kamwangamalu, 1987, p. 172)

Similarly, in example (16), there is doubling of nonfinite verb morphology: a prefix from the ML Acholi and a suffix from the English both mark the verb in the

second clause as nonfinite. In this example, there is also doubling of the plural: *lu-civilian-s*.

16. *Bene o-nwongo gi-using Swahili ka-terrorizing lu-civilian-s*
 also 3S-find.IMPERF 3PL-using Swahili INF-terrorizing PL-civilian-PL
 ‘One finds they are using Swahili to terrorize the civilians.’
 (Myers-Scotton, 2002, p. 93)

Late system morphemes: bridges and outsiders

In the 4-M model, there are two types of late SMs. They are accessed similarly in production: late SMs only become salient when larger grammatical constituents are assembled at the Formulator. Both types of late SMs participate in structuring the clause, although their roles are different. One type, called a ‘bridge’ late SM, joins together two units to create one. ‘Bridge system morphemes provide hierarchical structure to conceptual structure’ (Jake and Myers-Scotton, 2009, p. 224). Bridges specify syntactic patterns and dominance, but do not determine the thematic roles (as in *the destruction of the army*). In CS, the ML typically provides bridge SMs. In (17), Acholi, the ML, provides an associative *me* and other SMs of the grammatical frame, the locative *-i-* and the copula *tye*.

17. *Costs me education i private institutions tye high . . .*
 ‘Costs of education in private institutions are high. . .’
 (Myers-Scotton, 2005a, p. 9)

Bridges often come from the ML, especially if they coindex the relationship among the parts being combined into a larger structure. EL bridge SM prepositions occur rarely. An exception is *djal* in French-Moroccan Arabic CS (Bentahila and Davies, 1992).

18. *walakin ça dépend de quel degré de connaissance djal la personne . . .*
 ‘but that depends on the degree of knowledge of the person. . .’
 Moroccan Arabic-French (Bentahila and Davies, 1992, p. 450)

Another type of bridge SM connects clauses that do not otherwise receive a thematic role; *that* introduces a sentential complement whose thematic role is not specified by the matrix clause predicate. Bridges may be semantically empty, but can be pragmatically meaningful. For example, in (19), *que* ‘that’ connects two English clauses. In this case, the bridge SM comes from Spanish, the EL. In discussing such examples, Pfaff and others have remarked on ‘the solidarity-marking function of using Spanish function words’ (Pfaff, 1979, p. 314). Myers-Scotton and Jake (2015) discuss CS between clauses in some detail.

19. *It goes without saying I think que along with the picketing we are doing a boycott.*
 (Pfaff, 1979, p. 314)

The main point we make here is that while bridge SMs typically come from the ML, they can come from the EL under certain circumstances, when they are not coindexed with a superordinate structural category. The key feature of bridges is that they allow another constituent

to be integrated as a subordinate structure within a larger structure. An NP can be integrated within another NP, creating a larger, more complex nominal; a clause can be subordinated to a matrix clause with a bridge SM, if this is not already realized by a directly elected subordinator CM assigning a thematic role to the complement clause, as was the case in (10), where an EL CM preposition introduces a subordinate clause.

Outsider late system morphemes

In contrast with bridge SMs, outsider SMs depend on information that is outside of their immediate maximal projection. They are typically features of the hierarchically superordinate structure. Outsider SMs index relations holding across phrase and clause boundaries. The source of such information can be provided by verbs and prepositions and is often indicated with overt case or agreement.

Like bridge SMs, outsiders are elected when directions are sent to the Formulator to assemble larger constituents; these constituents are checked against requirements of morphemes elected at the conceptual level. For example, does a verb or a preposition assign dative or accusative case? Recall (2) from Swiss German-Italian CS, where the German definite determiner also indicates dative case, assigned by the preposition: . . . *mit em maiale* ‘. . . with the pig.’ Subject and/or object agreement is called at the level of the Formulator; information at the superordinate level from conceptually elected morphemes determines their form. Recall the Swahili inflections on the English verb in (1): *a-li-m-buy-i-a* ‘he bought for him.’ Some pronouns are outsider SMs (Jake, 1994). In (12), the Spanish object clitic *te* occurs in the verbal agreement complex of the finite verb, not in argument position: *te hicieron beat up* ‘they beat you up.’ It is provided by the ML, Spanish.

When can EL late outsider SMs occur in a bilingual clause? When they are part of an EL Island. In example (20), a Spanish EL late outsider preposition, the ‘personal’ *a*, which marks human or other direct objects with human-like properties, occurs as part of an EL Island, with an EL borrowing *boss*.

20. *They invite a el boss and then they don't keep their word*
 They invite their boss and then they don't keep their word
 (Moyer, 1992, p. 196)

When Spanish is the ML, the personal *a* typically introduces a mixed constituent, as in (21), where an EL noun occurs with the ML case-marking preposition *a* and an ML possessive pronoun. The EL noun occurs with an EL indirectly elected early SM, plural.

21. *Pero tú te refieres* .
 but 2S.SUBJ 2S.OBJ refer.2S.PRES
a tus coworkers . . .
 but 2S.SUBJ 2S.OBJ refer.2S.PRES
 ‘But are you talking about your coworkers. . . ?’
 (Jake, Myers-Scotton and Gross, 2002, p. 81)

Both bridge and outsider SMs are the elements referred to by the Differential Access Hypothesis, noted earlier. Information required to elect late SMs is not available until the level of the Formulator. Bridges build more complex structures and indicate subordination; outsiders overtly index the nature of the subordination. Moreover, outsiders are the subset of

SMs that must come from the ML, as specified by the SMP in classic CS, in line with the MLF model. Myers-Scotton and Jake (2017) present some empirical evidence that late SMs are not available at the same abstract level as CMs or early SMs based on asymmetry of morpheme type. See Section 4, next.

4. Current contributions and research

Thus far, this chapter has largely shown how the 4-M model provides an explanation of patterns and variation in CS data. In this section we employ the 4-M model to explain other types of contact phenomena, specifically convergence, creole development, and mixed languages.

Variation in morpheme integration depends on how a form is elected; this can vary across languages, but also within a language. As noted in section 3, a preposition heading an adjunct phrase can be a directly elected CM, but the same preposition can be an early SM, indirectly elected by a verb ((*look*) (*up the flagpole*) versus (*look up*) (*the answer*)). The 4-M model aims to clarify how lemmas are activated in building structure and how they are called in language production in two ways, conceptually and structurally. That is, a surface form is not tagged as a member of a class of morphemes; rather the way a morpheme is activated determines its classification in that particular grammatical frame, and a given form may be elected via more than one route, given different linguistic frames.

Patterns in naturally occurring, e.g., conversational, CS data reflect variation in how morphemes are realized and explain certain asymmetries in CS corpora. Here we consider two cases of asymmetry in the distribution of determiners in NPs in Spanish-English CS and in Italian-Swiss German CS. In these data sets, one language more frequently provides the determiner. This asymmetry reflects a difference in the level of abstract structure at which the determiner is elected. Determiners in many languages realize *phi*-features (person, number, and/or gender) as well as definiteness. *Phi*-features may correlate with semantic and pragmatic features as well.

The asymmetry in Spanish-English CS, with Spanish determiners and Spanish EL islands, reflects differences in levels of abstract structure (Myers-Scotton and Jake, 2017). In the corpora of determiner phrases studied, Spanish is always the ML; therefore, it is not surprising that Spanish determiners always dominate in all such phrases. This is the case even when an English noun occurs following a determiner. A secondary explanation is that in Spanish, *phi*-features are available at the abstract level of lexical-conceptual structure. The reason is that such features, as well as definiteness in determiner phrases, must be available for structuring other elements in a clause, notably pro-drop and agreement in verb phrases. Also the realization of subject and object affixes (late SMs) refers back to the determiner phrase. The way early SMs are elected in Spanish explains the high frequency of Spanish determiners, but also the relative absence of English NP EL islands. English determiners and some nominal elements do not become available until larger constituents become salient at the Formulator.

In CS in another language pair, Italian and Swiss German have determiners which are not equally available at the lexical-conceptual level. Determiners in both languages carry *phi*-features, of course, but German determiners have an extra feature, grammatical case, that is salient at the level of the Formulator. When Italian is the ML, either Italian or German nouns occur in full NPs with Italian determiners. Also, full Italian NP EL Islands occur frequently in clauses framed by a German ML. This is not the case for German nouns in clauses framed by Italian; full German EL NP Islands are rare. That is, Italian nouns fully specify the forms of determiners at the lexical-conceptual level. To repeat, while German nouns can occur in phrases fully specified by the ML, Italian, German determiners with German nouns are hardly

found at all in such a corpus (Preziosa-Di Quinzio, 1992). This is because German determiners are not fully specified until the level of the Formulator. Because they must specify grammatical case as well as gender, German determiners, and presumably determiners in other languages with grammatical case, are infrequent in CS except when such languages are the ML. Myers-Scotton (2002, pp. 305–306) refers to such elements as multi-morphemic, with the late SM they include taking precedence. This means the entire element shows distribution patterns as if it were a single late SM. The ‘drag down’ principle applies.

Convergence

Convergence occurs when abstract morphosyntactic structure from one language contributes to the structure of another variety. One consequence of the crossing over of features of conceptually salient morphemes from one variety into another is that shifts in grammatical structure in the framing language (ML) can occur. Lemmas underlying these morphemes provide directions to the Formulator, and this can affect, for example, the valance of a verb or the thematic roles and predicate-argument structures assigned by a verb. This, in turn, may lead to surface structures in one language, in which late (ML) SMs reflect directions from the other (EL) variety. The 4-M model limits which lexical items can have their abstract lemma entries modified in convergence.

The case of Valle d’Aostan French (VDA French) illustrates convergence of French and Italian (as well as VDA Patois) that results in a new variety (Sarullo, 1998). There is overt French-Italian-Patois CS as well; this kind of CS with morphosyntactic convergence is referred to composite CS (Myers-Scotton, 2006, pp. 271–278). Sarullo (1998) examines how different lexical categories show influence from Italian. Some influence is at the lexical-conceptual level, as is the case of a speaker using *faire* ‘to do’ instead of *suivre* ‘to follow’ to describe taking courses as a student. In Standard French *faire* indicates teaching, not taking a course, unlike in Italian, where context disambiguates the meaning.

22. *J’ai fait mes courses pour . . .*
 I’have.PRES.1S do.PP my.PL course.PL for. . .
 ‘I took my courses for. . .’
 (Sarullo, 1998, p. 110)

In other cases, the predicate-argument structure, as well as the lexical-conceptual structure of a French verb, reflects Italian influence, as is the case of *marier* ‘to marry.’ In (23), the speaker uses *marier* as a transitive verb to express a semantically reciprocal relationship. However, in Standard French, *marier* is a transitive performative in which an authority marries people; to express an agentive reciprocal, a passive-like structure with the auxiliary *être* ‘to be,’ a reflexive clitic and a prepositional phrase is employed instead: *je me suis marié avec elle* ‘I got married with/to her’ (Sarullo, 1998, p. 122).

23. . . . *et moi j’ai marié elle*
 . . . and me I’have.PRES.1S marry.PP.M.S her
 ‘. . . and I, I married her.’
 (Sarullo, 1998, p. 121)

In Italian, the verb *sposare* ‘to marry/get married’ can occur with a full object pronoun (or preverbal clitic). While the surface forms of (23) are in French, the abstract structure projected

by an Italian counterpart influences the lexical-conceptual structure and predicate-argument structure of the VDA French verb. This has consequences for the election of other lemmas required to realize the speaker's intentions in (23); the clause in (23) is not a passive-like structure, requiring *être*, but an active transitive structure, electing the auxiliary *avoir* instead. There is convergence involving not only the semantics and pragmatics of a CM, but also early SMs, the auxiliary verbs.

Sarullo (1998) makes the case for similar Italian influence at the abstract level for some early SMs in VDA French. For example, some VDA French nouns reflect the gender of their Italian counterparts. In (24), *couple* occurs with a feminine determiner, like its Italian counterpart *coppia*, instead of the Standard French masculine.

24. . . . *la couple participant* . . .
 ' . . . the participating couple. . . '
 (Sarullo, 1998, p. 152)

Examples from VDA French illustrate how convergence is different from CS: features associated with all morpheme types can change, although they are initially activated conceptually. The abstract reconfiguration of CMs in convergence is clear. In addition, the form of early SMs, such as determiners, reflects the abstract features of the CM electing them; if the gender of a noun changes in convergence, so does the gender of the determiner.

The Flensburg area of Schleswig-Holstein, Germany, is another convergence area. Historical, political, and social factors have created a context setting with areal features from South Jutish and Low German, with more dominant Standard German and Danish. Fredsted (2013) discusses factors that underlie the current linguistic situation and focuses on convergence involving verbs. Some bilinguals in the area exhibit composite CS because there is convergence as well as CS in their speech. One instance of convergence is surface case marking and verb agreement reflecting the overt ML, but whose form is directed by processes from a second language.

The example in (25) illustrates composite CS. There is CS because Danish *linje* 'line' occurs in place of German *Zeile* 'line' in a clause otherwise composed of German. There is convergence because a German experiencer pronoun occurs in nominative case and controls subject-verb agreement, as would occur in Danish. In Standard German, the experiencer is in dative and the theme controls subject-verb agreement.

25. *Ich fehl-e nur noch ein-e linje oder so*
 I lack-1S.PRES only still one-F line or so
 'I lack only just one line or so.'
 Standard German: *Mir fehlt nur noch eine Zeile oder so.*
 South Jutish: *Æ mangle kun en linje eller saan.*
 (Fredsted, 2013, p. 349)

The 4-M model allows for a more precise definition of convergence because it is based on a language production model. That is, convergence occurs when late SM features from one variety are realized in surface forms from another variety. In convergence, the impact of one language on another is limited to the levels of lexical-conceptual structure and predicate-argument structure because the morphological realization patterns of late SMs come from the other language, the ML, German, in the case of (25).

Fredsted (2013) also provides a convergence example illustrating the converse. The surface forms in (26) are entirely in Danish, but the Danish verb *fejle* ‘make mistakes’ occurs as a false cognate with German *fehlen* ‘lack,’ in place of Danish *mangle* ‘lack.’ In Standard Danish, both verbs take nominative subjects and direct objects, but in (26), the experiencer subject is in dative case, illustrating that the convergence situation in the Flensburg area is fluid. Here, there is convergence of Danish to German at the level of predicate-argument structure and lexical-conceptual structure. While this convergence structure involves changes in the abstract features of CMs, it has consequences for late SMs realizing case and agreement.

26. *mig fejl-er tre opgave-r*
 1S.DAT lack/(err)-PL three exercise-PL
 ‘I/(me) lack three exercises.’
 Standard Danish: *Jeg mangler tre opgaver*
 Standard German: *Mir fehlen drei Aufgaben*
 (Fredsted, 2013, p. 349)

Another example of convergence in the Flensburg area involves use of *und* ‘and’ in a purpose infinitive construction. This reflects influence from South Jutish and Low German, in which two constructions have merged. In Standard German, the purpose infinitive requires the prepositional forms *um zu* ‘in order to’/‘around to.’ The infinitive marker *zu* occurs before the verb stem or between a separable verbal prefix and the verb, as in *auf-zu-back-en* ‘up-to-bake-INF’/‘to freshen’ in the Standard German example provided in (27). In informal spoken German in the Flensburg area, some speakers use *und* in place of *um zu*. Although the purpose verb remains an infinitive, as in Standard German, the verb is not in final position, as in Standard German subordinate clauses. Thus, there is convergence toward South Jutish word order as well as merger of coordinator and subordinator for these speakers in their informal German.

27. *Ich bin erst bei und back-en die Brötchen auf*
 I am only by and bake-INF DET.PL.ACC roll.PL up
 ‘I am only by and [to] crisp the rolls up.’ (‘I have just started crisping up the rolls now’)
 Standard German: . . . *um die Brötchen aufzubacken*
 (Fredsted, 2013, p. 343)

While all the surface forms in (27) appear to come from German, South Jutish directions to the Formulator affect the surface realization of word order and the realization of German late SMs. The German conjunction *und* would require all verbal conjuncts to be inflected with similar TMA markers. However, in (27), the second verb is nonfinite, instead of finite, as would be expected if the verb were conjoined with Standard German *und* to the finite verb *bin* ‘am.’ This is an example of convergence where the abstract features of one variety partly determine features of the framing language at the level of morphological realization patterns.

One final example further illustrates how CS can lead to convergence and composite CS. In some varieties of Ecuadorian Quichua, CS with Spanish has consequences for the resulting grammatical structures. Muysken (1997) discusses examples from Media Lengua; Gómez-Rendón (2008) discusses borrowing resulting from Spanish-Ecuadorian Quichua contact. More recently, Lipski (2017) has examined the Imbabura Quichua and Media-Lengua of some bilingual speakers and argues they are distinct varieties for most bilinguals based on psycholinguistic experimental evidence. Example (28) illustrates CS with convergence, or composite CS (Jake and Myers-Scotton, 2009). The most obvious contributions from Spanish in (28) are

CMs: a Spanish verb *suced*er ‘to happen’ and a noun *culebra* ‘snake.’ In addition, a Spanish CM subordinator *porque* ‘because’ occurs. In Quichua, all subordinate clause verbs are nonfinite. However, in (28), the verbs of the complement clause of *porque* and its complement goal clause are finite, as they would be in Spanish.

28. *cai sucedi-ju-shca* [*porque cai culebra* *ri-shca*
 this happen-PROG-PST because this snake go-3PST
 [*chai mai-pi* *yacu tiya-n*] [*causa-ngapac*]]
 that where-LOC water exist-3PRES live-PURP.COREF
 ‘this was happening because the snake went to where there was water in order to live’
 (Jake and Myers-Scotton, 2009, pp. 235–236)

In (28), the TMA markers come from the ML, Imbabura Quichua, but the directions to the Formulator from the conceptually elected subordinator from the EL, Spanish, determine which TMA markers are suffixed to the Quichua verbs *ri-shca* ‘(it) went’ and *tiya-n* ‘(there) is.’ This illustrates composite CS, or CS with convergence. At the surface, late SMs come from the ML, Quichua, but their form is determined by procedures associated with an EL CM, *porque*. Thus, the EL CM is not simply inserted into the ML frame, as it would be in classic CS. Instead, it contributes to, or changes, the grammatical frame of the bilingual clause, although the EL still does not contribute any overt late SMs.

Mixed languages

The most radical outcome of language contact is a mixed language. We suggest that CS is the original mechanism involved in the development of mixed languages, but then convergence also enters the picture. One can identify implicational evidence of an ML and an EL in mixed languages, such as in Copper Island Aleut and Romani contact phenomena. The hypothesis is that the extensive convergence and composite CS leads to an ML turnover, which is then arrested (Myers-Scotton, 1998, 2002). Thus, there are late SMs from both varieties in a mixed system.

In Copper Island Aleut, for example, a favourable shift to greater social status for the Aleut language seems to have occurred: ‘the children may have found them themselves in a situation where they felt the need for a different language as the means of ensuring their identity. Russian, their native language, could no longer serve this purpose’ (Vakhtin, 1998, p. 324). Golovko (1994) implies there are two turnovers, a turnover from Aleut to Russian that is arrested, while CS is the unmarked choice. However, the turnover to Russian as ML is arrested due to a realignment of group identity. Russian verbal forms with late SMs are the evidence of this earlier dominance of Russian. They are the structurally assigned morphemes that have invaded the Aleut frame signalling an initial turnover to Russian (Myers-Scotton, 2003; Golovko, 1994).

Another example of a mixed language where the turnover is essentially complete is Ma’a, a language spoken in Tanzania. Ma’a has some Cushitic lexicon but Bantu (Mbugu) syntax. The hypothesis is that CS with convergence occurred and a Bantu ML structure became increasingly dominant. Some Cushitic vocabulary remains, especially in a variety referred to as Inner Mbugu, or Ma’a (Mous, 2003; Myers-Scotton, 1998, 2003).

Although CS and convergence is part of the ML turnover process, CS need not lead to a turnover of the ML. In most cases, classic CS occurs or CS with some convergence. In composite CS, the ML remains. That is, there is a shift of some morphosyntactic features of the

ML, but not a split of SMs, as in Copper Island Aleut, or a composite grammatical system, partly attested in Ma'a.

In many contact situations with Romani, there are mixed language structures. In Romani dialects, the outsider language is the source of much of the morphosyntax. Adamou and Granqvist (2015) discuss examples of finite inflected verbs from Turkish and Finnish in Romani bilingual clauses. We suggest that this type of Romani language mixing illustrates an early stage of mixed language formation that has not developed into an independent mixed language, owing to changes in the sociopolitical settings.

Pidgin and creole languages

Applying the 4-M model to pidgin and creole languages provides for a more precise explanation of what the lexifier (or superstrate) can contribute. Elsewhere we have discussed the implications of the 4-M model and the construct of the ML for creole development (Myers-Scotton, 2001). The main input to the creole frame is the reanalysis of conceptually activated lexifier morphemes that have the ability to build morphosyntactic structure. We assume creoles develop because speakers in an extreme contact setting lack access to lexifier late SMs because they are not salient at the level of lexical-conceptual structure. Conceptually salient morphemes (i.e., CMs and early SMs) from the lexifier are at least partly available and are reanalyzed to provide the creole with a developing grammatical frame. While substrate varieties cannot provide late SMs, because they are not the putative target language of the creole, the lexifier also cannot contribute late SMs, except as part of idiomatic chunks or periphrastic constructions, or if the lexifier form is also accessible conceptually, as in the case of possessive *vor* in Berbice Dutch (Kouwenberg, 1994), which also occurs as a thematic role-assigning preposition.

Haitian Creole has developed a robust TMA system created largely from reanalyzed lexifier CMs. In (29), the phonologically salient part of an accessible preposition in a French periphrastic (*être après à*) has been reanalyzed as the imperfective in Haitian. The salient semantic and pragmatic features of *après* provide the desired aspect features, imperfective. Late SMs, the copula and the infinitive marker, are not accessible and are blocked from reanalysis.

29. *Mari ap manje krab le*
 Mary IMPERF eat crab DET
 'Mary is eating the crab.'
 (Lefebvre, 1996, p. 240)

Tok Pisin further illustrates how lexifier conceptually salient morphemes are reanalyzed in creole development. In (30), two lexifier CM pronouns are reanalyzed to fit the substrate-based creole frame. *Em* (*him*) remains a CM in subject position; *he* is reanalyzed as a predicate marker *i* (used mostly with third person subjects). In addition, *him* is also reanalyzed as a transitive marker *-im*, occurring on transitive verbs in the same position as transitivizers in Austronesian Tigak (e.g., *vuak-i*, 'break-TZ').

30. *em i bruk-im diwai*
 3.S PM break-TZ stick
 'He broke/breaks the stick.'
 (Jenkins, 2000, p. 134)

Example (31), also from Tok Pisin, shows early SMs are accessible in creole development; in addition to the reanalyzed pronoun *me*, the CM verb *sing* and the indirectly elected early SM proposition *out* realize the creole lexical-conceptual structure. The reanalysis of the verb is clear from the position of the transitive marker *-im*. In addition, another lexifier CM, *finish*, has been analyzed as a TMA perfective marker *pinis*.

31. *Mi sing-aut-im pinis*
 I sing-out-TZ PERF
 'I called (out).'
 (Jenkins, 2000, p. 136)

Examples (32) and (33) from Mauritian illustrate how the semantic and pragmatic features of the creole determine how lexifier conceptually salient morphemes are reanalyzed. In (32), a CM emphatic pronoun plus a determiner-noun (an early SM with a CM) realize a true reflexive pronoun, occurring in argument position and receiving a thematic role: *mo lékorps*. In contrast, in (33), an intransitive verb occurs in place of a pseudo-reflexive construction from the lexifier. The lexifier pronoun clitic, essentially an agreement marker, is a late SM and not accessible in creole development. Examples (32) and (33) illustrate how the creole has developed a *theta*-role receiving reflexive out of early SMs and a CM, but that non-thematic reflexives are not constructed.

32. *mo va touye mo lékorps*
 'I shall kill myself.' (French: *se tuer*)
 (Corne, 1988, p. 88)
33. *mo souvini*
 'I remember.' (French: *se souvenir*)
 (Corne, 1988, p. 78)

Lexifiers of some creoles also have bridge SMs, such as weather *it*. Notably, such lexifier bridge SMs do not occur in creoles. Instead the argument position is typically filled by a pragmatically and semantically relevant CM, or reanalyzed early SM plus CM, as in the Haitian example in (34). Compare the French equivalent, in which a bridge late SM occurs.

34. *lapli ap tonbe*
 rain PROG fall
 'it is raining'
 French: *il pleut*
 (Holm, 1988, p. 88)

Other productive areas of analysis of creole structure from the perspective of the 4-M model include possessive constructions, serial verb constructions, linking subordinators, and TMA systems. Essentially, only conceptually salient morphemes from the lexifier (or other prominent varieties) are available for reanalysis to satisfy an abstract grammatical frame. The prediction is that CM and early SMs from the lexifier can be reanalyzed, but that late SMs from the lexifier cannot. In addition, while CMs or CMs with early SMs from the substrate can cross over as borrowings, late SMs cannot.

5. Summary

This review of the 4-M model and how it applies in the analysis of contact phenomena has shown that the nature of a particular morpheme type depends on two factors. First, both the abstract and structural features of morpheme type are critical. Second, and less studied, the psycholinguistic nature of morpheme types – that is, the abstract level in production at which a morpheme type is elected – is also critical. Recognizing the roles of these two features of morphemes enables us to account for the nature of linguistic data in bilingual communities. This means that, when languages are in contact, the structural nature of code-switching or other contact phenomena that can occur can be predicted.

Our analysis supports the following hypothesis: Only outsider late SMs from the ML (under the 4-M model) control structure in CS. Conceptually activated morphemes from any participating variety can appear in CS and other contact phenomena. What differentiates these morphemes from late SMs is how they are elected in language production. CMs are directly elected at the conceptual level, and many of their affixes or modifiers are indirectly elected at this same level as speaker intentions elect their CM heads. Our hypothesis is that late SMs are elected via a different route. Thus, they only become salient at the level of the Formulator where they participate in building larger constituents.

We have argued that the 4-M model makes predictions for the distributions of morpheme type in other contact phenomena. In some contact phenomena, e.g., convergence and creole development, restructuring of abstract structure of lemma entries occurs, but only conceptually elected morphemes are reanalyzed to realize speaker intentions. In convergence, abstract features of one variety partly determine features of the framing language at the level of morphological realization patterns. In creoles, CMs and early SMs from the lexifier can be reanalyzed to create grammatical structure.

6. Further reading

Amuzu, E. (2010). *Composite codeswitching in West Africa: The case of Ewe-English codeswitching*. Saarbrücken: Lambert Academic Publishing.

This volume makes a case for composite CS. It recognizes classic CS, in which one language is called the ML because it supplies all the critical predicate-argument structure in bilingual data. However, the main argument is that some bilingual data in Ewe-English switching show an ML that is a composite of abstract grammatical structure from more than one source variety. Amuzu's study recognizes that morpheme type is relevant to distinguishing CS with convergence, or composite CS, from classic CS. While entire morphemes elected at the lexical-conceptual level can occur in CS relatively freely, when features of those morphemes direct procedures at the level of the Formulator, there is composite CS.

Fredsted, E. (2013). Multilingualism and longitudinal language contact in the German-Danish border region. *Language Typology and Universals (Sprachtypologie und Universalienforschung)*, *STUF*, 66(4), pp. 331–353.

This article considers the structural similarity of neighbouring languages belonging to two different branches of Germanic languages spoken in the German-Danish border region. They have developed common features not found in other varieties of Danish and German spoken outside of the central area. These features, largely in verb phrases, are the result of contact and/or language shift(s). Fredsted's study illustrates that the abstract features of one variety can influence the surface forms of another variety, but that the actual late SMs are still provided by only one of the participating varieties.

Myers-Scotton, C. M. (2003). What lies beneath: split (mixed) languages as contact phenomena. In: Y. Matras and P. Bakker, eds., *The mixed language debate, theoretical and empirical advances*, 1st ed. Berlin: De Gruyter Mouton, pp. 73–106.

This chapter presents an argument that there are very few true split (mixed) languages. Such languages would have to show ‘systematic evidence that some of the abstract grammatical structure . . . comes from a source other than the major source of its lexicon’ (Myers-Scotton, 2003, p. 73). This abstract structure refers to what are called late outsider SMs in the 4-M model. These are the elements that co-index the main CMs in a clause (including subjects and objects), or those which mark grammatical case. Another way of looking at split (mixed) languages is that ‘split languages show a composite structure that goes beyond a composite at the level of lexical-conceptual structure’ (Myers-Scotton, 2003, p. 91). Copper Island Aleut is an example of such a language. The author claims that there are very few languages meeting these criteria. Instead, there are a number of languages in bilingual situations that show convergence. Myers-Scotton cites examples from Jake suggesting that Chaupi Lengua/Media Lengua spoken in Ecuadorian Quichua areas is moving towards a split structure because of a reduced distribution of some discourse affixes and grammatical morphemes. Chaupi Lengua/Media Lengua in an extreme case of composite CS, and not a mixed language, because all of the overt late SMs come from Quichua in mixed constituents in bilingual clauses.

Myers-Scotton, C. M. (2013). Paying attention to morpheme types: making borrowability more precise. In: de Féral, C., ed., *In and out of Africa, languages in question*, 1st ed. Louvain-la-Neuve and Walpole, MA: Peeters Publisher, pp. 31–42.

Myers-Scotton argues that grammatical morphemes as candidates for borrowing depends on both their surface forms and their abstract nature. This contribution considers the abstract structure of grammatical elements. Following the 4-M model of morpheme classification (Myers-Scotton and Jake, 2009, and elsewhere) this article differentiates conceptually activated morphemes from structurally assigned morphemes and argues that only morphemes that are elected at the abstract level of lexical-conceptual structure are easily borrowed. Myers-Scotton hypothesizes morphemes that are not salient until the level of the Formulator cannot be borrowed. These are called late SMs under the 4-M model; they co-index the main elements carrying semantics and, in so doing, make transparent relationships in the clause. Late SMs include morphemes that mark grammatical relations in verb phrases and grammatical case marking. Such morphemes are not only hard to borrow, but are rarely borrowed and generally only in exceptional contact conditions.

Wei, L. (2015). *Interlanguage, the abstract level of language acquisition*. Lewiston, NY and Lampeter, Wales: Edwin Mellen Press.

This volume goes beyond the description of Interlanguage (IL) to focus on several aspects of the abstract nature of IL. The author argues that the observable developmental sequences are not simply the result of cross-linguistic differences or language transfer, but arise from differential election of morphemes at an abstract level that determines morpheme acquisition, showing that late SMs are least accurate in IL. Wei shows that because languages in contact do not play equal roles in acquisition and because not all morphemes are equally accessible to learners, ‘the sources of . . . abstract linguistic structure that can contribute to IL development are predictably constrained’ (Wei, 2015, p. 251). One important proposal in this volume is the analysis of L2 learner errors in terms of cross-linguistic variation in lemmas underlying lexical entries in the bilingual or multilingual mental lexicon. Wei redefines IL transfer as lemma transfer. Another important proposal is that IL should be understood as a composite developing linguistic system, and it is this composite system which constrains IL performance and governs the IL developmental processes. Essentially, late SMs, especially late outsider SMs, are acquired last in second language acquisition.

7. Related topics

Bilingual language acquisition, cognitive factors, convergence, creoles, mixed languages

Abbreviations

| | |
|---------|---|
| 1 | first person or Bantu noun class for people |
| 2 | second person or Bantu noun class (plural) |
| 3 | third person or Bantu noun class (singular) |
| 6 | Bantu noun class (plural) |
| 8 | Bantu noun class (plural) |
| 9 | Bantu noun class (singular) |
| ACC | accusative |
| APPL | applied verbal extension |
| ASSOC | associative |
| CL | Bantu noun class |
| CM | content morpheme |
| COMP | complementizer |
| CONSEC | consecutive verb tense |
| COREF | coreference |
| CS | code-switching |
| DAT | dative |
| DEF | definite |
| DEM | demonstrative |
| DET | determiner |
| EL | Embedded Language |
| F | feminine |
| FUT | future |
| FV | final vowel (aspect marker) |
| IL | Interlanguage |
| IMP | imperative |
| IMPERF | imperfective |
| INDEF | indefinite determiner |
| INF | infinitive |
| LOC | locative |
| M | masculine |
| MANNER | manner |
| ML | Matrix Language |
| MLF | Matrix Language Frame |
| MOP | Morpheme Order Principle |
| NON.PST | non-past tense |
| NP | noun phrase |
| OBJ | object |
| PERF | perfective |
| PL | plural |
| PM | predicate marker |
| POSS | possessive |
| PP | past participle |
| PRES | present tense |
| PROG | progressive |
| PST | past |
| PURP | purpose |

| | |
|----------|-----------------------------|
| S | singular |
| SM | system morpheme |
| SMP | System Morpheme Principle |
| SUBJ | subject |
| SUBJUBCT | subjunctive |
| TMA | Tense, Mood, Aspect |
| TZ | transitivizer |
| USP | Uniform Structure Principal |
| VDA | Valle d'Aostan |

References

- Adamou, E. and Granqvist, K. (2015). Unevenly mixed Romani languages. *International Journal of Bilingualism*, 19(5), pp. 525–547.
- Amuzu, E. (2010). *Composite codeswitching in West Africa: The case of Ewe-English codeswitching*. Saarbrücken: Lambert Academic Publishing.
- Bacusk, A. (1996). *Two in one, bilingual speech of Turkish immigrants in the Netherlands*. Tilburg: Tilburg University Press.
- Bakker, P. (1997). *A language of our own: The genesis of Michif, the mixed Cree-French language of the Canadian Métis*. New York: Oxford University Press.
- Bentahila, A. and Davies, E. (1992). Codeswitching and language dominance. In: R. Harris, ed., *Cognitive processing in bilinguals*, 1st ed. Amsterdam: Elsevier, pp. 43–458.
- Bock, K. and Levelt, W. (1994). Language production: Grammatical encoding. In: M. Gernsbacher, ed., *Handbook of psycholinguistics*, 1st ed. New York: Academic Press, pp. 945–984.
- Bolonyai, A. (2002). Case systems in conflict: Syntactic and lexical case in bilingual child language. *The Southwest Journal of Linguistics*, 21(2), pp. 1–35.
- Corne, C. (1988). Mauritian creole reflexives. *Journal of Pidgin and Creole Languages*, 3(1), pp. 69–94.
- De Rooij, V. (1996). *Cohesion through contrast: Discourse structure in Shaba Swahili/French conversations*. Amsterdam: IFOTT.
- Fredsted, E. (2013). Multilingualism and longitudinal language contact in the German-Danish border region. *Language Typology and Universals (Sprachtypologie und Universalienforschung)*, STUF, 66(4), pp. 331–353.
- Golovko, E. (1994). Mednyj Aleut or Copper Island Aleut: An Aleut-Russian mixed language. In: P. Bakker and M. Mous, eds., *Mixed languages: 15 cases in language intertwining*, 1st ed. Amsterdam: IFOTT, pp. 113–121.
- Gómez-Rendón, J. (2008). *Typological and social constraints on language contact: Amerindian languages in contact with Spanish*. Utrecht: LOT.
- Hebblethwaite, B. (2007). *Intrasentential code-switching among Miami Haitian Creole-English bilinguals*. PhD. Bloomington, IN, Indiana University.
- Holm, J. (1988). *Pidgins and creoles*, vol. I. Cambridge: Cambridge University Press.
- Jake, J. (1994). Intrasentential code switching and pronouns: On the categorial status of functional elements. *Linguistics*, 32(2), pp. 271–298.
- Jake, J. and Myers-Scotton, C. (2009). Which language? Participation potentials across lexical categories in codeswitching. In: L. Isurin, D. Winford, and K. de Bot, eds., *Multidisciplinary approaches to code switching*, 1st ed. Amsterdam: John Benjamins, pp. 207–242.
- Jake, J., Myers-Scotton, C. and Gross, S. (2002). Making a minimalist approach to codeswitching work: Adding the matrix language. *Bilingualism: Language and Cognition*, 5(1), pp. 69–91.
- Jenkins, S. (2000). *Linguistic contact and composite structures in New Ireland Papua New Guinea*. PhD. University of South Carolina, Columbia.

- Kamwangamalu, N. (1987). French/vernacular code-mixing in Zaire: Implications for syntactic constraints. In: B. Need and A. Bosch, eds., *Papers from the 23rd annual meeting of the Chicago linguistic society*, 1st ed. Chicago, IL: Chicago Linguistic Society, pp. 166–180.
- Kouwenberg, S. (1994). *A grammar of Berbice Dutch Creole*. Berlin: Walter de Gruyter.
- Lefebvre, C. (1996). The tense, mood, and aspect system of Haitian Creole and the problem of transmission of grammar in creole genesis. *Journal of Pidgin and Creole Languages*, 11(2), pp. 231–312.
- Levelt, W. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
- Lipski, J.M. (2017). Ecuadoran Media Lengua: More than a ‘half’-language? *International Journal of American Linguistics*, 83(2), pp. 233–262.
- Matras, Y. (2009). *Language contact*. Cambridge: Cambridge University Press.
- Mous, M. (2003). *The making of a mixed language: The case of Ma’a/Mbugu*. Amsterdam: John Benjamins.
- Moyer, M. (1992). *Analysis of code-switching in Gibraltar*. PhD. Universitat Autònoma de Barcelona.
- Muysken, P. (1997). Media Lengua. In: S. Thomason, ed., *Contact languages: A wider perspective*, 1st ed. Amsterdam: John Benjamins, pp. 365–426.
- Myers-Scotton, C. (1993a, 1997). *Duelling languages, grammatical structure in codeswitching*. Oxford: Oxford University Press.
- Myers-Scotton, C. (1993b). *Social motivations for codeswitching: Evidence from Africa*. Oxford: Clarendon Press.
- Myers-Scotton, C. (1998). A way to dusty death: The matrix language turnover hypothesis. In: L. Grenoble and L. Whaley, eds., *Endangered languages: Language loss and community response*, 1st ed. Cambridge: Cambridge University Press, pp. 285–316.
- Myers-Scotton, C. (2001). Implications of abstract grammatical structure: Two targets in creole formation. *Journal of Pidgin and Creole Languages*, 16(2), pp. 217–273.
- Myers-Scotton, C. (2002). *Contact linguistics, bilingual encounters and grammatical outcomes*. Oxford: Oxford University Press.
- Myers-Scotton, C. (2003). What lies beneath: Split (mixed) languages as contact phenomena. In: Y. Matras and P. Bakker, eds., *The mixed language debate, theoretical and empirical advances*, 1st ed. Berlin: De Gruyter Mouton, pp. 73–106.
- Myers-Scotton, C. (2005a). Embedded language elements in Acholi-English codeswitching: What’s going on? *Language Matters*, 36(1), pp. 3–18.
- Myers-Scotton, C. (2005b). Supporting a differential access hypothesis: Code switching and other contact data. In: J. Kroll, and A. De Groot, eds., *Handbook of bilingualism, psycholinguistic approaches*, 1st ed. New York: Oxford University Press, pp. 326–348.
- Myers-Scotton, C. (2006). *Multiple voices, an introduction to bilingualism*. Malden, MA and Oxford: Blackwell.
- Myers-Scotton, C. (2013). Paying attention to morpheme types: Making borrowability more precise. In: C. de Féral, ed., *In and out of Africa, languages in question*. Louvain-la-Neuve and Walpole, MA: Peeters Publisher, pp. 31–42.
- Myers-Scotton, C. and Jake, J. (1995). Matching lemmas in a bilingual language competence and production model: Evidence from intrasentential code switching. *Linguistics*, 33(5), pp. 981–1024.
- Myers-Scotton, C. and Jake, J. (2000). Four types of morpheme: Evidence from aphasia, code switching, and second-language acquisition. *Linguistics*, 38(6), pp. 1053–1100.
- Myers-Scotton, C. and Jake, J. (2009). A universal model of code-switching and bilingual processing and production. In: B. Bullock and A. Toribio, eds., *Linguistic code-switching*. Cambridge: Cambridge University Press, pp. 336–357.
- Myers-Scotton, C. and Jake, J. (2014). Non-finite verbs and negotiating bilingualism in codeswitching: Implications for a language production model. *Bilingualism, Language, and Cognition*, 17(3), pp. 511–525.
- Myers-Scotton, C. and Jake, J. (2015). Cross-linguistic asymmetries in code-switching patterns: Implications for bilingual language production. In: J. Schwieter, ed., *The Cambridge handbook of bilingual processing*, 1st ed. Cambridge: Cambridge University Press, pp. 416–418.

- Myers-Scotton, C. and Jake, J. (2017). Revisiting the 4-M model: Codeswitching and morpheme election at the abstract level. *International Journal of Bilingualism*, 21(3), pp. 340–366.
- Preziosa-Di Quinzio, I. (1992). *Teoreticamente la Firma la Indietro: Frammestone di Italiano e Schwyzertötsch nella Conversazione Onedi Figli di Emigrati*. MA. University of Zurich.
- Pfaff, C. (1979). Constraints on language mixing: Intrasentential code-switching and borrowing in Spanish/English. *Language*, 55(2), pp. 291–318.
- Sakel, J. (2004). *A grammar of Mosestén*. Berlin: De Gruyter Mouton.
- Sarullo, P. (1998). *The linguistic situation in Valle d'Aosta: A study on the function and on the structure of codeswitching and convergence between Italian and French*. PhD. University of South Carolina, Columbia.
- Vakhtin, N. (1998). Copper Island Aleut: A case of language 'resurrection'. In: L. Grenoble and L. Whaley, eds., *Endangered languages: Language loss and community response*, 1st ed. Cambridge: Cambridge University Press, pp. 317–327.
- Van Hell, J.G., Litcofsky, K.A. and Ting, C.Y. (2015). Intra-sentential code-switching: Cognitive and neural approaches. In: J. Schwieter, ed., *The Cambridge handbook of bilingual processing*, 1st ed. Cambridge: Cambridge University Press, pp. 459–482.
- Wei, L. (2015). *Interlanguage, the abstract level of language acquisition*. Lewiston, NY and Lampeter, Wales: Edwin Mellen Press.
- Ziamari, K. (2008). *Le code switching au Maroc: l'arabe marocain au contact du français*. Paris: L'Harmattan.

Theoretical approaches to the grammar of code-switching¹

Jeff MacSwan

1. Introduction

Code-switching (CS) is a speech style in which bilinguals move between languages in a single social context, as illustrated in (1).

1. Mi hermano bought some ice cream.
 ‘My brother bought some ice cream.’

The term first appeared in print in Vogt’s (1954) review of Weinreich’s (1953) *Languages in Contact*. However, Einar Haugen claimed to have coined it even earlier at the 1962 Annual Round Table Meeting on Linguistics and Languages at Georgetown University. In these earliest examples, CS is presented as one of several language contact phenomena, leading to the emergence of a focused CS research literature in the late 1960s and early 1970s (Benson, 2001).

Scholarly efforts to connect with issues of social significance played a role in elevating CS as a field of research, as social and educational policy began to keenly focus on marginalized and economically disadvantaged groups (Riegelhaupt, 2000). A conventional perspective at the time on economic disparities was that poor families suffered due to inherent cultural, intellectual, and linguistic limitations (Raz, 2013), and bilingualism was among the many suspected underlying causes of poverty. Much as Labov (1970), Wolfram (1969), Fasold (1972), and others had shown through painstaking analysis that the stigmatized language of urban African-American communities was just as rich and complex as the language of the privileged classes, CS researchers showed that language mixing reflected deep linguistic competence, not confusion (Riegelhaupt, 2000; Lipski, 2014). An extensive body of research has now shown conclusively that bilinguals are exquisitely sensitive to the tacit rules which govern the interaction of their language systems in CS itself (for summaries, see MacSwan, 2013, 2014, 2016a; Bhatia and Ritchie, 2013).

Despite this consensus, researchers continued to debate about the nature of the underlying rule system. Early CS researchers constructed theories about the underlying system that

posited CS-specific constraints, echoing contemporaneous research in general syntactic theory; however, later researchers committed to engaging the study of CS in the same way as one would approach linguistic theory generally, with no CS-specific constraints permitted. These different points of view reflect a paradigmatic shift in CS research.

2. Historical overview

The constraint-based program

Among the earliest to observe that there are grammatical restrictions on language mixing were Gumperz and his colleagues (Gumperz, 1967, 1970; Gumperz and Hernández-Chávez, 1970), Hasselmo (1972), Timm (1975), and Wentz (1977). Timm listed several restrictions on Spanish-English CS, including, for example, a switch between a subject pronoun and the verb, as in (2), contrasting in grammaticality with a switch involving a lexical subject, as in (1).

2. *El *bought some ice cream.*
 'He bought some ice cream.'

Grammaticality effects such as these provided evidence that CS is rule-governed.

CS research conducted in the *Aspects* area (Chomsky, 1965) soon began offering 'constraints' like those developed in the contemporaneous syntactic literature to explain these effects. Chomsky had noted that the transformational component in a hybrid generative-transformational system had the disadvantage of vastly increasing the expressive power of the grammar, permitting the formulation of grammatical processes which did not occur in any language. Chomsky (1964, 1965) and other researchers posited *constraints* on transformations which restricted their application to phrase markers. Understood to be part of the grammar itself, this *theoretical* sense of 'constraint' contrasted with the common *descriptive* sense, which only implied that a grammatical pattern had been observed.

The idea of a *constraint* in the theoretical sense appealed to a number of CS researchers, and was used to articulate grammatical restrictions on CS. For example, Joshi (1985a) proposed the CS-specific constraint in (4).

4. Constraint on Closed-Class Items

Closed-class items (e.g., determiners, quantifiers, prepositions, possessives, Aux, Tense, helping verbs) cannot be switched.

Like several similar mechanisms, (4) makes explicit reference to (*code*)*switching*, which of course denotes a change from one language to another.

Reference to a *language* (explicitly or implicitly) in an actual grammatical rule like this is problematic. A language is a set of expressions defined by a grammar and cannot be a term in the grammar itself. That is, a grammar *G* defines a set of expressions *L*. We cannot insert *L* as part of the function *G*, as *L* is itself the output of *G*. Furthermore, constraints so formulated may serve to provide good linguistic description, but they do not provide explanations. Rather than explaining descriptive restrictions observed in CS data, CS-specific mechanisms simply note these restrictions within the grammar itself, and one is left still wondering what general principles of grammar might underlie the descriptions. We might define a *CS-specific constraint*, then, as a proposed grammatical mechanism which makes explicit reference to

(code)switching or language(s), and which is understood to be part of the actual linguistic competence of a speaker. Historically, CS researchers have consistently offered up CS-specific constraints, despite a clear and persistent intuition among many that a better theory of CS would do without them.

Pfaff (1979), for instance, appears to have been the first to consider the question of whether some mechanism external to either grammar is needed in our account of the facts of CS, concluding that no such device should be required: ‘It is unnecessary to posit a third grammar to account for the utterances in which the languages are mixed’ (p. 314). Echoing Pfaff, Woolford (1983, p. 522) similarly wrote, ‘There is no need to propose any sort of third, separate code-switching grammar.’ Lipski (1985, pp. 83–84) similarly observed that ‘preference must initially be given to modifications of existing grammars. . . , rather than to the formulation of a special bilingual generative mechanism.’

Other researchers expressed a similar discomfort with the idea that CS was governed by one or more CS-specific constraints continuing into the 1980s and 1990s, and sought to address the concern in their work. For example, Di Sciullo, Muysken and Singh’s (1986, p. 7) influential work on the Government Constraint postulated that CS ‘can be seen as a rather ordinary case of language use, requiring no specific stipulation.’ Clyne (1987, p. 279) similarly conjectured that CS is ‘governed by the kinds of structural constraints applying to monolingual performance.’ Belazi, Rubin and Toribio (1994, p. 234) proposed their Functional Head Constraint within the context of a view of CS as ‘constrained solely by Universal Grammar.’ Mahootian (1993), like Pfaff (1979) and Woolford (1983), argued against the ‘third grammar’ approach, claiming in Santorini and Mahootian (1995, p. 4) that ‘codeswitching sequences are governed by exactly the same principles of phrase structure as monolingual sequences.’

The demise of the constraint-based program

Despite this long-standing intuition among CS researchers that language mixing is not constrained by mediating mechanisms or a ‘third grammar,’ technological limitations available at the time made a *constraint-free (CF) approach* – one which eschews CS-specific mechanisms in accounting for bilingual data – difficult or impossible to implement. While a few examples of explicit endorsements of CS-specific constraints may be unearthed (e.g., Joshi, 1985a; Sankoff, 1998), the overwhelming perspective in the field has been that such mechanisms ought to be viewed with disdain.

CS researchers tended to take one of three courses in light of this predicament: (1) explicitly confront the limitations of the formal mechanism, and reluctantly but explicitly introduce CS-specific devices (e.g., Sankoff and Poplack, 1981); (2) leave the analytic framework inexplicit or inadequately developed so that the issue did not arise (e.g., Woolford, 1983; Mahootian, 1993); or (3) propose explicit CS-specific mechanisms, and argue that they are vacuously available in monolingual contexts too (e.g., Di Sciullo, Muysken and Singh, 1986; Belazi, Rubin and Toribio, 1994; Myers-Scotton, 1993).

Explicitly confronting the formal limitations

One of the most important early contributions to CS was Sankoff and Poplack’s (1981) formal implementation of the Equivalence Constraint. Several researchers had converged on the notion that CS is controlled by a syntactic equivalence condition (Lipski, 1978; Pfaff, 1979;

Poplack, 1978, 1981). Poplack (1981) proposed two complementary constraints which are among the best known:

5. *The Equivalence Constraint*

Codes will tend to be switched at points where the surface structures of the languages map onto each other.

6. *The Free Morpheme Constraint*

A switch may occur at any point in the discourse at which it is possible to make a surface constituent cut and still retain a free morpheme.

For Poplack, (5) stipulates that CS is allowed within constituents so long as the word order requirements of both languages are met at surface structure; it predicts that the switch in (7) is disallowed, because the surface word order of English and Spanish differ with respect to object pronoun (clitic) placement. The constraint in (6) defines a restriction on affixation in CS; it disallows (8), where an English stem is used with a Spanish bound morpheme without phonological integration.

7. *told *le*, *le* told, him *dije*, dije *him*
 told *to-him*, to-him *I-told*, him *I-told*, I-told *him*
 ‘(I) told him’

(Poplack, 1981, p. 176)

8. **estoy* eat-*iendo*
 I-am eat-ing

(Poplack, 1980, p. 586)

Research since Poplack’s initial proposals has found persuasive documentation that her Equivalence Constraint does not hold up to empirical tests (Stenson, 1990; Lee, 1991; Myers-Scotton, 1993; Mahootian, 1993; MacSwan, 1999; Chan, 1999; Muysken, 2000). Note, for example, the contrast in (9) (Belazi, Rubin and Toribio, 1994).

- 9a. The students *habían visto la película italiana*
 The students had seen the Italian movie

- 9b. *The student had *visto la película italiana*
 The student had seen the Italian movie

The word order requirements of Spanish and English are alike for these constructions, yet a switch between the auxiliary and the verb renders the sentence ill-formed in (9b) even though (5) predicts both examples to be well-formed.

Taken as a descriptive generalization, Poplack’s Free Morpheme Constraint, or observation of a ban on word-internal CS, has been attested across a wide range of language pairs (Bentahila and Davis, 1983; Berk-Seligson, 1986; Clyne, 1987; MacSwan, 1999). But it has also been somewhat controversial, with some CS scholars noting counter-examples (Bokamba, 1989; Myers-Scotton, 1993; Nartey, 1982; Chan, 1999; Jake, Myers-Scotton and Gross, 2002). However, in presenting counter-examples, researchers have often given too little attention to the specific syntactic and phonological characteristics of the examples cited, making it difficult

to determine whether they are in fact instances of word-internal CS or cases of (nonce) borrowing, involving phonological integration (Poplack, Wheeler and Westwood, 1989; Sankoff, Poplack and Vanniarajan, 1990; Poplack and Meechan, 1998; MacSwan, 2004). A reasonable consensus perspective in the field regarding the basic descriptive facts holds that word-internal CS is very rare (MacSwan, 2005a, 2005b; López, Alexiadou and Veenstra, 2017). We return later to this topic as a current area of debate in the field.

Although Sankoff and Poplack (1981) expressed a strong preference for avoiding CS-specific mechanisms to mediate between the two languages in contact, they nonetheless concluded that such a mechanism is necessary on empirical grounds. Otherwise, the authors argued, the free union of Spanish and English phrase structure grammars would yield ill-formed results. For instance, English places adjectives before nouns ($NP \rightarrow Det Adj N$), but Spanish puts them after ($NP \rightarrow Det N Adj$). If a speaker is free to select the Spanish rule and lexically insert English words, an ill-formed construction may result – for example, a Spanish noun and English adjective (**the casa white*), a licit code-switch, or even English lexical items for all categories (**the house white*). To prevent this, Sankoff and Poplack introduced a superscripting mechanism (sometimes called a *bilingual tag* or *language tag*) which restricted lexical insertion rules so that the grammar contributing the phrase structure rule would also be the grammar from which lexical insertion rules were drawn. Hence, under conditions of CS, the Spanish phrase structure rule would be annotated as in (10), generating (11). The superscripting conventions followed from *heritability conditions*, according to the authors, which essentially allowed phrase structure rules to look ahead and restrict the application of lexical insertion rules.

10. $NP \rightarrow Det N^{sp:n} Adj^{sp:adj}$

11. the casa blanca

Sankoff and Poplack do not make explicit the mechanisms for superscript insertion; rather, they indicate that phrase structure rules are so superscripted when they are selected in the generation of a CS utterance and are subsequently used to trigger language-specific lexical insertion rules. An important contribution of their work, however, was to probe the question of constraints on CS empirically by making their theoretical assumptions about such constraints explicit.

Leaving the analysis insufficiently explicit

Woolford (1983) emphasized that a theory of CS should avoid CS-specific mechanism. Like Sankoff and Poplack, she recognized that the rules of lexical insertion must be trained on their language-specific phrase structure rules. In her proposed system,

Phrase structure rules are drawn freely from both grammars during the construction of constituent structure trees, but the lexicon of each grammar is limited to filling only those terminal nodes created by phrase structure rules drawn from the same language. Nevertheless, in the event that there are phrase structure rules common to both languages, such rules belong simultaneously to both languages. Lexical items can be freely drawn from either language to fill terminal nodes created by phrase structure rules common to both languages.
(1983, p. 535)

Woodford's system implies that terminal nodes of a language-unique phrase structure rule (for instance, $NP \rightarrow Det N Adj$ for Spanish) could only be lexically filled by items from the

same language (predicting *the casa blanca* to be ill-formed for Spanish-English, contrary to the facts). In addition, Woolford does not present the formal mechanism that might be responsible for achieving these results; there is no explanation as to how the unique phrase structure rules get linked to language-specific lexical insertion rules. By contrast, Sankoff and Poplack's (1981) similar work, conducted just prior to Woolford's, made these issues and their limitations explicit.

Mahootian (1993) proposed what she termed 'the Null Theory' of CS, formulated within the framework of Tree Adjoining Grammars (TAG) originally introduced by Joshi (1985b) for applications in computational linguistics and natural language processing. TAG differs from mainstream generative grammar in that the lexical items encode partial tree structures and use operations of substitution and adjunction to assemble larger trees composed of multiple lexical items. For example, the verb *build* is represented in the lexicon along with its projection, and therefore the branching direction of its complement is lexically specified. A substitution operation allows a DP *a house* to integrate with *build* by substituting the DP (along with its category label) with the object category label of *build*, generating *build a house* along with its projecting tree structure.

Mahootian focused on the complement relation in phrase structure (see also Pandit, 1990 and Nishimura, 1997), and postulated that (12) accounts for CS phenomenon.

12. The language of a head determines the phrase structure position of its complements in code-switching just as in monolingual contexts.

Mahootian (1993, p. 152) used a corpus of Farsi-English CS data. In Farsi, objects occur before the verb (OV), contrasting with basic word order in English (VO). She observed that in CS contexts the language of the verb determines the placement of the object, as (13) illustrates.

13. Tell them you'll buy *xune- ye jaedid* when you sell your own house.

house POSS new

'Tell them you '11 buy a new house when you sell your own house.'

Mahootian argued that the TAG formalism provides an advantage for the analysis of CS data; because structures are encoded in the lexicon, no intervening control mechanism is needed to align lexical insertion rules with terminal nodes.

Though intended as a general theory of CS, Mahootian's proposal really concerns the role of phrase structure in CS. The analysis was restricted to head-complement configurations. Not only was (12) too narrow in this regard, failing to comment on CS in other domains of syntax, but it also proved to be insufficiently restrictive with respect to head-complement configurations. Note, for instance, the examples in (1) and (2); although all complements are in the correct positions assigned by heads, (1) is well-formed but (2) is not. Furthermore, note that in (9), *visto*, the complement of *habían/had*, is in the position assigned by its head, and therefore adheres to (12), yet (9a) is well-formed and (9b) is not. In (14) (MacSwan, 1999, p. 119), Spanish and Nahuatl word order is respected with regard to the placement of the verbal complement of negation, yet (14a) is ill-formed but (14b) is not.

- 14a. *No *nitekítitoc*
 no ni-tekítitoc
 not 1S-work-DUR
 'I'm not working'

- 14b. Amo *estoy trabajando*
 amo *estoy trabaja-ndo*
 not *be/3Ss work-DUR*
 ‘I’m not working’

One might argue that these are not complement relations, but rather represent extractions with landing sites to the left of their apparent complements. Still, this too makes the basic point: Much more affects grammaticality in CS than the head-complement relation. Like others before her, however, Mahootian’s basic intuition was right: A proper theory of CS should entertain no CS-specific constraints.

Bilingual constraints vacuously available to monolinguals

Di Sciullo, Muysken and Singh (1986) proposed the Government Constraint, which posited an anti-government condition on CS. Based on (15), a standard definition of government, they posed (16) as a condition on lexical insertion (where q indexes a category to a language-particular lexicon).

15. X governs Y if the first node dominating X also dominates Y, where X is a major category N, V, A, P and no maximal boundary intervenes between X and Y.
16. If X governs Y, ... X_q ... Y_q ...

Di Sciullo, Muysken, and Singh’s intuition was that (16) is an instance of (17), which they viewed as a common assumption in syntactic theory which is never made explicit.

17. All elements inserted into the phase structure tree of a sentence must be drawn from the same lexicon.

Based on these assumptions, CS ‘can be seen as a rather ordinary case of language use, requiring no specific stipulation’ (1986, p. 7). In order to permit the head carrying the language index q to percolate up to its maximal projection, Di Sciullo, Muysken and Singh formalized a condition on CS in the form of the Government Constraint in (18).

18. *Government Constraint*
- (a) If L_q carrier has index q , then Y^{\max}_q .
 - (b) In a maximal projection Y^{\max} , the L_q carrier is the lexical element that asymmetrically c-commands the other lexical elements or terminal phrase nodes dominated by Y^{\max} .

This formalism allows the language of a head to determine the syntax of its maximal projection and imposes the condition that two categories must be in the same language if the government relation holds between them.

Much like Sankoff and Poplack’s (1981) formalism, (18) is intended to trigger language-specific lexical insertion by identifying nodes within a phrase marker with a specific language label (called a *language index* here). Although the authors maintain that the mechanism underlying the language index is vacuously available to monolinguals too, it nonetheless appears to add few advantages over Sankoff and Poplack’s version. Similarly, as in Woolford (1983), we are not told how the grammatical system identifies the subset of lexical items comprising each lexicon.

Importantly, there are also well-known counter-examples to (18). For instance, because government holds between a verb and its object and between a preposition and its object, (18) predicts that a verb or preposition must be in the language of its complement. This is shown to be incorrect by examples in (19), where switches occur in case-marked (hence governed) positions.

- 19a. This morning *mi hermano y yo fuimos a comprar* some milk
 ‘This morning *my brother and I went to buy* some milk.’
 (Belazi, Rubin and Toribio, 1994, p. 222)

- 19b. Mi hermana *kitlasojtla in Juan* (Spanish-Nahuatl)
 mi hermana 0-ki-tlasojtla in Juan
 my sister 3S-3Os-love IN Juan
 ‘My sister loves Juan.’
 (MacSwan, 2013, p. 323)

See Halmari (1997) for further discussion of the Government Constraint.

Belazi, Rubin and Toribio (1994) proposed the Functional Head Constraint (FHC) drawing on Abney’s (1987) work on functional heads. According to Abney (1987), functional heads were responsible for selecting complements with specific feature matrices. For example, *for* is a C⁰ with a feature specification requiring its complement to be [-Tense]. Belazi, Rubin and Toribio developed the FHC, given in (20), as an intended refinement of Abney’s (1987) theory, and viewed it as vacuously available to monolinguals and bilinguals both.

20. The Functional Head Constraint

The language feature of the complement f-selected by a functional head, like all other relevant features, must match the corresponding feature of that functional head.

By *language feature*, the authors had in mind a label identifying the language from which an item was contributed, such as [+Spanish] or [+English]. If the features do not agree (a Spanish functional head with an English complement, for example), then the code-switch is blocked. Since (20) applies only to complements selected by functional heads, switches involving lexical heads are not constrained.

Mahootian (1993) and Muysken (2000) argued that the FHC was a further elaboration of the Government Constraint, in that it identified an independently motivated principle of grammar but incorporated language-specific identifiers (for the government constraint, a language index; for the FHC, a language feature). Belazi, Rubin, and Toribio, like Di Sciullo, Muysken and Singh (1986), maintained that the FHC does not constitute a CS-specific constraint. However, because the FHC, like related proposals before it, posited a language identifier as a component of the grammar itself, the FHC must also be seen as a CS-specific constraint.

There are also empirical counter-examples to Belazi, Rubin, and Toribio’s FHC. In (21), for instance, Mahootian (1993, p. 32) documents a Farsi complementizer, which is a functional head, with a sentential complement; (22) present a switch at the same boundary in French-Italian CS (Di Sciullo, Muysken and Singh, 1986, p. 15).

21. Anyway, I figured *ke* if I worked hard enough, I’d finish in the summer
 ‘Anyway, I figured that if I worked hard enough, I’d finish in the summer.’

22. No, *parce que* hanno donné des cours
no, because have given of the lectures
'No, because they have given the lectures.'

See Mahootian (1993), MacSwan (1999, 2013), and Muysken (2000) for additional discussion.

Finally, we turn to a class of CB proposals built around Levelt's (1989) *Speaking* model. Although similar proposals surfaced independently (de Bot, 1992; Azuma, 1991, 1993), Myers-Scotton's (1993) Matrix Language Frame (MLF) Model has stood out as the most influential performance-based CS account. The MLF model differentiates the languages involved in CS as the *matrix language* (ML) and the *embedded language* (EL). According to this approach, the matrix language defines the surface structure positions for content words and functional elements. It is the current dominant model among CB approaches to CS (Wardhaugh and Fuller, 2014).

The MLF model includes two primary principles which serve as constraints – the Morpheme Order Principle, which requires that morphemes within a bilingual constituent follow the order prescribed by the ML, and the System Morpheme Principle, which states that all 'system morphemes' – defined as morphemes which have grammatical relations with other constituents outside their maximal projections – come from the ML in any CS construction. From a theoretical point of view, we see that we immediately encounter the same difficulties as in some other approaches: The grammatical principles responsible for defining the distribution of CS explicitly reference language identifiers, and are therefore CS-specific constraints.

Myers-Scotton (1993) originally defined the ML as the language contributing the majority of the morphemes in an utterance, claiming that the ML 'may change across time, and even within a conversation' (p. 69). Critics expressed concern over the vagueness of this definition (Muysken and de Rooij, 1995; Bentahila, 1995; MacSwan, 1999, 2000, 2005a; Muysken, 2000), leading Jake, Myers-Scotton and Gross (2002, p. 73) to propose the Uniform Structure Principle, which offered a more structurally oriented definition: 'The ML may change within successive CPs, even within a multi-clausal sentence, but we stress that the ML does not change within a single bilingual CP.' Put differently, with a single CP single CP, all grammatical morphemes must come from one language only (System Morpheme Principle), and the language contributing the grammatical morphemes must define the surface order of the utterance (Morpheme Order Principle). Jake, Myers-Scotton and Gross (2002, p. 88) furthermore stressed that the ML/EL distinction is universal, existing 'in monolingual language as well as bilingual language.'

Myers-Scotton and colleagues further contended that an 'EL island' may occur below the CP where they 'are not inflected with ML system morphemes, although they occur in positions projected by the ML, following the Morpheme Order Principle' (Jake, Myers-Scotton and Gross, 2002, p. 77). That is, EL islands are essentially lawful violations of the System Morpheme Principle because they contain grammatical morphemes that are not in the ML, but an EL island must be a maximal projection and must remain true to the Morpheme Order Principle.

Like most other proposed constraints before it, the System Morpheme Principle does not hold up to empirical tests. Consider, for example, the French-Italian data in (23) and (24), reported in Di Sciullo, Muysken and Singh (1986, p. 15).

23. No, *parce que* hanno donné des cours
no, because have given of the lectures
'No, because they have given the lectures.'

24. Oui, alors j'ai dit que *si potev* aller comme ça
 yes so I have said that REF could walk like that
 'Yes, so I said that we could go like that.'

Note that in both cases a switch occurs between an auxiliary or modal and its complement. Because these forms have grammatical relations with other lexical heads within the structure, they meet the MLF model's definition of a system morpheme. Yet, contrary to the requirements of the System Morpheme Principle, each utterance involves system morphemes from different languages below the CP.

To rescue the MLF model, one might argue that [*donné des cours*] in (23) is an EL island, projected as a VP complement of the auxiliary, and that (24) similarly involves an EL island [*aller comme ça*], an IP complement of the modal. However, note that the examples in (23)-(24) contrast with Spanish-English data in (9) where a switch between an auxiliary and a participle is ill-formed. The construction in (9b) is eligible for the same structural analysis as (23), in which an EL island is hypothesized, yet it is ill-formed, contrary to the predictions of the MLF model.

In addition, consider the Spanish-Nahuatl examples in (14). Notice that Spanish negation (*no*) does not tolerate a Nahuatl complement, while Nahuatl negation (*amo*) permits a Spanish complement. Both the agreement morphology on the verbs and negation count as system morphemes since they enter into grammatical relations with other morphemes (in the less obvious case of negation, it c-commands a negative polarity item and may form a syntactic clitic with its verb). Hence, according to the System Morpheme Hypothesis, both (14a) and (14b) should be ill-formed because system morphemes are mixed below the CP, yet this is not so.

Myers-Scotton and colleagues might argue that NegP is an EL Island in (14a) but not in (14b), but with no independent evidence of the status of islands these claims appear to be mere *ad hoc* rationalizations. Jake, Myers-Scotton and Gross (2002, p. 76) furthermore allow 'internal EL islands,' defined as 'a constituent in the EL made up of EL morphemes following EL morpheme order, but smaller than a maximal projection.' In other words, not only can maximal projections be 'islands,' but structural units smaller than EL islands can too – sanctioning essentially any and all CS examples. Grammatical constraints, including proposed CS constraints, cannot be selectively applied; that is, one cannot reasonably claim that negation is an island, immune to the System Morpheme Principle in (14b) but not in (14a). Rather, once created, these mechanisms operate in all cases, and as such the MLF model creates a universe of expectations where essentially all CS is well-formed, contrary to the facts. (For further discussion of the MLF model, see MacSwan, 2005a, 2005b).

The history of CS reveals a common intuition among researchers that theories about CS should be free of CS-specific constraints. Sankoff and Poplack (1981) noted complications associated with the free union of two phrase structure grammars, then reluctantly introduced a CS-specific tagging mechanism to account for their CS data. Others did not address the issue and formal problems directly, while some researchers behind other proposed CS-specific constraints argued that the mechanisms applied to monolinguals as well as bilinguals. While some researchers (e.g., Woolford, 1983; Mahootian, 1993) claimed to have successfully implemented a CF approach, a closer look revealed that the proposed theoretical apparatus lacked adequate detail or was too narrow in scope to achieve that goal.

3. Critical issues and topics

The constraint-free program

A persistent challenge associated with developing a constraint-free (CF) approach to CS has been the Problem of Lexical Alignment. In a bilingual grammar, if the system makes

commitments to word order before inserting lexical items (LIs), then it will invariably permit the structure of either language regardless of how the terminal nodes are lexically filled, generating illicit switches such as **the casa white* as well as unacceptable word order for the languages under analysis (**house white, *blanca casa*). Preventing this result has motivated researchers working within the context of late lexical insertion models to rely on ad hoc mechanisms to bar lexical insertion in such cases. We noted that Sankoff and Poplack (1981) aligned phrase structure rules with lexical insertion rules by positing a *language tag*, and Di Sciullo, Muysken and Singh (1986) used a *language index* to similar effect. The FHC and MLF model also use language identifiers as primitives.

An advantage of early lexical insertion models for CS research is that the Problem of Lexical Alignment never arises. Chomsky's (1995) Minimalist Program (MP), which Chametzky (2003) described as a 'lexical entry driven' approach to syntax, draws items from the lexicon at the onset of syntactic generation and builds structure upward based on the abstract properties encoded in each lexical item (LI). In the MP there are two components of grammar: C_{HL} , a computational system for human language, generally believed to be invariant across languages; and a Lexicon, where parameters of linguistic variation are encoded as lexical features. An operation called *Select* picks LIs from the lexicon and introduces them into a *Lexical Array* (LA), a finite subset of the lexicon used to construct a derivation. *Merge* takes items from the LA and forms new, hierarchically arranged syntactic objects. Movement operations (*Internal Merge*) apply to syntactic objects formed by *Merge* to re-arrange elements within a tree (Chomsky, 1995, 2000). Phrase structure trees are thus built by the application of the operations *Select* and *Merge*. At a certain point in the derivation, an operation called Spell-Out applies to strip away from the derivation those elements relevant only to Phonetic Form (PF) or PHON(ology), prepared for articulation; what remains is mapped to Logical Form (LF) or SEM(antics) by further application of *Merge*.

An LI may be of two types: *lexical*, with substantive content, or *functional*, without substantive content. Each LI is a feature set. Chomsky (1995, p. 54) catalogued four specific kinds of features encoded in LIs:

25. a. Categorical features (N, V, A, P, T, C, . . .)
- b. Grammatical features (ϕ -features, case, . . .)
- c. Inherent semantic and syntactic features
- d. A phonological feature matrix

Note that (25a–c) are largely universal in nature but vary cross-linguistically to some measure. While all languages appear to share substantially the same set of substantive categories such as N, V, and A, languages may vary with regard to the value functional categories assign to grammatical features – some languages have no gender marking, some have two, others three; some have a two-way distinction for number, others a three-way, and still others have none. Perhaps varying most dramatically, phonological matrices (features of type (25d)) of LIs differ substantially cross-linguistically, and so do the respective phonological components responsible for mapping the structure to PF, readying it to be handed over to the articulatory-perceptual interface.

A bilingual grammar has the special challenge of representing potentially conflicting requirements in a single system. For instance, a Farsi-English bilingual will use Farsi OV word order with a Farsi verb, even if the object is English, and English VO word order with an English verb, even if the object is Farsi. Despite these contradictory requirements for Farsi and English, bilinguals expertly navigate the two systems, never deviating from these patterns.

To capture these facts, the architecture of a bilingual grammar must have the capacity to represent linguistic diversity in a way that captures the structure of each language separately and in interaction.

In a lexically driven approach to syntax like the MP, in which language-particular aspects of grammar are represented in the Lexicon, the natural candidate for discrete representation is of course the Lexicon. Features such as (25) pertain to the full range of LIs in a given language. While there are idiosyncrasies, no English noun has grammatical gender, and all Italian nouns do; German overtly marks accusative case on Ns, but Spanish does not. Clicks may be part of an LI's phonological matrix in Xhosa, but not in Romanian. When new items are added to our mental Lexicon, we do not acquire these specific characteristics anew; rather, new items take on these language-particular characteristics as a matter of course. This fact suggests that principles of word formation are at work in generating the language-particular characteristics of features such as (25). In the case of a bilingual, very different sets of LIs are generated with distinct feature sets, and these feature sets have distinct effects on syntactic structure. The distinct sets of LIs generated by the different sets of rules of word formation may be described as the discrete Lexicons associated with each language in a bilingual's repertoire. Hence, discrete Lexicons present organically as a result of the nature of mechanisms which generalize the language-particular properties associated with the elements in (25).

Moreover, not only are the phonological matrices and other features in (25) discrete for these reasons, but so are the phonological processes. Phonologists capture cross-linguistic variation in terms of differences in the rankings associated with phonological rules, called *constraints* in Optimality Theory (OT). As Prince and Smolensky (1993, p. 3) explained,

OT hypothesizes that constraints are prioritized with respect to each other on a language-specific basis. If a constraint A is prioritized above B, we will write $A \gg B$ and say that A is ranked above or dominates B. A ranking of the constraint set – a *constraint dominance hierarchy* – allows the entire set to evaluate alternatives.

Since language-particular phonologies differ with respect to their internal rankings, it follows that bilinguals will have discrete phonological systems, each with a distinct ranked order of constraints. If the systems were combined as one, as in a single model, the distinct rankings would not be preserved, and the phonological processes would not modify structure as expected. More concretely, building on Prince and Smolensky's explanation, if $A \gg B$ and $B \gg A$ are both part of a speaker's phonology, then a ranking paradox emerges, and A would have no priority relative to B. To avoid the paradox, the human language faculty organizes two discrete systems, one corresponding to the phonological requirements of each language. For the language faculty, these are just different constraint dominance hierarchies defined by their abstract properties.

Consider an illustration. Spanish /b, d, g/ are usually realized as stops when following another stop, a pause, or /l/ in the case of /d/ (e.g., *cuando* [kwaɲdo] 'when,' *tengo* [teŋgo] 'I have') but as continuants in intervocalic contexts (e.g., *hada* [aða] 'fairy,' *haga* [aɰa] 'do-subj.3sg') (Lipski, 1994). English does not have this distribution. This difference is represented phonologically by ranking Spanish stricture agreement higher than it is ranked in English, rendering the constraint ranking *Agree(stricture) >> Ident-IO(continuant)*, *Ident-IO(sonorant)* for Spanish and *Identity-IO(continuant)*, *Ident-IO(sonorant) >> Agree(stricture)* for English.

In MacSwan and Colina (2014), we empirically evaluated the theory that phonological systems are discretely represented for Spanish-English bilinguals using this specific potential

conflict. In one experiment, we tested whether Spanish intervocalic approximant allophones of /b, d, g/ would occur in CS contexts when situated between a Spanish vowel and an English vowel at word boundaries (e.g., *Hablamos de mi ghost yesterday*). The goal was to discover whether Spanish-English bilinguals (N = 5, adult simultaneous bilingual Arizonans) would allow a Spanish phonological process to modify English word structure (in the example, /g/ in *ghost*). A second experiment examined whether an English segment could trigger a Spanish phonological process (/s/-voicing) to modify a Spanish word (e.g., *mis ghosts*). The results of a phonetic analysis showed that our bilingual participants switched seamlessly and effortlessly at language boundaries but maintained separation of their phonological systems; participants applied the Spanish phonological processes exclusively to Spanish segments, even in a bilingual triggering environment.

The PF Interface Condition (MacSwan, 2009; MacSwan and Colina, 2014), building on the PF Disjunction Theorem (MacSwan, 1999, 2000), captures this restriction as an epiphenomenon, or emergent property of the linguistic system:

26. *PF Interface Condition (PFIC)*

- i. Phonological input is mapped to the output in one step with no intermediate representations.
- ii. Each set of internally ranked constraints is a constraint dominance hierarchy, and a language-particular phonology is a set of constraint dominance hierarchies.
- iii. Bilinguals have a separately encapsulated phonological system for each language in their repertoire in order to avoid ranking paradoxes, which result from the availability of distinct constraint dominance hierarchies with conflicting priorities.
- iv. Every syntactic head must be phonologically parsed at Spell-Out.
- v. Therefore, the boundary between heads (words) represents the minimal opportunity for code-switching.

With these considerations in mind, Figure 5.1 (MacSwan, 2016a) represents the architecture of a bilingual grammar, where the phonological component is understood to operate discretely in the manner indicated by the PFIC. The model posited here provides a framework for a CF approach to CS.

This system handles the ban on word-internal switching effortlessly since the PFIC implies that bilinguals have separately encapsulated phonological systems, and every syntactic head must be phonologically parsed in one fell swoop. It therefore subsumes the Free Morpheme Constraint in (6) and accounts for data such as (8) and does so without stipulating CS-specific constraints. Also see Wang (2017), Bessett (2017), and Stefanich and Amaro (2018).

However, perhaps more importantly, it has consequences for numerous syntactic phenomena involving head movement, or the merger of syntactic heads, since every head, even complex heads, must be parsed as single objects by the phonological system. Consider once again the contrast noted in (14), where we observed an asymmetrical switch in which Spanish negation would not tolerate a Nahuatl verb, but Nahuatl negation would allow a Spanish verb. Zagana (1988) observed that Spanish *no* functions as a syntactic clitic and forms part of the Spanish verbal complex as a result of head merger. To make a case for this analysis, she points out that Spanish *no* must be fronted with the verb in (27), unlike the adverb in (28).

27. ¿Qué no dijo Juan?
 what not say/1Ss/PAST Juan
 ‘What didn’t Juan say?’

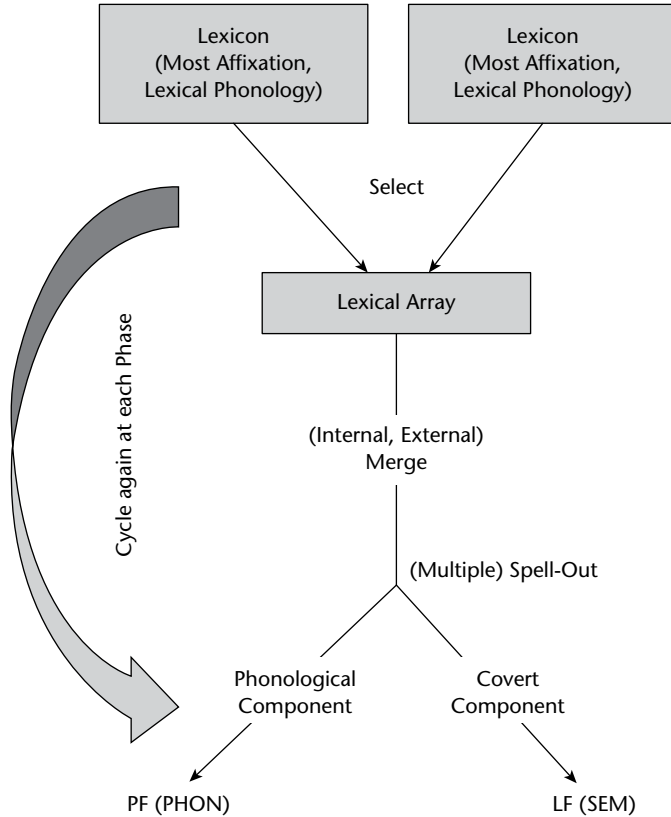


Figure 5.1 A minimalist model of bilingual code-switching

28. *¿Qué sólo leyó Juan?
 what only read/1Ss/PAST Juan
 ‘What did Juan only read?’

Zagona further observed that Spanish *no* cannot be contrastively stressed in (29a) as its English counterpart in (29b) can be, owing to the fact that clitics are inherently unstressable. Based on these and other facts, Zagona concluded that while negation cliticized or merged with the verb in Spanish, it did not do so in English.

- 29a. *Juan no ha *no* hecho la tarea
 Juan not has not done the task
 ‘Juan hasn’t not done the task.’

- 29b. Juan hasn’t *not* done the task

Nahuatl patterns with English in this regard, indicating that negation in Nahuatl is not a clitic:

- 29c. Amo nio *amo* niktati nowelti
 amo ni-o amo ni-k-tati no-welti

not 1S-go amo 1S-3Os-see my-sister
 ‘I’m not going to *not* see my sister.’

Returning now to the CS data in (14), observe that cases involving Spanish negation are ill-formed, but those involving Nahuatl negation are not. This is precisely the outcome expected given the PFIC in (26): Different phonological systems cannot be simultaneously invoked to parse a single (complex) head, causing the derivation in (14a) to crash. For other examples of CS in head movement contexts, see MacSwan (2009, 2013).

Finally, we return to (13), an example of CS between Farsi, an OV language, and English, a VO language. Mahootian (1993) observed that the language of the verb determines word order in these cases: If the verb is English, VO order results, regardless of the language of the object; if the verb is Farsi, OV order results, regardless of the language of the object. This CS pattern has been observed across a wide range of OV-VO language pairs (see MacSwan, 2009 for discussion).

OV and VO languages differ with respect to properties of little *v*. In OV languages, *v* probes the object, triggering movement into its specifier position, but in VO-languages *v* does not. Besides this, in both OV and VO contexts, V syntactically merges with *v* as it raises to T to value its features. This complex head, T+*v*+V, must be phonologically parsed as a single unit in one fell swoop, guaranteeing that in a mixed-language expression all three components will be of the same language, as implied by the PFIC. As a result, the language of the verb is associated with the language of *v* and its abstract features; hence, the language of the verb will determine the position of the object. No similar conditions apply to the DP object, which raises to the specifier position of *v* in OV languages. In this way, we predict that in CS contexts the language of the verb will determine the position of the object, regardless of its language. Numerous other sample derivations could be offered, available in MacSwan (2013, 2016a, 2016b), Cantone and Müller (2005, 2008), van Dulm (2007, 2009), van Gelderen and MacSwan (2008), Cantone and MacSwan (2009), Sánchez (2012), Finer (2014), MacSwan and Colina (2014), Milian (2014), Moro Quintanilla (2014), Di Sciullo (2014), Toribio and González-Vilbazo (2014), and Giancaspro (2015), among others.

In this way the CF approach to CS research derives the facts of language mixing from independently motivated properties of grammar, using the same tools of linguistic analysis as used by linguists working on theories of I-language generally. Departing from the CB tradition, it eschews any proposed principle which identifies languages as primitives of the grammatical system, directly or indirectly (by reference to *(code)switching*, *switch points*, *language feature*, *tag*, or *index*, or any similar concept).

4. Current contributions and research

A bilingual grammar is an integrated system which internally includes shared and discrete systems; syntax and semantics are generally shared systems across a bilingual’s two language systems, with language-particular details specified in the morphology and phonology. Accordingly, two recent topics of research in CS have focused on (1) the nature of bilingual morphology, exploring non-lexicalist approaches to CS, and (2) word internal CS, which touches on the nature of bilingual phonology.

Non-lexicalist approaches to code-switching

Distributed Morphology (DM) was introduced in Halle and Marantz (1993), about the same time that the classical version of the Minimalist Program (MP) was emerging (Chomsky,

1995); while DM focused on the nature of word formation and the MP on the nature of syntax, both included critical assumptions the morphology-syntax interface and the timing of lexical insertion. The MP was explicitly lexicalist, calling upon lexically encoded features to drive syntactic derivations, whereas DM called for late lexical insertion in much the same way as earlier generativist theories had done. The leading aim of DM is to eliminate morphology as a dedicated component of grammar responsible for word formation, and to instead entertain a single morphosyntax which assembles all complex objects, whether they are sub-word elements or phrasal elements.

DM has attracted considerable interest among CS researchers as a tool for the analysis of bilingual data. Examples include González-Vilbazo and López (2011, 2012), Pierantozzi (2012), Bandi-Rao and den Dikken (2014), Grimstad, Lohndal and Åfarli (2014), Lillo-Martin, Müller and Chen Pichler (2016). López, Alexiadou and Veenstra (2017), Riksem, Grimstad, Lohndal and Åfarli (2019), and Grimstad et al. (2018), among others. These researchers are committed to a CF approach, but seek to develop it using DM as an analytic framework. The Problem of Lexical Alignment under a CF approach is as much of a challenge for DM-oriented researchers as it has been throughout the history of the field of CS: Once the syntax has made specific commitments to word order, aligning lexical items to prevent ill-formed constructions (e.g., **the house white*, **the house new you'll buy*) requires a rich feature specification that sufficiently distinguishes between competing items from the languages under analysis, putting pressure on the analyst to add otherwise unnecessary complexity to the feature matrix. DM-style CS research has made some progress in attending to this historic concern; see, for example, recent work by Grimstad, Riksem, Lohndal and Åfarli (2018). Whether CS data are better analyzed within a DM or lexicalist framework is a current topic of debate in the field. (See MacSwan, 2016b for further remarks.)

Word-internal code-switching

Another current topic of interest in the CS literature is word-internal switching. As previously noted, Poplack first codified the ban on word-internal switching as part of her Poplack's Free Morpheme Constraint. In the CB era of CS research, when scholars typically offered counter-examples to constraints while proposing succeeding CS-specific constraints, some offered apparent counter-examples (Bokamba, 1989; Myers-Scotton, 1993; Nartey, 1982; Chan, 1999; Jake, Myers-Scotton and Gross, 2002) while others offered confirmation (Bentahila and Davis, 1983; Berk-Seligson, 1986; Clyne, 1987). MacSwan accepted the Free Morpheme Constraint and general ban on word-internal CS as a descriptive generalization, and offered the PF Disjunction Theorem (MacSwan, 1999), later termed the PF Interface Condition (MacSwan, 2009), as a CF solution.

A debate about word-internal CS has re-emerged in the context of lexicalist and non-lexicalist CF approaches to CS, with advocates of non-lexicalist (DM) schools of thought frequently exploring treatments which sanction (at least some) word-internal CS. As noted, we know that word-internal CS is very rare and might be an illusion where it has been attested (MacSwan, 2005a). The appropriate empirical question is not whether it exists, but where it exists; the appropriate theoretical question is how its distribution may be explained.

To date, empirical research on word-internal CS has been very messy, with scholars often accepting orthographic transcriptions of CS as equivalent to transcriptions of phonetic detail. One cannot reliably extract instances of word-internal CS from corpora which were not transcribed with phonetic detail, as these may in fact be instances of (nonce) borrowing, which involve phonological integration. Further research on word-internal CS should seek to discover

theoretically significant differences in the distribution of empirically well-documented word-internal CS.

5. Future directions

A strong consensus in the field of CS, nearly since its inception, has been that a good theory of CS, minimally, is one that would not appeal to a ‘third grammar’ (Pfaff, 1979; Mahootian, 1993) or a CS-specific device to regulate the interaction of the two systems. The MP provides a framework which permits us to abandon the quest for constraints on CS and engage in the linguistic analysis of mixed-language utterances in the same way in which we engage in the analysis of I-language generally. This research program has enhanced our understanding of the nature of bilingualism, CS, and the architecture of the bilingual language faculty. It has permitted us to move beyond traditional battles over which proposed CS-specific constraint is accurate and which is not, calling upon us to examine CS in the context of specific constructions, operations, and grammatical features, paying close attention to the feature compositions of the specific language pairs under analysis.

The goal, as the field continues on its present course, is to propose increasingly better theories about the nature of the bilingual language faculty as a reflection of the facts of CS, informing the fields of bilingualism, language contact, and general linguistics, focusing narrowly on the nature of the relationship between language-general systems, such as syntax and semantics, and language-specific systems, such as morphology and phonology.

6. Further reading

Bullock, B. E. and Toribio, A. J., eds. (2009). *The Cambridge Handbook of linguistic code-switching*. Cambridge: Cambridge University Press.

This handbook provides a wide-ranging survey of the code-switching literature.

MacSwan, J., ed. (2014). *Grammatical theory and bilingual codeswitching*. Cambridge, MA: MIT Press.

This volume includes cutting edge contributions on the syntax, morphology, phonology, and psycholinguistics of code-switching.

7. Related topics

Language contact, bilingualism, code-mixing, bilingual education, translanguaging

Abbreviations

| | |
|-----|---------------------------------|
| 1S | Subject, first person |
| 1Ss | Subject, first person, singular |
| 3S | Subject, third person |
| 3Os | Object, third person singular |
| CB | constraint based |
| CF | constraint-free |
| CP | complementizer phrase |
| CS | code-switching |
| DP | determiner phrase |
| DUR | durative |
| EL | embedded language |

| | |
|------|----------------------------|
| FHC | Functional Head Constraint |
| G | grammar |
| IP | inflectional phrase |
| LF | logical form |
| LI | lexical items |
| ML | matrix language |
| MLF | Matrix Language Frame |
| MP | Minimalist Program |
| NP | noun phrase |
| OT | Optimality Theory |
| OV | object-verb |
| PAST | past tense |
| PHON | phonology |
| PF | phonetic form |
| PFIC | PF Interface Condition |
| POSS | possessive |
| REF | reflexive |
| SEM | semantics |
| TAG | Tree Adjoining Grammars |
| VO | verb-object |
| VP | verb phrase |

Note

1 Many thanks to Katharine Glanbeck for feedback and assistance.

References

- Abney, S.P. (1987). *The English noun phrase in its sentential aspect*. PhD. MIT.
- Azuma, S. (1991). *Processing and intrasentential code-switching*. PhD. University of Texas, Austin.
- Azuma, S. (1993). The frame-content hypothesis in speech production: Evidence from intrasentential code switching. *Linguistics*, 31, pp. 1071–1093.
- Bandi-Rao, S. and den Dikken, M. (2014). Light switches: On v as a pivot in codeswitching, and the nature of the ban on word-internal switches. In: J. MacSwan, ed., *Grammatical theory and bilingual codeswitching*, 1st ed. Cambridge, MA: MIT Press, pp. 161–184.
- Belazi, H.M., Rubin, E.J. and Toribio, A.J. (1994). Code switching and X-bar theory: The functional head constraint. *Linguistic Inquiry*, 25(2), pp. 221–237.
- Benson, E.J. (2001). The neglected early history of codeswitching research in the United States. *Language & Communication*, 21(1), pp. 23–36.
- Bentahila, A. and Davis, E.E. (1983). The syntax of Arabic-French code-switching. *Lingua*, 59, pp. 301–330.
- Bentahila, A. (1995). Review of *Duelling languages: Grammatical structure in codeswitching*. *Language*, 71(1), pp. 135–140.
- Berk-Seligson, S. (1986). Linguistic constraints on intrasentential code-switching: A study of Spanish-Hebrew bilingualism. *Language in Society*, 15, pp. 313–348.
- Bessett, Ryan M. (2017). *The integration of lone English nouns into bilingual Sonoran Spanish*. PhD. University of Arizona.
- Bhatia, T.K. and Ritchie, W.C. (2013). *The handbook of bilingualism and multilingualism*. Malden, MA: Blackwell.
- Bokamba, E.G. (1989). Are there syntactic constraints on code-mixing? *World Englishes*, 8, pp. 277–292.

- Cantone, K. and MacSwan, J. (2009). The syntax of DP-internal codeswitching. In: L. Isurin, D. Winford, and K. de Bot, eds., *Multidisciplinary approaches to codeswitching*, 1st ed. Amsterdam: John Benjamins, pp. 243–278.
- Cantone, K.F. and Müller, N. (2005). Codeswitching at the interface of language-specific lexicons and the computational system. *International Journal of Bilingualism*, 9(2), pp. 205–225.
- Cantone, K.F. and Müller, N. (2008). Un nase or una nase? What gender marking within switched DPs reveals about the architecture of the bilingual language faculty. *Lingua*, 118(6), pp. 810–826.
- Chametzky, R.A. (2003). Phrase structure. In: R. Hendrick, ed., *Minimalist syntax*, 1st ed. Oxford: Blackwell, pp. 192–225.
- Chan, B.H.-s. (1999). *Aspects of the syntax, production and pragmatics of code-switching with special reference to Cantonese-English*. PhD. University College London.
- Chomsky, N. (1964). *Current issues in linguistic theory*. The Hague: Mouton.
- Chomsky, N. (1965). *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.
- Chomsky, N. (1995). *The minimalist program*. Cambridge, MA: MIT Press.
- Chomsky, N. (2000). Minimalist inquiries: The framework. In: R. Martin, D. Michaels, and J. Uriagereka, eds., *Step by step: Essays on minimalist syntax in honor of Howard Lasnik*, 1st ed. Cambridge, MA: MIT Press, pp. 89–155.
- Clyne, M. (1987). Constraints on code switching: How universal are they? *Linguistics*, 25, pp. 739–764.
- de Bot, K. (1992). A bilingual production model: Levelt’s “speaking” model adapted. *Applied Linguistics*, 13(1), pp. 1–24.
- Di Sciullo, A.-M., Muysken, P. and Singh, R. (1986). Government and code-switching. *Journal of Linguistics*, 22, pp. 1–24.
- Di Sciullo, A.M. (2014). On the asymmetric nature of the operations of grammar: Evidence from codeswitching. In: J. MacSwan, ed., *Grammatical theory and bilingual codeswitching*, 1st ed. Cambridge, MA: MIT Press, pp. 63–87.
- Fasold, R. (1972). *Tense marking in Black English*. Arlington, VA: Center for Applied Linguistics.
- Finer, D.L. (2014). Movement triggers and reflexivization in Korean-English codeswitching. In: J. MacSwan, ed., *Grammatical theory and bilingual codeswitching*, 1st ed. Cambridge, MA: MIT Press, pp. 37–62.
- Giancaspro, D. (2015). Code-switching at the auxiliary VP boundary: A comparison of heritage speakers and L2 learners. *Linguistic Approaches to Bilingualism*, 5(3), pp. 380–408.
- González-Vilbazo, K. and López, L. (2011). Some properties of light verbs in code-switching. *Lingua*, 121, pp. 832–850.
- González-Vilbazo, K. and López, L. (2012). Little v and parametric variation. *Natural Language and Linguistic Theory*, 30(1), pp. 33–77.
- Grimstad, M., Lohndal, T. and Áfarli, T.A. (2014). Language mixing and exoskeletal theory: A case study of word-internal mixing in American Norwegian. *Nordlyd*, 41, pp. 213–237.
- Grimstad, M., Riksem, B.R., Lohndal, T. and Áfarli, T.A. (2018). Lexicalist vs. exoskeletal approaches to language mixing. *The Linguistic Review*, 35(2), pp. 187–218. doi: 10.1515/tlr-2017-0022
- Gumperz, J. (1967). On the linguistic markers of bilingual communication. *Journal of Social Issues*, 28(2), pp. 48–57.
- Gumperz, J. (1970). Verbal strategies and multilingual communication. In: J.E. Alatis, ed., *Georgetown round table on language and linguistics*, 1st ed. Washington, DC: Georgetown University Press, pp. 129–147.
- Gumperz, J. and Hernández-Chávez, E. (1970). Cognitive aspects of bilingual communication. In: W.H. Whitely, ed., *Language and social change*, 1st ed. Oxford: Oxford University Press.
- Halle, M. and Marantz, A. (1993). Distributed morphology and the pieces of inflection. In: K. Hale and S.J. Keyser, eds., *The view from building 20*, 1st ed. Cambridge, MA: MIT Press, pp. 111–176.
- Halmari, H. (1997). *Government and code switching: Explaining American Finnish*. New York: John Benjamins.
- Hasselmo, N. (1972). Code-switching as ordered selection. In: E.S. Firchow, ed., *Studies for Einar Haugen*, 1st ed. The Hague: Mouton, pp. 261–280.

- Jake, J., Myers-Scotton, C. and Gross, S. (2002). Making a minimalist approach to codeswitching work: Adding the matrix language. *Bilingualism: Language and Cognition*, 5(1), pp. 69–91.
- Joshi, A. (1985a). Processing of sentences with intrasentential code switching. In: D.R. Dowty, L. Karttunen, and A.M. Zwicky, eds., *Natural language parsing: Psychological, computational and theoretical perspectives*, 1st ed. Cambridge: Cambridge University Press, pp. 190–205.
- Joshi, A. (1985b). How much context-sensitivity is necessary for assigning structural descriptions? Tree adjoining grammars. In: D. Dowty, L. Karttunen, and A. Zwicky, eds., *Natural language parsing: Psychological, computational and theoretical perspectives*, 1st ed. Cambridge: Cambridge University Press, pp. 206–250.
- Labov, W. (1970). The logic of non-standard English. In: F. Williams, ed., *Language and poverty*, 1st ed. Chicago: Rand McNally.
- Lee, M-H. (1991). *A parametric approach to code-switching*. PhD. State University of New York at Stonybrook.
- Levelt, W.J.M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
- Lillo-Martin, D., Müller, R. and Chen Pichler, D. (2016). The development of bimodal bilingualism: Implications for linguistic theory. *Linguistic Approaches to Bilingualism*, 6(6), pp. 719–755.
- Lipski, J. (1978). Code-switching and the problem of bilingual competence. In: M. Paradis, ed., *Aspects of bilingualism*, 1st ed. Columbia, SC: Hornbeam Press.
- Lipski, J. (1985). *Linguistic aspects of Spanish-English language switching*. Tempe, AZ: Center for Latin American Studies, Arizona State University.
- Lipski, J. (1994). *Latin American Spanish*. London and New York: Longman.
- Lipski, J. (2014). Spanish-English code-switching among low-fluency bilinguals: Towards an expanded typology. *Sociolinguistic Studies*, 8(1), pp. 23–55.
- López, L., Alexiadou, A. and Veenstra, T. (2017). Code-switching by phase. *Languages*, 2(3), p. 9.
- MacSwan, J. (1999). *A minimalist approach to intrasentential code switching*. New York: Garland Press.
- MacSwan, J. (2000). The architecture of the bilingual language faculty: Evidence from codeswitching. *Bilingualism: Language and Cognition*, 3(1), pp. 37–54.
- MacSwan, J. (2004). Code switching and linguistic theory. In: T.K. Bhatia and W. Ritchie, eds., *Handbook of bilingualism*, 1st ed. Oxford: Blackwell, pp. 283–311.
- MacSwan, J. (2005a). Codeswitching and generative grammar: A critique of the MLF model and some remarks on “modified minimalism”. *Bilingualism: Language and Cognition*, 8(1), pp. 1–22.
- MacSwan, J. (2005b). Comments on Jake, Myers-Scotton and Gross’s response: There is no ‘matrix language’. *Bilingualism: Language and Cognition*, 8(2), pp. 277–284.
- MacSwan, J. (2009). Generative approaches to code-switching. In: B.E. Bullock and A.J. Toribio, eds., *Cambridge handbook of linguistic codeswitching*, 1st. ed. New York: Cambridge University Press, pp. 309–335.
- MacSwan, J. (2013). Code switching and linguistic theory. In: T.K. Bhatia and W. Ritchie, eds., *Handbook of bilingualism and multilingualism*, 2nd ed. Oxford: Blackwell, pp. 223–350.
- MacSwan, J., ed. (2014). *Grammatical theory and bilingual codeswitching*. Cambridge, MA and London: MIT Press.
- MacSwan, J. (2016a). Code-switching in adulthood. In: E. Nicoladis and S. Montanari, eds., *Bilingualism across the lifespan: Factors moderating language proficiency*, 1st ed. Washington, DC: American Psychological Association, pp. 183–200.
- MacSwan, J. (2016b). Codeswitching and the timing of lexical insertion. *Linguistic Approaches to Bilingualism*, 6(6), pp. 786–791.
- MacSwan, J. and Colina, S. (2014). Some consequences of language design: Codeswitching and the PF interface. In: J. MacSwan, ed., *Grammatical theory and bilingual codeswitching*, 1st ed. Cambridge, MA: MIT Press, pp. 185–210.
- Mahootian, S. (1993). *A null theory of code switching*. PhD. Northwestern University.
- Milian, H.S. (2014). Codeswitching and the semantics-syntax interface: The role of aspectual features in constraining intrasentential codeswitching involving the verb. In: J. MacSwan, ed., *Grammatical theory and bilingual codeswitching*, 1st ed. Cambridge, MA: MIT Press, pp. 227–254.

- Moro Quintanilla, M. (2014). The semantic interpretation and syntactic distribution of determiner phrases in Spanish-English codeswitching. In: J. MacSwan, ed., *Grammatical theory and bilingual codeswitching*, 1st ed. Cambridge, MA: MIT Press, pp. 213–226.
- Muysken, P. (2000). *Bilingual speech: A typology of code-mixing*. Cambridge: Cambridge University Press.
- Muysken, P. and de Rooij, V. (1995). Review of social motivations for code-switching: Evidence from Africa, and Duelling languages: Grammatical structure in codeswitching. *Linguistics*, 33, pp. 1043–1066.
- Myers-Scotton, C. (1993). *Dueling languages: Grammatical structure in code switching*. Oxford: Clarendon Press.
- Nartey, J.S. (1982). Code-switching, interference or faddism? Language use among educated Ghanaians. *Anthropological Linguistics*, 24, pp. 183–192.
- Nishimura, M. (1997). Japanese/English code-switching: Syntax and pragmatics. In: I. Rauch, ed., *Berkeley insights in linguistics and semiotics*, vol. 24, 1st ed. New York: Peter Lang.
- Pandit, I. (1990). Grammaticality in code switching. In: R. Jakobson, ed., *Codeswitching as a worldwide phenomenon*, 1st ed. New York: Peter Lang, pp. 33–70.
- Pfaff, C. (1979). Constraints on language mixing: Intrasentential code-switching and borrowing in Spanish/English. *Language*, 55, pp. 291–318.
- Pierantozzi, C. (2012). Agreement within early mixed DP: What mixed agreement can tell us about the bilingual language faculty. In: C. Gabriel and K. Braunmüller, eds., *Multilingual individuals and multilingual societies*. Amsterdam: John Benjamins, pp. 137–152.
- Poplack, S. (1978). Quantitative analysis of constraints on code-switching. *Centro working papers*. New York: City University of New York, Centro de Estudios Puertorriqueños.
- Poplack, S. (1980). “Sometimes I’ll start a sentence in Spanish *y termino en español*”: Toward a typology of code-switching. *Linguistics*, 18, pp. 581–618.
- Poplack, S. (1981). The syntactic structure and social function of code-switching. In: R. Durán, ed., *Latino language and communicative behavior*, 1st ed. Norwood, NJ: Ablex, pp. 169–184.
- Poplack, S. and Meechan, M. (1998). How languages fit together in codemixing. *International Journal of Bilingualism*, 2(2), pp. 127–138.
- Poplack, S., Wheeler, S. and Westwood, A. (1989). Distinguishing language contact phenomena: Evidence from Finnish-English bilingualism. *World Englishes*, 8(3), pp. 389–406.
- Prince, A. and Smolensky, P. (1993). Optimality theory: Constraint interaction in generative grammar. In: A. Prince and P. Smolensky, eds., *Optimality theory in phonology: A reader*, 1st ed. Malden, MA: Blackwell Pub, pp. 3–71.
- Raz, M. (2013). *What’s wrong with the poor?: Psychiatry, race, and the war on poverty*. Chapel Hill: University of North Carolina Press Books.
- Riegelhaupt, F. (2000). Codeswitching and language use in the classroom. In: A. Roca, ed., *Research on Spanish in the United States: Linguistic issues and challenges*, 1st ed. Somerville, MA: Cascadilla Press, pp. 204–217.
- Riksem, B.R., Grimstad, M.B., Lohndal, T., and Áfarli, T.A. (2019). Language mixing within verbs and nouns in American Norwegian. *The Journal of Comparative Germanic Linguistics*, 22, pp. 189–209.
- Sánchez, L. (2012). Convergence in syntax/morphology mapping strategies: Evidence from Quechua – Spanish code mixing. *Lingua*, 122(5), pp. 511–528.
- Sankoff, D. (1998). A formal production-based explanation of the facts of code-switching. *Bilingualism: Language and Cognition*, 1(1), pp. 39–50.
- Sankoff, D. and Poplack, S. (1981). A formal grammar for code-switching. *Papers in Linguistics*, 14, pp. 3–45.
- Sankoff, D., Poplack, S. and Vanniarajan, S. (1990). The case of the nonce loan in Tamil. *Language Variation and Change*, 2(1), pp. 71–101.
- Santorini, B. and Mahootian, S. (1995). Codeswitching and the syntactic status of adnominal adjectives. *Lingua*, 96, pp. 1–27.

- Stefanich, S. and Amaro, J.C. (2018). Phonological factors of Spanish/English word internal code-switching. In: Luis López, ed., *Code-switching – experimental answers to theoretical questions: In honor of Kay González-Vilbazo*, 1st ed. Philadelphia: John Benjamins, pp. 195–222.
- Stenson, N. (1990). Phrase structure congruence, government, and Irish-English code switching. *Syntax and Semantics*, 23, pp. 167–197.
- Timm, L.A. (1975). Spanish-English code-switching: el porqué and how-not-to. *Romance Philology*, 28, pp. 473–482.
- Toribio, A.J. and González-Vilbazo, K.E. (2014). Operator movement in English-Spanish and German-Spanish codeswitching. In: J. MacSwan, ed., *Grammatical theory and bilingual codeswitching*, 1st ed. Cambridge, MA: MIT Press, pp. 87–118.
- van Dulm, O. (2007). *The grammar of English-Afrikaans code switching: A feature checking account*. Utrecht: Netherlands Graduate School of Linguistics (LOT).
- van Dulm, O. (2009). English-Afrikaans intrasentential code switching: Testing a feature checking account. *Bilingualism: Language and Cognition*, 12(2), pp. 193–212.
- van Gelderen, E. and MacSwan, J. (2008). Interface conditions and codeswitching: An F-movement analysis of pronouns and lexical DPs. *Lingua*, 118(6), pp. 765–776.
- Vogt, H. (1954). Review of languages in contact. *Word*, 10, pp. 79–82.
- Wang, S-L. (2017). The PF disjunction theorem to Southern Min/Mandarin code-switching. *International Journal of Bilingualism*, 21(5), pp. 541–558.
- Wardhaugh, R. and Fuller, J. M. (2014). *An introduction to sociolinguistics*. 7th ed. Oxford: Wiley.
- Weinreich, U. (1953). *Languages in contact*. The Hague: Mouton.
- Wentz, J. (1977). *Some considerations in the development of a syntactic description of code-switching*. PhD. University of Illinois, Urbana-Champaign.
- Wolfram, W. (1969). *A sociolinguistic description of Detroit Negro speech*. Washington, DC: Center for Applied Linguistics.
- Woolford, E. (1983). Bilingual code-switching and syntactic theory. *Linguistic Inquiry*, 14(5), pp. 520–536.
- Zagona, K. (1988). *Verb phrase syntax: A parametric study of English and Spanish*. Boston: Kluwer Academic Publishers.

Usage-based approaches

Ad Backus

1. Introduction and definitions

A relative newcomer to the field of linguistic theories, the usage-based approach has a lot to offer to our efforts to account for language contact phenomena. The goal of this chapter is to outline this potential and exemplify it with some of the work that has already been done. Needless to say, a lot more work is waiting to be done, and the chapter will go into some of the challenges that lie ahead if we are to realize the potentially ground-breaking possibilities the usage-based approach offers for moving contact linguistics forward.

Various concepts will play a prominent role in this chapter. Some of those are, of course, shared with most other chapters in this volume:

- Code-switching: the use of lexical material from more than one language in the same conversation, sometimes by alternating utterances in each language and sometimes by combining this material in the same utterance or clause;
- Borrowing: when elements (e.g., a word or a grammatical construction) that originate in one language become a conventional part of another language;
- Interference (or transfer): the use of non-lexical material (e.g., a grammatical pattern, a semantic feature, etc.) that originated in one language in an utterance of another language.

A few other concepts are central to the usage-based approach. The most important ones are the following:

- Usage: an all-inclusive reference to language use, covering both speaking and hearing, and all levels of language (sounds, words, grammatical patterns, discourse styles, etc.);
- Knowledge: the mental representation that makes language use possible, including the full inventory of all linguistic knowledge that an individual speaker possesses and that can therefore be activated while communicating;
- Unit: the form-meaning combinations that form the basis of our knowledge. All forms have a meaning: forms can be short or long, and may be internally complex; and the meaning can be specific or abstract;

- Entrenchment: an indication of the degree to which a particular unit (a word, a pattern, etc.) is mastered by an individual speaker.

2. Historical overview

As many pages in the current volume show, contact linguistics is a discipline that boasts several decades of productive research. Few studies have explicitly adopted a usage-based approach, though this approach is by now firmly embedded in linguistics. However, many of the findings of contact linguistics are very compatible with a usage-based perspective. Much of contact linguistics adopts a general descriptive linguistic framework rather than a more specific theory, and usage-based accounts align with this approach as they are supposed to account for all language use. Also, the findings reported in contact linguistic work tend to point to explanatory factors that can fruitfully be articulated within a usage-based account, especially the relevance of past experience and its impact on degrees of entrenchment and ease of activation, and the impact on activation of salient aspects of meaning, including pragmatic intention and social indexicality.

Language contact studies address many different topics, as this volume testifies. Taking their cue from the pioneers Weinreich and Haugen, scholars have addressed questions of psycholinguistic, sociolinguistic or general linguistic nature, leading to the current situation in which many different subfields can be distinguished within bilingualism studies. It is a matter of interpretation whether those are best seen as islands with little communication between them, or as an archipelago with tenuous but real connections, or, to prolong the geographical metaphor for a bit, as a connected land mass with relatively self-supporting settlement areas, including well-populated cities as well as smaller isolated settlements. Communication is as good as the road connections between them.

As a result, different research traditions have developed what may be gathered under the disciplinary heading of ‘contact linguistics’ but which are only loosely connected. They deal with different aspects of the same phenomenon, language contact, but do not share an overarching theory that pulls the different traditions together. Instead, each contact linguistic sub-discipline gets theoretical and methodological inspiration from its ‘parent’ discipline and constructs its own models. In sociolinguistics, for example, there is intense interest in the way in which language use reflects and impacts social inequality: sociolinguistic studies of language contact, therefore, have illuminated the relationship between the languages in contact, often framing it in terms of power and solidarity. The empirical focus has often been on the determinants of language choice in conversation among multilinguals and, in the long run, its implications for language maintenance or shift. Given that language choice is often not clear-cut, an interest in the social and pragmatic motivations for speaking the languages side-by-side, i.e., in code-switching, was inevitable. Likewise, psycholinguistic studies of bilingualism have echoed the parental discipline’s interest in the mechanics of speech production and comprehension. This has naturally led to models that attempt to account for the possibility of switching between two systems, and the inhibition needed to keep it from happening in the many circumstances in which code-switching is not deemed appropriate.

General linguistic approaches, while not dismissive of sociolinguistic and psycholinguistic determinants of language use, generally aim to document and account for what kinds of contact phenomena are possible and which ones are not. Here too, characteristics of the larger discipline have informed the themes chosen in contact-linguistic work. The focus on structural generalizations in linguistic typology has inspired a tradition in contact linguistics on

uncovering typical ways in which grammatical systems respond to the pressure of another language: what grammatical aspects are easily copied, which ones are vulnerable to outside influence, and under what circumstances can language contact affect a language's overall typological profile (see for example Matras, 2009; Heine and Kuteva, 2005)? Similarly, generative linguistics, which approaches the explanation of grammatical structure through the discovery of constraints on variation, has led to a decades-long search for universal or language-particular syntactic constraints on how languages can be combined in bilingual speech (cf. Muysken, 2000). More recent times have seen the development of the 'Heritage Languages' paradigm, which echoes the general trend in linguistics to pay more attention to the impact of extralinguistic factors (such as the age at which speakers become bilingual and the degree to which they use both languages) and their interaction with linguistic factors, such as the relative cognitive difficulty, and therefore vulnerability, of constructions that rely on both syntactic and pragmatic information (the 'Syntax-Pragmatics Interface'), cf. Pires and Rothman (2009) or Montrul (2016). This has led to a closer integration with both sociolinguistic and psycholinguistic approaches. The usage-based approach goes even further, suggesting that all of linguistics should reduce to an integrated sociolinguistic + psycholinguistic framework (Backus, 2015).

3. Critical issues and topics

In my perspective, contact linguistics has been dominated by two main questions: how can two languages be combined in bilingual speech, and how do languages change due to external influence? From the vantage point of usage-based linguistics, it may be argued that these are two aspects of the same question. Accounting for change is the central concern, and to understand how change unfolds one needs to understand how speakers combine elements from two or more languages. The various sub-disciplines (the 'islands' in my metaphorical representation) all deal with different aspects of the question. Thanks to the common superordinate topic, the connections are, if not fully developed or even articulated, at least present in rudimentary form. Footpaths can develop into well-travelled highways.

The reason why a usage-based lens forces us to integrate the questions of mixing and change is that it incorporates a tight link between synchronic language use and diachronic language development (cf. Croft, 2000). Seen through the prism of how synchronic and diachronic dimensions of language relate to each other, connections between the different sub-disciplines also become more visible. In linguistics, the two dimensions have largely been allocated to different corners of the discipline, and their interaction has not been a primary concern. Diachronic linguistics has looked for the origins of change primarily in the structural characteristics of grammatical constructions, some of which may make a construction vulnerable to change or influence from another language. Speaker behaviour has received relatively little attention. In a usage-based approach this separation is impossible, because of its insistence on the assumption that synchronic language use has immediate effects on mental representation. On the other side of the fence, studies of synchronic language use have largely been conducted without much concern for how it relates to diachronic change. Studies of code-switching, notably, have shied away from linking the phenomenon to changes in the languages involved, focusing instead on the linguistic or psycholinguistic mechanisms that produce bilingual speech.

The essence of the usage-based idea is that language is like other culturally transmitted tools, and that therefore all our linguistic knowledge results from how usage is processed by our innate cognitive learning and processing skills (Geeraerts, 2006). It was first formulated

in Langacker (1987) as one aspect of a paradigm that soon came to be known as Cognitive Linguistics. It was positioned and widely understood as an alternative to generative linguistics, because of its commitment to avoiding the assumption of an innate language faculty, and soon gave rise to a number of research traditions and theories, including a reinvigorated semantics and the syntactic theories known collectively as Construction Grammar (Goldberg, 2006). Of course, there are some aspects of human behaviour that do seem to be innate, and this includes both our social need to communicate and our domain-general cognitive skills, such as pattern recognition and the ability to abstract away from individual events and form schematic templates on the basis of perceived similarities across events.

The approach parts company with previous influential schools of thought in linguistics in two ways. First, it sees all structure as emergent, as the solidification of usage, rather than as following from the architecture of Universal Grammar (Bybee, 2010). Second, the strict division of linguistic knowledge into different modules is seen as unwarranted. The definition of language that is adopted is startlingly different from conventional definitions, and one that is very accommodating to the findings of contact linguistics. ‘Language’ is the accumulation of many individuals’ languages: the inventory of units in a speaker’s memory. Various aspects of this definition are crucial for a proper understanding of the implications for contact linguistics. First, a unit is always a combination of a form and a meaning. The form can be ‘specific’ or ‘schematic,’ and if it is more than a morpheme or a monomorphemic word, it is internally complex. This means that many units will be ‘partially schematic’: constructions of which one or more elements are lexically (or ‘morphemically’) fixed. Second, the units in the inventory are mastered to different degrees. Units that are well entrenched are known so well that activating them, whether for production or comprehension, is easy; other units are not as well entrenched and require more cognitive effort to activate. Third, the inventory is not static. Obviously, when you learn a new word it gets added to the inventory, but more important is that degrees of entrenchment are claimed to fluctuate. The more a unit gets activated, the more entrenched it gets; the longer it goes without getting activated (i.e., without being said or heard), the less entrenched it gets. Finally, what is normally referred to as a ‘language’ is the common denominator of the shared inventories of a group of people (which can be as small as a handful of individuals or as large as millions) that all live in the same speech community.

To gauge the usefulness of the usage-based approach for contact linguistics, the remainder of this chapter will go over various prominent lines of research and see what a usage-based account can add, without claiming exhaustiveness. We will explore whether it provides a coherent account of familiar findings, suggests possible explanations for issues that have proved troublesome for other approaches, and opens up new avenues of research.

Sociolinguistics

Perhaps the most basic of all questions in multilingualism research is how speakers use their languages. However, obvious as the existence of entities called languages seems to be, given how we talk about language in everyday life and in education, defining what a language is has proven notoriously difficult, both in modern sociolinguistics and in usage-based linguistics. Both fundamentally see language as a social construct, an abstract category that emerges from behaviour. This makes language choice a social act, and this presupposes that speakers recognize individual words and structures as belonging to a mental category referred to with the name of a particular language. The category exists because it has proven useful in life. The social question is what that use is; the cognitive question is how we are able to do this. These

questions have given rise, respectively, to sociolinguistic studies on language choice and code-switching, and psycholinguistic studies on language separation.

Bilinguals choose their languages according to community conventions as to what is appropriate in particular settings and with particular interlocutors, and in many cases the vernacular is a bilingual register, i.e., a way of speaking that includes abundant code-switching. Speakers may choose to follow or violate these conventions; conventional and deviant choices alike may index an intended pragmatic or social meaning. In the majority of conversations, the choice of language is so obvious that people do not have to think about it. This means the pattern is so entrenched that the choice is made automatically. One may routinely speak English at work and one's native language at home, for example, and it would be weird to make any other language choice. However, in bilingual communities, people often contest the ideologies behind what is supposedly right or desired, for example because choosing the home language conveys group solidarity and the societal language conveys embracing modernity, and individual speakers may not wish to choose between the social identities these two poles index. As a result there may be variation in language choice patterns, and people may mix the languages to avoid the binary choice; these choices may go hand-in-hand with contact effects and produce new varieties, e.g., a heavily influenced variety of the minority language or an ethnolectal variant of the majority language, which may have indexical potential of their own.

All of this can be couched in usage-based terms relatively easily. Community conventions are by definition recurrent patterns of behaviour, and any recurrent pattern will be represented as schematic knowledge in an experienced speaker's mind, similar to other behavioural scripts that describe common aspects of behaviour, such as how to dress for work or how to talk to children. Whenever a linguistic choice attracts attention, for example when it violates a convention, it is potentially salient; this will be particularly prominent in cases where the conventions are backed up by ideologies of language choice, as is often the case in multilingual settings. Salient choices have the potential to set in motion some kind of social change. Language revitalization efforts, for example, are often based on turning around the negative associations attached to speaking the language that people are shifting away from.

Despite the ubiquitous nature of language choice, a logical implication of the usage-based approach is that there is nothing inherently special or different about multilingual language use. Speaking is basically a matter of choosing a message to convey and of activating the units that together convey that message. The basic units are form-meaning combinations; that they 'belong' to a particular language may be part of their meaning if that is relevant, as it often is in multilingual settings. This makes for an interesting connection between usage-based linguistics and the literature on the pragmatic effects of code-switching (or 'translanguaging'; cf. Garcia and Wei, 2014). The basic insight behind these approaches is that language, from the most specific level of the individual word to the most schematic level of the entire language, has the potential to index social meaning, capitalizing on associations people have with speakers who typically use that unit or that language. Empirical questions include when linguistic units or languages evoke this social meaning, what exactly those social meanings are, how widespread or individual they are, how constant or dynamic, etc. The relevant literature contains many findings that are of interest to usage-based theorists.

A predominant finding has been that a switch in language often conveys pragmatic meaning but that the marking potential for an individual switch goes down as the frequency of code-switching goes up. If code-switching is frequent, this way of talking itself conveys social meaning, such as marking a bilingual or bicultural identity.

Typical for 'translanguaging' studies is attention for the indexicality of individual features, though their indexicality is often inherited from the indexicality of the language that it belongs

to. This is a classic case of the link between a specific unit (the individual feature) and the schematic category (the language) it instantiates. The usage-based prediction would be that not just the form but also the meaning of the schematic category will be more schematic, as it is the result of generalizing ('schematizing,' 'abstracting') over many similar but not identical instances. The meaning of the individual features, as used in specific usage events (an utterance in a conversation, a song lyric, a TV advert, a social media post, etc.), will be much richer in detail.

The question remains where the indexical meaning that may be associated with a language comes from. A child growing up in a monolingual English-speaking environment will not start out including the semantic feature 'belongs to English' to the word 'chair' when learning it. In contrast, a bilingual child who learns English and German at the same time in a One Parent One Language household may add language indexes to the meanings of the words 'chair' and 'Stuhl' early in development, because it may notice that the parents differ in the way they talk, and perhaps start explaining that the difference is that they speak different 'languages.' Finally, a child growing up in a bilingual Turkish-German family in which abundant mixing of the languages is routine may be slow to add a language index to the meaning of the form 'Stuhl' and use it indiscriminately in utterances that we could recognize as grammatically 'Turkish' or 'German.' The language index may get added nevertheless as soon as the child has seemingly synonymous forms for the same meaning (the Turkish and the German word) and becomes aware of the existence of two languages, an awareness impossible to avoid in modern societies with their emphasis on language separation. Our sample monolingual child may never attach much social information to the word 'chair,' but that is only because it is a word used in all registers and not only in particular ones; other words do get social information represented in their stored meaning, such as 'informal' for a word only acceptable in colloquial discourse. Associations with languages, dialects, and registers all contribute to the entrenchment of the social aspects of meaning for specific words and expressions as well as the maximally schematic forms themselves. This is what ultimately allow studies of code-switching to conclude that a particular speaker switched at a particular point in discourse from Language A to Language B to index, say, the power associated with Language B.

It is difficult to find communities in the world in which the emergence of socio-cultural entities called languages did not happen. Just like with other social categories people find useful, the idea of a language and therefore also of distinct languages, have proven to be useful abstractions from everyday experience. Most likely this is because the more basic processes of accommodation, entrenchment and conventionalization make people talk more differently the less interaction they have, and behavioural differences between groups makes distinguishing between these groups easier. The result is that we experience the distinguishing of separate languages and registers as relatively natural. The bigger the need to make the distinction, the greater the perceived usefulness of language separation.

Psycholinguistics

It is perhaps inevitable that the study of bilingualism will always have the tendency to compare bilinguals to monolinguals, since the difference from monolingualism is the very reason why the field exists in the first place. In fact most people in the world are bi- or multilingual but at the same time it may be argued that all of our engrained ideas about language are based on a monolingual mindset, on a worldview in which it is natural to distinguish separate languages, to keep them separate in usage, and to learn them as separate tools of communication. As we have seen, sociolinguistics investigates how deviations

from this monolingual practice manifest in bilinguals' lives and how these deviations give rise to contact phenomena; likewise, psycholinguistic studies of bilingualism investigate to what extent bilinguals keep their languages separate in the mind, as this is where they differ from monolinguals who do not experience the effects of the co-activation of another language and the cross-linguistic influence this may give rise to, or the need to inhibit activation of the other language. Important individual background factors that tend to account for the performance of bilinguals on psycholinguistic experimental tasks include the degree to which speakers are used to keeping the languages separate in everyday life, and the degree to which they are proficient in their languages; important experimental factors that co-determine outcomes are whether speakers are induced to be in a rather bilingual or a monolingual mode, whether any of their verbal output is primed, and whether the linguistic structure under consideration is considered robust or vulnerable (e.g., Kootstra, Van Hell and Dijkstra, 2012; Kroll and Bialystok, 2013; Gollan, Starr and Ferreira, 2015). Up to a point, one could argue that most of psycholinguistics implicitly adopts a usage-based approach. Psycholinguistics and usage-based linguistics share that they attribute importance to memory strength for the activation of units.

Linguistics: code-switching and borrowing

Since a usage-based lens forces the analyst to consider synchrony and diachrony in relation to each other, it forces a perspective on the difference between code-switching and lexical borrowing that is at odds with much of the literature in contact linguistics, and perhaps points a way out of a terminological battle that has long dominated the field (Backus, 2014b). Code-switching would be defined as a synchronic phenomenon, as it refers to the use of elements that originate in two different languages in the same discourse. Borrowing, on the other hand, would be a strictly diachronic phenomenon: the more often an individual uses or encounters a particular foreign-origin element, the more entrenched it becomes. This process of increasing entrenchment in individual speakers, and if the same process unfolds in the minds of many different speakers also the process of increasing conventionalization in a community, holds for all lexical elements; loanwords are only special because they originated in the other language. In a usage-based perspective, therefore, the question is not whether a particular word is a code-switch or a borrowing; instead, there are two independent questions: is the word used to highlight the indexical social meaning associated with the language it originates from (in which case it is a code-switch), and is it highly entrenched in the minds of the individual speaker who used it, and those of other speakers with similar backgrounds (in which case it is a borrowing or loanword)? The independence of these questions should not be overstated: the more entrenched the borrowing is, the bigger the chance that the foreign origin is not a salient aspect of its meaning: as a result it cannot easily be used to index the associations that go with its language of origin, and thus it cannot be regimented easily for code-switching (or translinguaging) purposes. Since most individual foreign-origin words in bilingual data sets occur only once or just a handful of times (so-called nonce loans; cf. Poplack and Dion, 2012), they probably have relatively low levels of entrenchment in general, but that does not mean they cannot count as established borrowings. Most of the time, loanwords have relatively specific semantics, and are therefore not very frequent in overall language use, a characteristic they share with all other non-basic content vocabulary. Such words are entrenched for the speakers who use them or hear them, just not as well perhaps as words with greater frequency, such as basic vocabulary items and function words.

Linguistics: modularity

What was said earlier about the relationship between code-switching and lexical borrowing also holds for the relationship between synchronic grammatical interference and diachronic grammatical change (Backus, 2014a) and between lexical interference and lexical contact-induced change (or ‘loan translation’; Backus and Dorleijn, 2009).

From a usage-based perspective the traditional division between lexical and syntactic borrowing, or Matter and Pattern Borrowing (Matras and Sakel, 2007), needs rethinking. If all units can be arranged on continuums of schematicity and internal complexity, then the way these factors play out in contact-induced change has to be explored. Traditionally, lexical borrowing is studied in connection to code-switching and grammatical borrowing or convergence as a separate field (often with a strong connection to historical linguistics, the traditional home for studies of language change). Loan translations are generally not a prominent point of attention, though examples of calques are often found in the same data sources. This division into different research traditions despite the co-occurrence of the underlying phenomena is presumably related to the modularity of linguistic description, with its separate sub-fields of phonology, lexicon, morphology, syntax, semantics, and pragmatics. A usage-based perspective forces the analyst to examine the similarities and differences between the different contact phenomena because all units in the inventory are attributed the same structure (a combination of a form and a meaning), and to account for the differences with reference to the only characteristics of these units that are variable: (1) the degree of phonological specificity (high for words, word combinations, and expressions; low or zero for morphological, syntactic, and discourse structures); (2) the degree of semantic specificity (high for content words with relatively specific meaning, cf. Backus, 2001; Winter-Froemel, 2019, and discourse structures with pragmatic salience, such as discourse markers, momentarily flagged words and expressions, and marked word orders; low for basic vocabulary and grammatical patterns); and (3) the degree of internal complexity (high for grammatical patterns and multiword expressions and low for single words).

Linguistics: equivalence

One of the major tasks of linguistics is to explain how speakers manage to produce output that is ‘well-formed,’ in the sense that it makes sense (‘felicitous’) and that it follows grammatical rules (‘grammatical’). The reformulation of this in a usage-based framework is that linguistics needs to discover and explain the conventions found in human communication. Explanations are suggested in terms of fundamentals of human communication and of human cognition. Using conventional units makes you understood, is cognitively easy (Clark, 2013), and shows that you belong to the community. Applied to the basic issues of contact linguistics, this suggests two different routes through which borrowing can occur, one typical for specific units (roughly overlapping with ‘lexicon’: you select the units that best make you understood) and one typical for schematic ones (‘grammar’: the easiest activated units get used). In both cases, though, some assessment of semantic equivalence has to be established between a foreign-origin unit and a native one (a ‘pivot,’ cf. Matras, 2009). People use foreign-origin words if they can assume that they are conventional in the community (i.e., already established as loanwords) or are necessary for conveying information that cannot be conveyed by a more conventional form (i.e., a translation equivalent in the base language). When a foreign word is making its first inroads, speakers presumably compare its meaning to those of what is available in host

language form-meaning units, and it is logical that the more specific the meaning of the putative loanword, the better the chance that it will be used when needed.

Schematic units, on the other hand, tend to have very general meaning and thus they are not likely to be activated through a conscious selection effort: there will normally be a host language grammatical pattern that already conveys the exact same schematic meaning. The foreign-origin structure will be used ‘in the other language’ (i.e., with words from that language) only when it is entrenched enough in the mind of a speaker. Since entrenchment depends for a large part on frequency, and frequency will only be high enough for speakers who use the other language a lot, and therefore highly proficient in it, grammatical borrowing tends to occur only when the level of bilingualism is high (one of Thomason and Kaufman’s 1988 generalizations). Therefore, lexical borrowing is sensitive to salience and grammatical borrowing to frequency, and this is related to the nature of the meaning of specific units (high in specificity) and that of schematic units (high in schematicity).

Having surveyed some of the ways in which a usage-based perspective relates to the traditional concerns of contact linguistics, the next section will present some relevant recent work.

4. Current contributions and research

Bearing in mind that much of contact linguistics adopts theoretical frameworks that are compatible with the usage-based approach, it is impossible to present an overview of usage-based contact linguistics. In this section I merely summarize some of the work that explicitly self-identifies as such.

Along with the field of Cognitive Sociolinguistics (Geeraerts, Kristiansen and Peirsman, 2010), a cognitive-linguistic approach to language contact has been developing since the late 1990’s. Initially, this mostly meant the incorporation of concepts such as frequency, entrenchment, salience and specificity into accounts of code-switching (Backus, 1996), but gradually the concepts, issues and methods of Cognitive Linguistics have increasingly been applied to all contact phenomena. Zenner, Backus and Winter-Froemel (2019) collects a number of articles that represent this trend. Together, they build on the advances made in Cognitive Sociolinguistics, help overcome the monolingual habitus of most work in Cognitive Linguistics (Vaid and Meuter, 2018), and seek cognitive underpinnings of phenomena often accounted for in structuralist terms (Verschik, 2019, pp. 76–77).

Meaning

A usage-based perspective forces attention on meaning, and on the conceptualization process that gives rise to it. This is uncontroversial for the study of lexical issues (cf. Backus, 2001; Verschik, 2008; Zenner, Speelman and Geeraerts, 2014; Aaron, 2015; Serigos, 2017 on semantic specificity as a factor predicting borrowability of words and expressions). Research on grammatical borrowing, however, tends to focus on form only (e.g., clausal word order). The meaning of patterns and forms often does figure in the analyses, under the heading of ‘grammatical function,’ but explanations for transfer tend to focus on the inadvertent activation of foreign structure. However, since change does not occur randomly throughout the grammatical system, meaning or function are likely to be part of the explanation. Relevant for a usage-based analysis is the insight from grammaticalization studies that an increase in frequency of use, a widening range of uses, and semantic erosion often go hand in hand (Heine and Kuteva, 2005). This may have to do with the increasing likelihood that a speaker will not devote much conscious attention to a form-meaning unit that is well entrenched, easily activated, and has

general meaning. It is activated almost unconsciously, produced almost automatically, and the need for this is a design feature of human communication with its reliance on fast and efficient processing (Christiansen and Chater, 2016; Clark, 2013).

Frequency

Many studies explore the theoretical importance of frequency of occurrence. Backus (2003) noted that among inserted Dutch nouns in Turkish-Dutch code-switching data, many had preserved their Dutch plural marker. Tellingly, these were nouns that in everyday Dutch tended to be used most often in their plural form. Other inserted nouns were in their singular base form and inflected with the Turkish plural marker. This suggests these plural nouns are stored as such in the mind and get activated as fully formed units. Otherwise it would be hard to explain why sometimes a Dutch plural marker is found and sometimes a Turkish one. High frequency helps entrench multimorphemic combinations. Hakimov (2016a) provides much more sophisticated empirical underpinnings for this claim, based on a careful comparison of similar plural nouns from German inserted into Russian clauses in the speech of Russian immigrants to Germany, with reference to independent frequency counts of the relevant words in a large corpus of spoken German. In Hakimov (2015), a similar analysis, with similar outcomes, is carried out for adjectives that often co-occur with particular nouns and get inserted along with the noun in so-called Embedded Islands, i.e., multiword insertions. Further evidence comes from Brown (2015), who studies the greater phonological reduction in words that have a shared representation because of cognate status, and thus enjoy higher usage frequency.

Constructional borrowing

The focus in usage-based linguistics on grammatical constructions rather than on an abstract syntactic architecture finds an echo in usage-based studies on constructional borrowing. Like earlier contact linguistic studies (e.g., Silva-Corvalán, 1994; Verschik, 2008), usage-based studies of contact-induced change often involve exhaustive quantitative analyses of corpora. The difference lies largely in how the findings are accounted for. Doğruöz and Backus (2009), for example, analyzed their corpus of Turkish as spoken by second generation immigrants in the Netherlands for the occurrence of a number of constructions, both in their inherited and in their Dutch-influenced form, providing a picture not just on what is changing in the language but also on how far along the change in question is. This way they were able to show that most changes are at an early stage, and often reflect the literal translation of a Dutch expression, which suggests that grammatical changes start out as more concrete instances of literally translated, specific, expressions. Backus, Doğruöz and Heine (2011) take this one step further, comparing rates of change thus derived to those of changes considered more advanced, such as the development of an indefinite article in Upper Sorbian, a language with many *centuries* of contact history (with German) rather than a few *decades*. The difference is mainly one of quantity: the younger contact situation has some examples in which the indefinite article is used in an innovative way, and probably triggered by its normal (i.e., entrenched) usage in Dutch. Turkish can use the indefinite article *bir* to indicate pragmatic contrast: *akustik bir gitar* contrasts with *akustik gitar* (both mean ‘an acoustic guitar’) in that it emphasizes that the guitar is acoustic (as in ‘he has an *acoustic* guitar, not an electric one’) while Dutch simply uses the indefinite article in the great majority of indefinite noun phrases. Due to Dutch influence bilingual speakers sometimes used the indefinite article without this contrastive meaning. However, most of the time they used the indefinite article in the same way as speakers in

Turkey do. In the older contact setting of Sorbian, the change has more or less reached completion in the sense that the indefinite article construction it borrowed from German is used all the time, just like in German.

Many recent studies analyze ongoing contact-induced change in contact settings by examining variation in the use of particular constructions affected by foreign influence: they compare the rates at which different groups of speakers (e.g., generations), including a monolingual control group where possible, use the construction. Onar Valk (2015), Moro (2016), and Irizarri van Suchtelen (2016) all develop usage-based explanations for their findings on contact-induced change in immigrant languages spoken in the Netherlands: subordinate clause structure in Turkish, various clausal structures in Malay, and dative structures in Spanish, respectively. They each attribute the constructional change to the conceptual influence implied in Slobin's (1991) 'Thinking for Speaking' proposal: the utterance is planned in the majority language and then realized with the morphemes and words from the minority language that best convey the content. Structural borrowing may be the inevitable by-product of conceptual influence, of myriad cases of calqued expressions with highly specific meanings rather than the direct activation of a fully schematic foreign-origin pattern with highly schematic meaning.

Quick, Backus and Lieven (2019) analyzed the bilingual language use of one child learning German and English simultaneously. Although this child was largely raised with a One Parent One Language strategy, he produced quite a lot of switching. At first sight, this could be seen as evidence against a usage-based approach: if the child does not hear mixing in the input, then it must be constructing the bilingual utterances by combining two lexicons and two grammatical systems. However, application of the 'traceback' method (Lieven, Salomo and Tomasello, 2009) to the data, in which all recorded output of the child (words, multiword combinations and constructions) are checked against all previous data and parental input data, showed that many portions of the child's utterances could be traced back to the input in the two languages. Both the child's bilingual utterances and the parental input utterances contain many recurrent utterance-initial chunks, some of them completely fixed and others 'partially schematic,' meaning they consist of one or more fixed elements and one or more open slots. This is in keeping with other usage-based work on acquisition, which generally shows a lot of repetition and imitation; presumably this is how children build up their inventory (Tomasello, 2003). Code-switching in this case usually consists of the creative combination of two or more chunks that are each stored on the basis of past frequency and originate in different languages, or the use of an item from one language used to fill a slot in a partially schematic construction anchored by functional words or morphemes from the other language. Quick et al. (2019, p. 21) conclude that two usage-based assumptions receive support from their data: first, usage matters in determining what gets stored, since children's language use correlates heavily with input (see Gaskins, Backus and Quick, 2019 for similar results). Second, syntactic knowledge seems to emerge on the basis of a slow process of abstraction or schematization, on the basis of the storage of lexical units and partially schematic constructions.

Ease of activation

Usage-based explanations of speaking posit a central role for the ease of activation, as speakers tend to select the form-meaning units that are the easiest to activate, often because they are the best entrenched. Ease of activation is often researched through priming, also used in psycholinguistic studies on language alternation (e.g., Kootstra, Van Hell and Dijkstra, 2012). A few studies have looked at priming 'in the wild,' as a phenomenon that helps determine what

we say in everyday language use. Like many other studies, Travis, Torres Cacoullos and Kidd (2017) aimed to find out how Spanish-English bilinguals expressed the subject in Spanish clauses. They found that there is a priming effect, meaning that the 1st person singular subject pronoun *yo* was used way more often if it had been used in one of the immediately preceding clauses and, interestingly, that the effect extends across languages, but in attenuated form. That is, if the English pronoun *I* was used, *yo* was used significantly more often than chance would predict. This suggests that the speakers operate both with partially schematic constructions [*yo* + clause] and a schematic construction that abstracts away from the actual form of the subject pronoun [overt pronoun + clause]. This provides some insight into an important open question: to what degree does a speaker's inventory contain fully schematic constructions at all, in addition to partially schematic ones? The cross-linguistic priming effect supports the cognitive reality of the maximally schematic construction (for similar corpus-based studies investigating priming effects, see Marzo, Zenner and Van de Mieroop, 2019 and Quick et al., 2019). Höder (2014) and Wasserscheidt (2016) articulate attempts to reformulate Construction Grammar to account for such cross-linguistic constructional similarities.

5. Main research methods

In a rudimentary sense, many contact linguistic studies have always been usage-based in their methodology, as they are often based on data that reflect natural language use, in the form of collected recordings of spontaneous or elicited conversation, or judgements about ordinary language use collected through traditional linguistic fieldwork methods. Theoretical claims are often based on empirical generalizations made on the basis of such data and framed through the traditional structuralist ways of describing language. Given the straightforward usability of corpora in usage-based studies, such theoretical claims could obviously be framed in a usage-based framework. Examples of such work on language contact that fits squarely in the traditions of corpus-based usage-based linguistics are the aforementioned studies Quick et al., 2018a, 2018b, 2019a, 2019b on bilingual child language acquisition and Hakimov's series of studies on German-Russian contact (Hakimov, 2015, 2016a, 2016b). The latter is a textbook example of the derivation of testable hypotheses from usage-based theory, and their statistical testing through comparisons of frequencies in a bilingual corpus and monolingual reference corpora (also see Broersma and De Bot, 2006 for a pioneering example).

Mainstream linguistic studies often make use of corpora that are much bigger than what will normally be possible in contact linguistics. Depending on the degree to which a corpus can be seen as representative, its analysis allows generalizations about the manifestations and frequencies of grammatical constructions and about the kinds of socially stratified variation that forms the basis of sociolinguistic theory. However, a large corpus is easiest to assemble if the input data are from written sources; written language is, of course, not always available as some languages are not written and others only sporadically used in specific registers. In addition, written language will generally be different to some extent from spoken language, making its usefulness limited for research questions that deal with the impact of everyday language use on mental representation and vice versa.

As contact situations often involve socially disadvantaged languages with low prestige and no or limited manifestations in writing, contact linguists will rarely have the funding and manpower to build corpora that allow full use of the analysis techniques of corpus linguistics (Backus, 2014b). Nevertheless, the last decade has seen vast improvement.

Adamou (2016) may serve as an example of what can be done with relatively small corpora. In several studies of different minority languages, including Ixcatec, Romani, a variety

of Croatian spoken in Austria, Molise Slavic, and Upper Sorbian, it explores the degree to which corpus linguistics holds lessons for contact linguistics. Another example is Torres-Cacoulos and Travis (2018), part of a concerted team effort to build a corpus of New Mexican Spanish, and to use it to produce quantitative and qualitative analyses of English influence, or its absence, on various aspects of Spanish grammar, especially the use or absence of subject pronouns. Like other work referenced previously, they derive a testable hypothesis from linguistic theory (namely that pro-drop Spanish will increase its use of subject pronouns under the influence of non-pro-drop English), and use corpus linguistic analytical methods to test it (it does not). In this case, the analysis is not phrased in usage-based terms, but usage-based approaches, of course, should be able to account for the same results.

While the issues discussed so far belong to the time-honoured tradition of contact linguistics, usage-based theory sometimes calls for additional questions, and this has started to impact work on language contact. Shared with all usage-based work is the hunt for the best way of measuring entrenchment. If explanations refer to ease of activation, and therefore to degrees of entrenchment, entrenchment levels must be measured: if we were to account for a frequently found phenomenon through a high degree of entrenchment and then measure degree of entrenchment through frequency of occurrence that would be circular. One way to strengthen the claim would be to search for correlations between the use of an element in bilingual data and its frequency of occurrence in an independent referent corpus that may be seen as representative for the language use of the speaker (see again Hakimov's studies).

However, direct measurements of ease of activation, which can then be compared to corpus measures, yield more robust evidence for or against usage-based claims. One could argue, of course, that if something occurs in a corpus, it was, at that moment in time and for that particular speaker, the most easily activated unit. However, independent evidence can be found through carefully designed experiments.

Indirect evidence can be found through metalinguistic judgements about what is common, possible, or grammatical, or indeed by extrapolating from raw data on language use, but the preference seems to be, where possible, to use the more direct methods used in psychology and psycholinguistics. There has been a vibrant psycholinguistic approach to bilingualism for decades. Use of experimental methods has also been necessitated by other theoretical developments within contact linguistics, which has witnessed the natural shift from documenting phenomena to testing hypotheses about when a particular phenomenon will occur and when not. Testable hypotheses call for some degree of experimental control, for instance about the types of bilingual speakers and the types of language use investigated. This shift is visible in the development of the research tradition referred to as 'Heritage Languages,' in which the performance of 'Heritage Speakers' on particular linguistic tasks is compared to that of a monolingual control group, or different types of Heritage Speakers are compared to each other. This could be seen as a continuation of earlier work on generational differences in minority languages, in which different generations are often located at different stages of a language shift continuum. The increase in the degree of experimental control desired has led to a decreased reliance on spontaneously recorded conversation and increased use of the types of experimental methods employed in psycholinguistics, where they are used to tap into mental representation, and in applied linguistics, where they are used to compare the performance of different groups of learners on standardized tests. Methods frequently employed include controlled narratives (where participants for example relate what happens on pictures or videos presented to them), lexical decision tasks, elicited imitation experiments and many others.

6. Future directions

As this chapter has hopefully illustrated, the usage-based approach offers contact linguistics an innovative perspective on contact phenomena, which in turn provide usage-based linguistics with relevant data, especially on language change. A full usage-based account of language contact awaits, but requires further conceptual and methodological development. While several of the aforementioned works have developed the basic assumptions and principles that will underlie usage-based theorizing about language contact, many of these assumptions and the hypotheses they give rise to await substantial empirical exploration.

Conceptually, usage-based linguistics presents the study of language as the combination of sociolinguistics and psycholinguistics. Social and cognitive factors co-determine how we use language in an ever-turning cycle of usage and storage. What is activated gets used and what is used gets stored, and what is stored well stands a good chance of getting activated again. Having this as the common basis for all language use helps weaken the idea that bilingualism is ‘special.’ However, it is also clear that bilingualism and language contact do give rise to interesting phenomena, such as the need to separate languages and thus index words and structures as belonging to particular languages, the various ways of synchronically mixing such words and structures of various origins, and the diachronic types of change this inexorably leads to.

The emerging theoretical account of mixing and contact-induced change requires substantiation not adequately served by sticking only to the traditional research questions and methodological designs of contact linguistics. What will be needed is further articulation of relevant questions about language separation and fusion, about when people find language indexing relevant, about the degree to which foreign-origin units are entrenched and the implications of entrenchment, etc. This in turn requires further adaptation and development of corpus and experimental methods to the empirical realities of contact situations, in which basic requirements for such research, such as baseline data, ecological validity, and the large number of data points needed to reach sufficient statistical power, are not always easily fulfilled. The accumulating experience will no doubt be of great benefit to the development of linguistics in general.

7. Further reading

Backus, A. (2015). A usage-based approach to codeswitching: the need for reconciling structure and function. In: Stell, G. and Yakpo, K. eds., *Code-switching between structural and sociolinguistic perspectives (Lingua and litterae, Vol. 43)*, 1st ed. Berlin: De Gruyter Mouton, pp. 19–37.

This chapter contains a more elaborate introduction to the ideas presented in the current chapter about how well usage-based linguistics and contact linguistics go together, and what the implications are of their combination.

Bybee, J. (2010). *Language, usage and cognition*. Cambridge: Cambridge University Press.

This monograph is a relatively recent full articulation of the usage-based approach. It explains its foundations and presents empirical usage-based work on phonological, morphological, and syntactic topics.

Croft, William (2000). *Explaining language change: An evolutionary approach*. Harlow, England and New York: Longman.

This monograph is a book-length attempt to rethink state of the art knowledge on language change in a usage-based framework. Contact-induced change is placed within a broader framework of change in general.

Geeraerts, D., Kristiansen, G. and Peirsman, Y. (2010). *Advances in cognitive sociolinguistics*. Berlin and New York: De Gruyter Mouton.

This book contains chapters that combine the topics and methods of Labovian sociolinguistics and those of Cognitive Linguistics, providing many examples of the compatibility of these fields and the degree to which they benefit from each other's expertise.

Zenner, E., Backus, A. and Winter-Froemel, E., eds. (2019). Cognitive contact linguistics. Placing usage, meaning and mind at the core of contact-induced variation and change. *Cognitive linguistics research*, vol. 62, 1st ed. Berlin: De Gruyter Mouton.

This book brings together articles that use a usage-based framework to investigate familiar topics in language contact research.

8. Related topics

Processing multilingual data, language contact in the lab, variationist methods, borrowing, code-switching, convergence.

References

- Aaron, J.E. (2015). Lone English-origin nouns in Spanish: The precedence of community norms. *International Journal of Bilingualism*, 19(4), pp. 459–480.
- Adamou, E. (2016). *A corpus-driven approach to language contact: Endangered languages in a comparative perspective*. Boston and Berlin: De Gruyter Mouton.
- Backus, A. (1996). *Two in one. Bilingual speech of Turkish immigrants in the Netherlands*. Tilburg: Tilburg University Press.
- Backus, A. (2001). The role of semantic specificity in insertional codeswitching: Evidence from Dutch Turkish. In: R. Jakobson, ed., *Codeswitching worldwide II*, 1st ed. Berlin: De Gruyter Mouton, pp. 125–154.
- Backus, A. (2003). Units in codeswitching: Evidence for multimorphemic elements in the lexicon. *Linguistics*, 41(1), pp. 83–132.
- Backus, A. (2014a). Towards a usage-based account of language change: Implications of contact linguistics for linguistic theory. In: R. Nicolaï, ed., *Questioning language contact. Limits of contact, contact at its limits*, 1st ed. Leiden and Boston: Brill, pp. 91–118.
- Backus, A. (2014b). A usage-based approach to borrowability. In: E. Zenner and G. Kristiansen, eds., *New perspectives on lexical borrowing (Language contact and bilingualism 7)*, 1st ed. Berlin: De Gruyter Mouton, pp. 19–39.
- Backus, A. (2015). A usage-based approach to codeswitching: The need for reconciling structure and function. In: G. Stell and K. Yakpo, eds., *Code-switching between structural and sociolinguistic perspectives (Lingua and litterae, Vol. 43)*, 1st ed. Berlin: De Gruyter Mouton, pp. 19–37.
- Backus, A., Doğruöz, A.S. and Heine, B. (2011). Salient stages in contact-induced grammatical change: Evidence from synchronic vs. diachronic contact situations. *Language Sciences*, 33, pp. 738–752.
- Backus, A. and Dorleijn, M. (2009). Loan translations versus code-switching. In: B.E. Bullock and A.J. Toribio, eds., *The Cambridge handbook of linguistic code-switching*, 1st ed. Cambridge: Cambridge University Press, pp. 75–93.
- Broersma, M. and De Bot, C. (2006). Triggered codeswitching: A corpus-based evaluation of the original triggering hypothesis and a new alternative. *Bilingualism: Language and Cognition*, 9(1), pp. 1–13.
- Brown, E.L. (2015). The role of discourse context frequency in phonological variation: A usage-based approach to bilingual speech production. *International Journal of Bilingualism*, 19(4), pp. 387–406.
- Bybee, J. (2010). *Language, usage and cognition*. Cambridge: Cambridge University Press.
- Christiansen, M. and Chater, N. (2016). The now-or-never bottleneck: A fundamental constraint on language. *Behavioral and Brain Sciences*, 39, p. 62.
- Clark, A. (2013). Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences*, 36(3), pp. 181–204.

- Croft, W. (2000). *Explaining language change: An evolutionary approach*. Harlow, England and New York: Longman.
- Doğruöz, A.S. and Backus, A. (2009). Innovative constructions in Dutch Turkish: An assessment of on-going contact-induced change. *Bilingualism: Language and Cognition*, 12(1), pp. 41–63.
- Garcia, O. and Wei, L. (2014). *Translanguaging: Language, bilingualism and education*. London: Palgrave Macmillan.
- Gaskins, D., Backus, A. and Quick, A.E. (2019). Slot and frame schemas in the language of a Polish and English-speaking child: The impact of usage patterns on the switch placement. *Languages*, 4(8).
- Geeraerts, D. (2006). *Cognitive linguistics: Basic readings*. Berlin and New York: De Gruyter Mouton.
- Geeraerts, D., Kristiansen, G. and Peirsman, Y. (2010). *Advances in cognitive sociolinguistics*. Berlin and New York: De Gruyter Mouton.
- Goldberg, A.E. (2006). *Constructions at work: The nature of generalization in language*. Oxford: Oxford University Press.
- Gollan, T.H., Starr, J. and Ferreira, V.S. (2015). More than use it or lose it: The number-of-speakers effect on heritage language proficiency. *Psychonomic Bulletin and Review*, 22(1), pp. 147–155.
- Hakimov, N. (2015). *Explaining Russian-German code-mixing: A usage-based approach to code-mixing*. PhD. University of Freiburg.
- Hakimov, N. (2016a). Explaining variation in plural marking of German noun insertions in Russian sentences. In: H. Behrens and S. Pfänder, eds., *Experience counts: Frequency effects in language*, 1st ed. Berlin and Boston: De Gruyter, pp. 21–60.
- Hakimov, N. (2016b). Effects of frequency and word repetition on switch-placement: Evidence from Russian-German code-mixing. In: J.A. Robinson and M. Reif, eds., *Culture and cognition in bilingualism*, 1st ed. Berlin and Boston: De Gruyter, pp. 91–125.
- Heine, B. and Kuteva, T. (2005). *Language contact and grammatical change (Cambridge approaches to language contact)*. Cambridge: Cambridge University Press.
- Höder, S. (2014). Constructing diasystems: Grammatical organisation in bilingual groups. In: T. Afarli and B. Maehlum, eds., *The sociolinguistics of grammar (Studies in language companion series 154)*, 1st ed. Amsterdam: John Benjamins, pp. 137–152.
- Irizarri van Suchtelen, P. (2016). *Spanish as a heritage language in the Netherlands: A cognitive linguistic exploration*. Utrecht: LOT, Netherlands Graduate School.
- Kootstra, G.J., Van Hell, J.G. and Dijkstra, T. (2012). Priming of code-switching in sentences: The role of lexical repetition, cognates, and language proficiency. *Bilingualism: Language and Cognition*, 15, pp. 97–819.
- Kroll, J.F. and Bialystok, E. (2013). Understanding the consequences of bilingualism for language processing and cognition. *Journal of Cognitive Psychology*, 25, pp. 497–514.
- Langacker, R.W. (1987). *Foundations of cognitive grammar, Vol. I: Theoretical prerequisites*. Stanford, CA: Stanford University Press.
- Lieven, E., Salomo, D. and Tomasello, M. (2009). Two-year-old children's production of multiword utterances: A usage-based analysis. *Cognitive Linguistics*, 20(3), pp. 481–507.
- Marzo, S., Zenner, E. and Van de Mierop, D. (2019). When sociolinguistics and prototype analysis meet: The social meaning of sibilant palatalization in a Flemish urban vernacular. In: E. Zenner, A. Backus, and E. Winter-Froemel, eds., *Cognitive contact linguistics. Placing usage, meaning and mind at the core of contact-induced variation and change. Cognitive linguistics research*, vol. 62. Berlin: De Gruyter Mouton, pp. 127–156.
- Matras, Y. (2009). *Language contact*. Cambridge: Cambridge University Press.
- Matras, Y. and Sakel, J. (2007). Investigating the mechanisms of pattern replication in language convergence. *Studies in Language*, 31, pp. 829–865.
- Montrul, S. (2016). *The acquisition of heritage languages*. Cambridge: Cambridge University Press.
- Moro, F.R. (2016). *Dynamics of Ambon Malay: Comparing Ambon and the Netherlands*. Utrecht: LOT, Netherlands Graduate School.
- Muysken, P. (2000). *Bilingual speech. A typology of code-mixing*. Cambridge: Cambridge University Press.

- Onar Valk, P. (2015). *Transformation in Dutch Turkish subordination? Converging evidence of change regarding finiteness and word order in complex clauses*. PhD. Tilburg University.
- Pires, A. and Rothman, J. (2009). Disentangling sources of incomplete acquisition: An explanation for competence divergence across heritage grammars. *International Journal of Bilingualism*, 13(2), pp. 211–238.
- Poplack, S. and Dion, N. (2012). Myths and facts about loanword development. *Language Variation and Change*, 24, pp. 279–315.
- Quick, A.E., Backus, A. and Lieven, E. (2019). Partially schematic constructions as engines of development: Evidence from German-English bilingual acquisition. In: E. Zenner, A. Backus, and E. Winter-Froemel, eds., *Cognitive contact linguistics. Placing Usage, meaning and mind at the core of contact-induced variation and change. Cognitive linguistics research*, vol. 62, 1st ed. Berlin: De Gruyter Mouton, pp. 279–303.
- Quick, A.E., Hartmann, S., Backus, A. and Lieven, E. (2019, aop). Entrenchment and productivity: The role of input in the code-mixing of a German-English bilingual child. *Applied Linguistic Review 2019*.
- Quick, A.E., Lieven, E., Backus, A. and Tomasello, M. (2018b). Constructively combining languages: The use of code-mixing in German-English bilingual child language acquisition. *Linguistic Approaches to Bilingualism*, 8(3), pp. 393–409.
- Quick, A.E., Lieven, E., Carpenter, M. and Tomasello, M. (2018a). Identifying partially schematic units in the code-mixing of an English and German speaking child. *Linguistic Approaches to Bilingualism*, 8(4), pp. 477–501.
- Serigos, J. (2017). Using distributional semantics in loanword research: A concept-based approach to quantifying semantic specificity of anglicisms in Spanish. *International Journal of Bilingualism*, 21(5), pp. 521–540.
- Silva-Corvalán, C. (1994). *Language contact and change: Spanish in Los Angeles*. Oxford: Clarendon Press.
- Slobin, D. (1991). Learning to think for speaking: Homeland language, cognition, and rhetorical style. *Pragmatics*, 1, pp. 7–26.
- Thomason, S.G. and Kaufman, T. (1988). *Language contact, creolization, and genetic linguistics*. Berkeley, CA: University of California Press.
- Tomasello, M. (2003). *Constructing a language*. Cambridge, MA: Harvard University Press.
- Torres-Cacoullous, R. and Travis, C. (2018). *Bilingualism in the community. Code-switching and grammars in contact*. Cambridge: Cambridge University Press.
- Travis, C., Torres Cacoullous, R. and Kidd, E. (2017). Cross-language priming: A view from bilingual speech. *Bilingualism: Language and Cognition*, 20(2), pp. 283–298.
- Vaid, J. and Meuter, R.F.I. (2018). Languages without borders: Reframing the study of the bilingual mental lexicon. In: M. Libben, M. Goral, and G. Libben, eds., *Bilingualism. A framework for understanding the mental lexicon*, 1st ed. Amsterdam: John Benjamins, pp. 7–26.
- Verschik, A. (2008). *Emerging bilingualism: From monolingualism to code-copying*. London: Continuum.
- Verschik, A. (2019). English-Estonian code-copying in blogs: Combining a contact-linguistic and cognitive approach. In: E. Zenner, A. Backus, and E. Winter-Froemel, eds., *Cognitive contact linguistics. Placing usage, meaning and mind at the core of contact-induced variation and change. Cognitive linguistics research*, vol. 62, 1st ed. Berlin: De Gruyter Mouton, pp. 51–80.
- Wasserscheidt, P. (2016). Construction grammar and code-mixing. In: M. Reif and J.A. Robinson, eds., *Cognitive perspectives on bilingualism*, vol. 17. Berlin: Walter de Gruyter, pp. 65–90.
- Winter-Froemel, E. (2019). Reanalysis in language contact: Perceptive ambiguity, salience, and cat-achrestic reinterpretation. In: E. Zenner, A. Backus, and E. Winter-Froemel, eds., *Cognitive contact linguistics. Placing usage, meaning and mind at the core of contact-induced variation and change. Cognitive linguistics research*, vol. 62. Berlin: De Gruyter Mouton, pp. 81–126.
- Zenner, E., Backus, A. and Winter-Froemel, E., eds. (2019). *Cognitive contact linguistics. Placing usage, meaning and mind at the core of contact-induced variation and change. Cognitive linguistics research*, vol. 62. Berlin: De Gruyter Mouton.
- Zenner, E., Speelman, D. and Geeraerts, D. (2014). Core vocabulary, borrowability, and entrenchment: A usage-based onomasiological approach. *Diachronica*, 31(1), pp. 74–105.

Part 2

Processes and dimensions



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Social factors

Kofi Yakpo

1. Introduction

I use ‘social factors’ as a cover term for a set of interdependent characteristics relating to economic structure, political participation and social organization, demographic distribution and ideological superstructure within a society. Contact ecologies and outcomes may be construed as intersecting spaces along a continuum. On one end, economically and politically more egalitarian linguistic ecologies create the conditions for the early acquisition of multiple languages. Balanced child and adult multilingualism, gradual population growth and modest mobility favour language maintenance, long-term structural accommodation and cumulative change, and the emergence of linguistic areas. This may be supported by fluid ethnolinguistic identities, exogamy, and dense and multiplex social networks across linguistic boundaries. Such sprachbund-type outcomes have, for example, been studied in the Balkans (Kopitar, 1829; Sobolev, 2004), Vanuatu (François, 2011), the Ethiopian highlands (Hayward, 1991), the West African littoral zone (Ameke, 2005), Amazonia (Epps, 2006), and India (Masica, 1976).

On the other end, we find hierarchical and non-egalitarian ecologies with rather sharp ethnolinguistic boundaries, such as those created by European colonialism and other enterprises of conquest (e.g., the Han expansion in China, and the expansion of Arabic in Western Asia and North Africa). Concentrations of economic and political power in the hands of few, population displacement, genocide, and enslavement may engender skewed demographic distributions and lead to rapid and large-scale language acquisition by adults. These scenarios typically involve far-reaching and fast-paced linguistic change, often in the course of language shift, ushering in the emergence of heavily restructured and new languages. The European-lexifier creoles of the southern hemisphere, the Romance languages of Europe, and the Arabic varieties of the Maghreb represent some of the possible outcomes of such scenarios.

Other ecologies recombine characteristics of the social factors mentioned previously in various constellations, leading to a whole range of heterogeneous contact outcomes between these two idealized extremes, e.g., the emergence of urban youth languages in Africa and urban multi-ethnolects in Europe. In this chapter, I assess some of the social factors underpinning the processes and outcomes of language contact. I discuss explanations that have been

proposed to account for the differences between these outcomes. My own research strengths lie in the cultural and linguistic contacts between Africa and Europe in the age of European conquest and domination from the sixteenth century till today. Much of the following will therefore revolve around the outcomes of African-European language contact in the Atlantic basin, and specifically creolization. The various scenarios in which Afro-European language contact has taken place contain many of the ingredients that allow conclusions of a more general nature about the role of social factors in language contact.

2. Historical overview

Theories on the role of social factors in language contact and change began to flourish in early comparative historical linguistics. Hypotheses were advanced on how ‘invasions’ of ‘Indo-Europeans’ in Europe and South Asia ushered in the various sub-families of the Indo-European language family. In its earliest manifestations, the invasion paradigm postulates language contact as an epiphenomenon of demic diffusion alone, that is, the movement of ‘superior’ invading groups into the territory of ‘inferior’ groups. The invasion paradigm is intimately linked with nineteenth-century Social Darwinist thought, in which the status quo at the height of European imperialism and settler colonialism was projected in a recursive fashion on the historical relations between population groups and the languages they speak (see e.g., Aytürk, 2004, pp. 1–5 for an overview). This paradigm, together with the monoparental family tree model it beget, cast long shadows into the twentieth century, making it difficult to distinguish between scenarios where an element of demic diffusion is likely, as in the Bantu expansion (see de Filippo et al., 2012) and those where invasions as the source of language dispersal and differentiation are highly improbable, as with the ‘Indo-European invasions’ of India (Kivisild et al., 2003) and Europe (Demoule, 2014). In the intellectual climate of the time, the existence of European-lexifier creole languages was particularly unsettling because they proved difficult to align with the monoparental family tree model and showed an unexpected agency of the dominated in the restructuring of European languages. Only a few linguists at the time, among them Hugo Schuchardt (Schuchardt and Gilbert, 1980), could escape the nativist and racist European pseudoscience that proliferated around creoles and pidgins in particular (see Mufwene, this volume).

A new take on social dimensions began to emerge with the crystallization of language contact studies as an independent sub-discipline of linguistics. Psychological factors such as individual proficiency and language attitudes still dominate the enquiry of Weinreich (1953), but social factors such as relative group size, for example, are also taken into account. Van Coetsem’s (1988) theory of transmission in language contact includes both psycho-cognitive aspects in the notion of agentivity as well as social ones, in the notions of social dominance and prestige of a language. In Thomason and Kaufman’s (1988) typology of contact, the pendulum swings to the role of social factors in language contact, indicative of a maturation of the various subfields of contact linguistics. One field differentiated almost entirely into the psychological, and with the development of new technologies, into the neurological realms without much interaction with the social field, as noted by Stell and Yakpo (2015) for the study of code-switching. Studies like those of van Coetsem, and Thomason and Kaufman however remained vague on the exact nature of the social factors mediating contact processes and outcomes. The five levels of intensity of contact proposed by Thomason and Kaufman (1988) of (1) ‘casual,’ (2) ‘slightly more intense,’ (3) ‘more intense,’ (4) ‘strong cultural pressure,’ and (5) ‘very strong cultural pressure’ implicitly suggest a continuum ranging from more symmetrical to more asymmetrical power relations between speaker populations in

contact but the authors fail to explore this more systematically. In a similar vein, the analysis of social factors appears circular when degrees of ‘intensity of contact’ serve to characterize contact outcomes.

The study of the grammatical and typological ramifications of language contact has been the most richly differentiated field, with a large number of significant publications on contact hotspots and areal linguistics in specific world regions like Amazonia (e.g., Aikhenvald, 2004), Africa (Güldemann, 2018 and the sources contained there), and South Asia (Masica, 1976), or on specific processes such as borrowing (e.g., Matras and Sakel, 2007) and specific functional domains, such as morphology (Vanhove et al., 2012). Such works contain valuable information on social context even where they do not develop broader social typologies of contact. The convergence of unrelated languages observed in linguistic areas struck linguists as peculiar, further challenged the family tree model, and required some attention to the social backdrop to this kind of language change (e.g., in Gumperz and Wilson, 1971). Muysken (2008b, pp. 11–20) mentions migrations, trade, political transformations, European colonial expansion and slave trading, demographic restructuring, and urbanization as social factors in the emergence of a broadly defined Atlantic linguistic area encompassing Europe, the Americas, and West Africa. Aikhenvald and Dixon (2007, pp. 37–42) list a mix of factors on different levels of aggregation (e.g., size of speaker group vs. mythology), some of which are more psychosocial (language attitudes), others more socio-structural (e.g., types of subsistence, division of labour between the sexes). With other factors again, there is the recurrent tendency toward circularity when contact factors (degree of knowledge of each other’s language and balanced or displacive contact) are adduced to explain patterns of contact. The ‘lifestyles’ listed (hunter-gatherers, village-dwelling agriculturalists, nomadic cattle herders, or largely urbanized groups) also indicate that Aikhenvald’s sociocultural parameters are historically oriented and geared towards classifying contact processes and outcomes in small-scale societies.

From earliest manifestations in comparative historical linguistics, most appraisals of social factors therefore do not define causalities, micro-, meso-, and macro-features and parameters and their relation to each other. *En gros*, they fail to provide a socio-structural analysis of *social systems* in determining contact outcomes. We are therefore still waiting for a more systematic appreciation of social factors in language contact, and here more cross-fertilization with anthropology, sociology, and political science and economics would be useful. Areal linguistics has shown that egalitarian relations and exchange between linguistic groups leading to long-drawn and gradual diffusion of linguistic features are important in many language contact scenarios (but certainly not all, see Muysken’s Atlantic linguistic area), thus anticipating a later line of enquiry into small-scale multilingualism (see Section 3). In creole linguistics, the investigation of social factors in language contact has been, conversely, oriented towards the inegalitarian face of language contact, since most research in this area has been conducted on languages that emerged during the European colonization of the Americas, Africa, and the Pacific. It is this attention to demography and power dynamics that has allowed creolists to develop some of the more sophisticated social explanations for language contact and change compared to other branches of linguistics.

In the following sections, I will therefore be mainly concerned with the role of the macro-level social factors of demography and socio-economic structure in the processes and outcomes of language contact. In doing so, I carve out a space of enquiry separate from the occupation of variationist sociolinguistics with social variables like gender, social class or ethnicity at the (individual) micro-, and (small group) meso-levels (Labov, 2001) and the social

meaning of such variables (Eckert, 2008). I also pass by the interactional micro-level proper to research on social factors in pragmatics (Auer, this volume).

3. Critical issues and topics

I now focus on demography and socio-economic structure, two social factors that feature in various manifestations in the literature, and which have been deemed central for determining language contact outcomes. These outcomes are discussed further in section 4 in relation to current research in the field.

DEMOGRAPHY. Fluctuations in the size and composition of populations, and migrations play an important role in explaining social, cultural and economic change in the social sciences at large (see Morland, 2019 for an overview). Demography has also been a key concern of theories attempting to explain systematic differences in grammatical structure between languages, and to address the question of structural complexity. In an influential work, Thurston (1987) claims that languages spoken in exoteric (hence outwardly oriented) societies tend to be more analytic, and therefore adapted to L2 learning by adults. Languages spoken in esoteric (inwardly oriented) societies, on the other hand, are adapted to L1 learning by children, and therefore tend to be more synthetic. Building on such earlier work, Kusters (2003) shows that languages serving as *lingua francas*, and have therefore integrated large numbers of L2 speakers in the course of their history, tend to lose inflectional and derivational morphology. For various neuro-cognitive reasons, adult speakers, as opposed to children, are said to find the learning of elaborate morphology difficult (see e.g., Klein and Perdue, 1997), and therefore tend to substitute synthetic (morphological) structures with analytic (syntactic) ones. Lupyán and Dale (2010) baptize these correlations the ‘linguistic niche hypothesis’ and seek to lend quantitative support to them. Both the statistics and the cognitive-neurological assumptions of the ‘linguistic niche hypothesis’ have been criticized (e.g., Koplenig, 2019) but such work broadens the empirical and theoretical foundations for studying the relation between language contact and change, and social structure.

Demography and language contact and change have been particularly close bed mates in creole linguistics, maybe because creolists have access to European historical records that contact scholars working in other regions often do not (e.g. Holm 2014). Demographic research on the provenance of the African and European populations during European colonialism in countries like Haiti (Singler, 1995) and Suriname (Arends, 1995) has invigorated the study of the detailed contributions of substrate (e.g., Lefebvre, 1998; Migge, 1998), superstrate (e.g., Chaudenson, 1992; Mufwene, 1996; Blasi, Michaelis and Haspelmath, 2017) and areal features (see Muysken and Smith, 2015 for a collection of recent studies; also Yakpo, 2017c, 2017d) to Afro-European creoles, and has also played a role in debates about the complexity of creoles.

SOCIO-ECONOMIC STRUCTURE. While the links between demography and language structure have been made more directly, those between socio-economic and linguistic structure have been raised in more diffuse ways because they are more multi-layered in nature. Trudgill (2011) addresses social factors in addition to relative population sizes by turning to the nature of social networks in small vs. large linguistic groups. Face-to-face societies, typically pre-colonial hunter-gatherer, small-scale agriculturalist, and other decentralized societies organized in dense and multiplex social networks are believed to have morphologically more complex synthetic languages. In contrast, larger centralized societies characterized by overarching political structures and ‘loose networks’ tend to produce languages with more analytic structures. Trudgill’s (2011) sociolinguistic typology seeks to provide a dynamic analysis of the social determinants of linguistic structure, and in doing so, focuses on the causes of

language change. He identifies five social factors that determine the outcomes of language change with respect to ‘complexification’ and ‘simplification.’ These are group size, density of social networks, amounts of shared information, degree of social stability, and degree of contact. Like Thomason and Kaufman before, Trudgill’s typology of social factors is circular, when linguistic factors (e.g., ‘much contact’) are adduced to explain linguistic outcomes (e.g., ‘simplification’). The model is also patchy in picking out certain factors (e.g., ‘network density’) without elaborating on socio-economic context (dense societies can be egalitarian or hierarchical, for example). Very few sources explore how group size and density of social networks form part of the economic base of a society, relate to its social stratification, and may engender and be shaped by a specific ideological superstructure (e.g., Adamou, 2016, pp. 184–210). Yet these social macro-factors probably have the most profound impact on the processes and outcomes of language contact.

High levels of socio-economic stratification are linked to political centralization and the presence of states, elite capture and the extraction of economic resources from commoners through various forms of (bonded and forced) labour, and social endogamy (Johnson and Earle, 2000). The centralization of socio-economic power often takes place when elites and their auxiliaries conquer and colonize other polities (the ‘invasion’ model referred to in Section 2). Conquest and colonization ‘punctuate’ an existing ‘equilibrium’ and can lead to extensive language contact. (Thurston, 1987, p. 40). The long-term contact outcome typically is language shift, both from dialects and languages within the new polity to the superstrate of the elites, and hence an eventual decrease in linguistic diversity (but see the cases covered under ‘linguistic nationalism’ and ‘grassroots heteroglossia’ that follow).

Political centralization and socio-economic stratification are, in turn, intimately linked to ‘language rationalization’ (Laitin, 1992). Since the emergence of the nation-state in Western Europe, and its imposition on the rest of the world during European colonialism, linguistic nationalism (Blommaert and Verschueren, 1998) and the invention and engineering of standard languages have become the norm (Makoni and Pennycook, 2007); I brush aside the problematic ideological ramifications of the term ‘standard.’ Much remains to be discovered about how fast languages change under differing social conditions (cf. Bakker et al., 2017, pp. 135–136), and how innovations percolate through social networks (cf. Yu, 2013), be they exposed to heavy cross-typological contact, and serve as vehicular languages in multilingual ecologies, or not. When elites codify their lect as part of a (sub-)nationalistic enterprise this may lead to its structural arrest until the next punctuation event. Despite large-scale adult immigration into all Western European countries from Eastern Europe and the Global South, natural contact-induced change including morphological levelling has largely stalled in the European standard languages. The intellectual elites of socially stratified nation-states may also seek to expunge traces of contact from the standard, for example in the ill-fated attempt to engineer Katharevousa, a standard Greek language purified of Turkish and other ‘foreign’ influences (Demoule, 2014, p. 585), and the elimination of Arabo-Persian lexicon in the creation of Standard Hindi (Khan, 2006, pp. 225–250).

As a result of language rationalization, a temporary equilibrium may emerge with an engineered and fixed standard language spoken natively by political and economic elites, and to varying degrees by other sections of a polity’s population in accordance with a tangled amalgam of factors like class, ethnicity, gender, geographic location, and access to educational, economic, and sociocultural resources. Other lects spoken in the polity may continue coexisting with the standard superstrate for varying periods of time, leading to situations known in the literature under processes and outcomes such as diglossia (Ferguson, 1959), the creole continuum (DeCamp, 1971; Rickford, 1987), dialect levelling, and diaglossia (Bellmann, 2009).

More recent research has attenuated the eschatological premise of these approaches that convergence with the standard inescapably ushers in the elimination of non-standardized varieties and languages. As sites of individual agency and linguistic creativity, the heteroglossic practices embodied in non-standardized varieties can have a significant impact on standardized ones. This is shown, for example, by the far-reaching influence of African-American Vernacular English on Standard American English (e.g., Paris, 2009) and the role of non-standardized varieties in nourishing a global pop culture in places as far apart as Kinshasa and Berlin (see Nassenstein, Hollington and Storch, 2018). Speakers of non-standard varieties can also contest and subvert the national standard, and develop their own regional standards (e.g., Hinskens, 2009 for a regiolect of Dutch).

In spite of an overall tendency of non-standard lects to become more similar to superstrate standard lects, such ‘advergence’ (Mattheier, 1996) is therefore not monolithic and regularly paced, but fractured, decentred, and halting.

4. Current contributions and research

GRASSROOTS HETEROGLOSSIA. Recent research has highlighted the decentred evolution and fragmentation that exists beside the linguistic homogenization tendencies inherent to globalization. In Africa South of the Sahara and the Caribbean, for example, diffuse and non-concerted language policies by weak postcolonial states and the failure of education dispensed in colonial languages has led to the appropriation and transformation of colonial languages by common people through grassroots heteroglossia (e.g., for Dutch in Suriname, see Muysken, 2017). This has effectively put a hold on the spread of European standard languages and the monolingual ideologies driving their expansion beyond an administrative-formal domain controlled by small elites (Mufwene and Vigouroux, 2008).

Often, such heteroglossic patterns in postcolonial nation-states are modelled on the ancient experiences of people practising subsistence economies in egalitarian small-scale societies (as in Vanuatu, see François, 2012), but sometimes they are also more recent adaptive responses of small-scale societies to external threats (e.g., in the Casamance, see Lüpke, 2016). While this leads to morphosyntactic assimilation of the participating languages, a complex array of socio-cultural (François, 2011) and cognitive factors (Ellison and Miceli, 2017) may simultaneously result in lexical and phonetic dissimilation. Maintaining and reproducing ethno-linguistic identities and boundaries, in spite of linguistic and cultural convergence in a region, can be crucial when a group seeks to secure its access to, and control over land and fishing grounds, trading networks, jobs, political power, and knowledge systems (e.g., Paulston and Paulston, 1980; the studies in Brooks, 1993; Michael, 2008; Rüsçh and Nassenstein, 2016).

In the mega-cities of the Global South, small languages and lingua francas, creoles, pidgins, youth sociolects, cryptolects and other paradigm-defying means of communication are indispensable tools for securing livelihoods through the mobilization of local, regional, professional, and ethnic networks. Vast linguistically non-normalized spaces therefore remain in Africa, Amazonia and the Caribbean, South and South East Asia, Oceania, even in the heart of Europe (e.g., the case of Luxembourgish, see Gilles, 2009).

EXPANSIVE LINGUA FRANCAS. At least one other sociolinguistic scenario needs to be considered that falls between the extremes of language standardization in nation-states and grassroots heteroglossia. In various parts of the globe, we witness the expansion of unstandardized or weakly standardized lingua francas spoken by tens of millions of people and upwards that thrive and grow without top-down standardization and state enforcement. These languages

of wider communication are contact languages par excellence. They are generally spoken by fewer L1 than L2 speakers, are acquired outside of the parental fold by many, and often later in childhood or adolescence, and serve as outgroup languages spoken in various multilingual constellations throughout the day, include intense code-switching and show contact effects in various subsystems. This is particularly the case in Africa, with its generally weak penetration of standard languages. Some of the largest African lingua francas are the continua of non-coastal varieties of Kiswahili in Eastern-Central Africa, Lingala in the DRC and the Republic of the Congo, Mande, Hausa and Arabic in West and North-East Africa.

A striking case is the continuum of the African English-lexifier contact languages (AECs) spoken along the West African littoral in mutually intelligible varieties known in the literature as Nigerian Pidgin, Cameroon Pidgin, Ghanaian Pidgin English, Krio, and Pichi. From humble beginnings in the nineteenth century, their speaker numbers have exploded in the past few decades reaching over one hundred million today (Yakpo, 2016a), and are likely to grow to several hundred million speakers in the next few decades (Yakpo, 2017b). The growth of these languages is driven by a recurring constellation of social factors linked with ‘modernization’: accelerated demographic growth, rural-urban migration and galloping urbanization, the expansion of basic education, the degradation of rural economies and changes in employment structure, the expansion of transport, telecommunications, and the digital media revolution (see Yakpo, 2015, pp. 265–266, for a study of a similar set of factors in Suriname). As a consequence of these dynamics, AECs have penetrated all but the most formal domains (i.e., government, and other highly asymmetrical spaces of interaction) in all countries where they are spoken (Yakpo, 2016b). AECs dominate the pop music soundscapes of Nigeria, Ghana, Sierra Leone, and ‘anglophone’ Cameroon. Nigerian Pidgin is deployed in patterns of plurilingual communication together with Nigerian English in movies, TV shows, comedy shows, and on YouTube channels in the burgeoning media industry of Nigeria. Due to their role as means of horizontal communication, languages like Nigerian and Cameroon Pidgin are not claimed by specific social formations, e.g., ‘an ethnic group,’ are rarely identified as ‘mother tongues’ by their speakers, and speech practices are not as embedded in ‘traditional’ social structures as other African languages. A shared lexicon, rampant code-switching, and lexical and structural borrowing between English and the AECs challenge monoglossic notions of separate languages. There is also some evidence that these lingua francas change faster than expected (for Ghanaian Pidgin English, see Corum, 2015; for Sranan (Suriname), see Yakpo, 2017a).

SOCIAL FACTORS IN CREOLIZATION OUTCOMES. Creole studies is the field of linguistics in which the most comprehensive suggestions for the role of social factors in language change have been proposed. The simplicity of creole languages vis-à-vis non-creoles is a leitmotif that has found its way into textbooks far beyond linguistics. It assumes a ‘terra nullius’ scenario, in which subjugated populations could or would not acquire the European lexifier language, and therefore contracted it to the rudimentary system of a pidgin, later expanding it to a creole in the course of inter-generational transmission (most radically formulated by Bickerton, 1981 and subsequent work) (but see Mufwene, this volume, for an opposing view of the pidgin-to-creole cycle). The focus on (morphological) form in discussions about complexity in the outcomes of language contact has been criticized from many angles, for example by syntacticians for its privileging of ‘surface’ complexity (e.g., Aboh, 2015), its disregard for semantic complexification accompanying morphological simplification (e.g., Koplenig, 2019, also see Bisang 2015), its indifference to the complexity of plurilingual grammars (Yakpo, 2009), and for a superficial view on the analytic typology of their substrates (Szeto, Lai and Ansaldo, 2019).

An important strain of newer historically oriented work has focused on the agency of enslaved and marginalized peoples to subvert, refashion and co-determine socio-economic, and concomitantly, linguistic structures even under conditions of extreme oppression (Faraclas, 2012; Linebaugh and Rediker, 2013). Faraclas et al. (2007) expand earlier work by (Alleyne, 1971, 1980; Mintz, 1971) to show how different modes of economic production, differing ideologies about religion, ethnicity and race, and divergent legal-political dispensations decisively shaped the outcomes of Afro-European language contact in the European colonies of the Americas. The meagre opportunities for socio-economic advancement that ‘Northern’ European colonial production regimes (British, Dutch, and partly French) offered to enslaved Africans, versus the comparatively better opportunities offered by the ‘Southern’ European (Portuguese and Spanish) models engendered very different Afro-European language contact outcomes (e.g., Brazil vs. the USA, see Carvalho, 2018; Klein, 2018; Mattos and Grinberg, 2018). As a consequence, African substrate features are pervasive in the English-lexifier creoles, intermediate in the French-lexifier creoles, and comparatively few in the American vernacular varieties of Spanish and Portuguese.

The political economic perspective on the outcomes of language contact and change can be combined with a typological approach that embeds the evolution of creoles and other contact languages in their areal context (Yakpo, 2017c). This can be exemplified with respect to the fate of prosody in Afro-European contact languages in Africa and the Americas (i.e., creoles, pidgins, and the transplanted colonial varieties of European languages that emerged through the interaction of Africans and Europeans in the Atlantic area). These languages feature prosodic systems ranging from tone to stress to mixed systems incorporating both. In the literature, tone is often seen as a crosslinguistically marked feature that either gets lost or is reduced in language contact (Salmons, 1992; McWhorter, 1998; Trudgill, 2010). A systematic look at a larger range of prosodic contact constellations, however, shows no evidence that stress invariably trumps tone (Bordal Steien and Yakpo, 2020). Instead we find a continuum of prosodic systems stretching across the Atlantic basin from Africa to the Americas with outcomes determined by degrees of socio-economic stratification (in this case along racialized lines), and degrees of areal-typological dominance of tone languages in the ecology (Lim, 2009).

One group contains the Afro-European creoles of the West African littoral zone as well as the European colonial colloquial varieties spoken in the tonal ecologies of Africa. These all feature lexical tone systems. Isolated Afro-European contact varieties spoken in the Americas also feature tone systems (Saamaka, Ndyuka, Paamaka, Aluku, Kwinti). A second group consists of the European-lexifier creoles of the Caribbean and many of the colloquial varieties of the European colonial languages of the wider region, including South America. Many of these contact languages have ‘residual’ tone systems (Berry, 1972), where distinctive tone is limited to specific semantic fields and subsystems, and to specialized functions, in prosodic systems otherwise characterized by the use of stress. The English-lexifier Creoles of the Caribbean, for example, have been described to possess grammatically distinctive tone with person forms (see e.g., James, 2003 for Tobagonian Creole), in ideophones and in reduplications (see Smith and Adamson, 2006 for Sranan).

A third group consists of languages with stress systems that nevertheless seem to point to some substratal transfer from African tone languages in their evolutionary trajectory. Brazilian Portuguese features ‘tonal events not linked to stressed syllables’ which are ‘reminiscent of the intonational characteristics of languages like Japanese and Korean’ (Frota and Vigário, 2000). In Afro-Bolivian Spanish (Rao and Sessarego, 2016),

Afro-Peruvian Chincha Spanish (Sessarego, Rao and Butera, 2017) and the Spanish-lexifier creole Palenquero, spoken in Colombia, (Hualde and Schwegler, 2008), ‘accented’ syllables consistently carry high (contour) pitch. This is quite different from European and Euro-American varieties, where word-level pitch contours can be significantly altered by the pragmatic functions of intonation (see Hualde and Prieto, 2015 for an overview). The first holistic take on the evolution of the prosodic systems of Afro-European contact languages on the basis of both social and linguistic factors is presented in Bordial Steien and Yakpo, 2020.

With a refinement of methods, and after making room for the countless variations in these scenarios, approaches based on demography and social stratification including factors at a lower level of aggregation (e.g., network density, exogamy vs. endogamy, cultural, identity, and gender features) may explain differences in outcomes of other large linguistic continua with a history of intense contact (e.g., the Romani continuum, Sinitic, the northern Indian continuum of languages, Arabic, and others).

5. Main research methods

Much twentieth-century work on the social context of language contact focused on colonial ecologies in the Americas and in the Asia-Pacific region, and therefore drew on methods of historical research next to structural linguistics. Such work had important ramifications, for example, for our understanding of the linguistic demography in the early colonial Caribbean (e.g., Singler, 1995), thereby calling into question the social foundation of the pidgin-to-creole cycle in the region and the idea of ‘abrupt’ creolization (Arends, 1995), instead establishing, for example, the existence of ‘ordinary’ genealogical links between the Caribbean (e.g., Baker, 1999) and West African English-lexifier creoles (Huber, 1999). The reconstitution of ‘histories from below’ through the use of neglected archival sources and oral histories reveals the complex socio-cultural, economic, and linguistic settings in societies featuring a diverse range of populations that shaped contact languages that arose in the Atlantic (e.g., Walicek, 2009) and Pacific (e.g., Drechsel, 2014) regions.

In the last two decades or so, the production of numerous grammars with the toolbox of descriptive linguistics and linguistic typology of languages profoundly marked by contact beyond the traditional purview of pidgin and creole studies is revealing the great diversity of contact ecologies and contact outcomes (e.g., Nordhoff, 2009; Litamahuputty, 2012; Souag, 2014). Such studies also seem to suggest that contact languages, innovative as they often are, largely reflect areal and genealogical proclivities in their structures rather than constituting a typological class onto themselves (cf. van Sluijs, van den Berg and Muysken, 2016; Blasi, Michaelis and Haspelmath, 2017).

Beyond, that social network analysis and anthropological methods of participant observation have provided insights into the social context, and the spread of contact-induced features at the level of individual communities and social groups (e.g., Beyer and Schreiber, 2013).

Significant changes in the study of language contact have also come, as in other sciences, from the integration of core linguistics with advanced statistics, cognitive and brain sciences. Equally, advances in natural language processing are allowing increasingly sophisticated mining and analysis of the gargantuan amounts of linguistic data created through digital and social media platforms and applications (also see Section 6).

6. Future directions

The main challenge of future research is to consolidate and refine an emerging multifactorial methodology to investigate the social factors that co-determine the processes and outcomes of contact. A key goal is to understand how micro-, meso-, and macro-social factors interact with each other, and with respect to different structural-linguistic features (see e.g., Sinnemäki and Di Garbo, 2018). The current trend to model the impact of social (and environmental) factors on language contact and change with the help of data driven sciences will continue. Works that integrate computational linguistics and data visualization (e.g., Hammarström and Güldemann, 2014), genetics (e.g., Pakendorf, de Filippo and Bostoen, 2011), GIS (e.g., Eriksen, 2011), anthropology and archaeology (e.g., O'Connor and Muysken, 2014), have allowed the testing of correlations between linguistic, social, and environmental factors at increasing levels of refinement, and contact-induced change and areal convergence will remain at the core of much of such work.

Linguistic data sciences are also being harnessed to set up and mine an increasing number of uncurated (e.g., Heyd and Mair, 2014) and curated corpora of contact languages (e.g., for Cameroon Pidgin English, see Green, Ayafor and Ozón, 2016; for Nigerian Pidgin, see Caron, 2019). Structured data banks like APiCs (Michaelis et al., 2013), eWave (Kortmann and Lunkenheimer, 2011) and other corpora in tandem with careful analysis of the socio-political, economic, and cultural context of a particular contact ecology can also be harnessed to uncover correlations between linguistic structure and social factors (e.g., Adamou and Granqvist, 2015). Studies informed by computational methods will allow new insights into variation, register, and style (e.g., Manfredi et al., 2019). The expansion of online media and texting is driving the emergence of vernacular orthographies and the development of formal registers in contact and minority languages alike (e.g., Eberhard, 2019), and is creating a widening gap between on-the-ground facts of language use and rigid official language policies (see e.g., the role of the BBC Pidgin Service in West Africa, Kasraee, 2017).

New developments can also be expected in qualitatively oriented research. Most contact languages studied are located in the Global South, in regions once colonized by Europe, or are spoken in the Northern hemisphere by marginalized populations with cultural and historical ties with the Global South. Yet, the vast majority of contact research is still conducted by Northern researchers. This raises problems with respect to research agendas, not only because representation is a prerogative of social justice, but also with respect to the value of resulting research. The absence of deeper intuitions about the languages they study and the often rudimentary biographic experience of multilingualism by researchers from countries with monoglossic traditions may result in long learning curves and significant gaps in findings (cf. Chomsky, 1997); the institutional self-containment of research in the North limits the impact on policy-making in Global South countries in dire need of reform of the destructive language regimes inherited from colonialism (see e.g., Wa Thiong'o, 1992; Muthwii and Kioko, 2004); the lack of a personal stake and an apolitical stance may lead Northern researchers to unwittingly perpetuate colonialist and racist tropes, as shown by DeGraff (2005) with respect to the epistemic violence (Chakravorty, 1999) inherent to the argument on creole simplicity.

The global expansion of tertiary education and increasing academic mobility will, however, continue to render a growing amount of research on contact languages by 'bearers of the corresponding cultures' (Huttar, 2019, p. 178), linguistically and culturally at ease in the communities they work with, and whose research often explicitly or implicitly includes decolonial approaches and social activist agendas (e.g., Carpenter and Devonish, 2010). Such work has, for example, engendered more nuanced, insider views of the links between social forces and

linguistic structure in the emergence of Afro-European contact languages, for example with respect to morphology in Haitian Creole (DeGraff, 2001), complex patterns of suppletion in the African English-lexifier creoles (Yakpo, 2019), the subtle interplay between substrate and superstrate influences in creole syntax (Aboh, 2015), and the identity factors driving grammatical change in Ghanaian Pidgin English (Osei-Tutu, 2016).

Dramatic socio-economic transformations and far-reaching changes in the demographic distribution of the world's population are likely to cause further reorientations in research priorities. By the end of the twenty-first century, the population of Africa will have risen to more than four billion, and the populations of Africa and Asia will together constitute about 90% of the world's population (United Nations, 2017). This will go hand in hand with rapid urbanization, and an acceleration of regional and international migration out of, into, and within Africa and Asia, and the growth of a young and mobile population numbering billions. The self-authored linguistic evolution of expansive lingua francas numbering hundreds of millions of speakers like West African Pidgin English, Hausa, Swahili, Lingala, Arabic, Hindi-Urdu, Malay-Indonesian, and regional varieties of English, French, Portuguese, and Spanish will bring crucial insights into the nature of linguistic accommodation and change, and the differing roles in these processes of individual actors and social networks, institutions, and the media in a more globalized, and simultaneously more decentred world. These trends are going to fundamentally shake up established views and the fairly limited understanding we still have of how social factors shape language contact and change.

7. Further reading

Thurston, W. (1987). *Processes of change in the languages of North-Western New Britain (Pacific Linguistics: Series B 99)*. Canberra: Research School of Pacific and Asian Studies, Australian National University.

This pioneering work on language contact in Papua New Guinea proposes a link between the degree of rich morphology and closedness 'esoteric' (L1-oriented) vs. 'exoteric' (L2-oriented) functions of linguistic systems, foreshadowing later work by Kusters, 2003 on the development of morphology in large lingua francas, and on the 'linguistic niche hypothesis' by Lupyán and Dale, 2010.

Faraclas, N., Walicek, D.E., Alleyne, M., Geigel, W. and Ortiz, L. (2007). The complexity that really matters: the role of political economy in creole genesis. In: U. Ansaldo, S. Matthews, and L. Lim, eds., *Deconstructing creole (Typological Studies in Language 73)*, 1st ed. Amsterdam: John Benjamins, pp. 227–264.

Expanding earlier work by Alleyne (1971) and Mintz (1971), this book chapter offers a refreshingly holistic take on the outcomes of language contact and creolization in the Caribbean due to differing economic structures, political systems, and ideological superstructures practised by European colonizers on the one hand, and African colonized/enslaved peoples on the other.

Aboh, E. O. (2015). *The emergence of hybrid grammars*. Cambridge: Cambridge University Press.

Taking earlier work on linguistic ecologies further (Mufwene, 2001; Ansaldo, 2009), Aboh presents ample socio-historical and linguistic data to buttress his analysis of the structure of contact languages, arguing that hybridization, rather than simplification, drive their evolution in ways no different from languages that emerge in more monolingual environments.

Bordal Steien, G. and Yakpo, K. (Forthcoming). Romancing with tone: on the outcomes of prosodic contact. *Language*, 96(1).

Arguing against claims that tone necessarily cedes to stress during language contact and creolization (Salmons, 1992; Trudgill, 2010), this study of contact between African tone and European intonation-only languages suggests that certain social (linguistic demography and socio-economic stratification), and linguistic factors (areal-typological dominance) are pivotal for explaining the presence of tone

systems in almost all Afro-European contact languages including creoles and African varieties of French, Spanish, and English.

8. Related topics

Pragmatic factors, typological factors, borrowing, convergence, creoles and pidgins, urban youth speech styles, West Africa

Abbreviations

- AEC African English-lexifier contact languages
L1 first language
L2 second language

References

- Aboh, E.O. (2015). *The emergence of hybrid grammars*. Cambridge: Cambridge University Press.
- Adamou, E. (2016). *A corpus-driven approach to language contact: Endangered languages in a comparative perspective*. Boston and Berlin: De Gruyter Mouton.
- Adamou, E. and Granqvist, K. (2015). Unevenly mixed Romani languages. *International Journal of Bilingualism*, 19(5), pp. 525–547.
- Aikhenvald, A.Y. (2004). *Language contact in Amazonia*. Oxford: Oxford University Press.
- Aikhenvald, A.Y. and Dixon, R.M.W., eds. (2007). *Grammars in contact: A cross-linguistic typology*. Oxford: Oxford University Press.
- Alleyne, M.C. (1971). Acculturation and the cultural matrix of creolization. In: D. Hymes, ed., *Pidginization and creolization of languages*, 1st ed. Cambridge: Cambridge University Press, pp. 169–186.
- Alleyne, M.C. (1980). *Comparative Afro-American: An historical-comparative study of English-based Afro-American dialects of the new world*. Ann Arbor: Karoma Publishers.
- Ameka, F.K. (2005). Multiverb constructions on the West African littoral: Microvariation and areal typology. In: M. Dimitrova-Vulchanova and T.A. Åfarli, eds., *Grammar and beyond: Essays in honour of Lars Hellan*, 1st ed. Oslo: Novus, pp. 15–42.
- Ansaldò, U. (2009). *Contact languages: Ecology and evolution in Asia*. Cambridge: Cambridge University Press.
- Arends, J. (1995). Demographic factors in the formation of Sranan. In: J. Arends, ed., *The early stages of creolization*, 1st ed. Amsterdam: John Benjamins (Creole Language Library (CLL), 13), pp. 233–285.
- Aytürk, İ. (2004). Turkish linguists against the West: The origins of linguistic nationalism in Atatürk's Turkey. *Middle Eastern Studies*, 40(6), pp. 1–25.
- Baker, P. (1999). Investigating the origin and diffusion of shared features among the Atlantic English Creoles. In: P. Baker and A. Bruyn, eds., *St. Kitts and the Atlantic creoles: The texts of Samuel Augustus Mathews in perspective*, 1st ed. London: University of Westminster Press, pp. 315–364.
- Bakker, P. et al. (2017). *Creole studies – phylogenetic approaches*. Amsterdam: John Benjamins.
- Bellmann, G. (2009). Between base dialect and standard language. *Folia Linguistica*, 32(1–2), pp. 23–34.
- Berry, J. (1972). Some observations on “residual tone” in West Indian English. In: *Conference on creole languages and educational development*. St. Augustine, Trinidad.
- Beyer, K. and Schreiber, H. (2013). Intermingling speech groups: Morpho-syntactic outcomes of language contact. In: I. Léglise and C. Chamoreau, eds., *The interplay of variation and change in contact settings*, 1st ed. Amsterdam: John Benjamins, pp. 107–134.
- Bickerton, D. (1981). *Roots of language*. Ann Arbor: Karoma Publishers.

- Bisang, Walter. 2015. Hidden complexity – The neglected side of complexity and its implications. *Linguistics Vanguard* 1(1). <https://doi.org/10.1515/lingvan-2014-1014>.
- Blasi, D.E., Michaelis, S.M. and Haspelmath, M. (2017). Grammars are robustly transmitted even during the emergence of creole languages. *Nature Human Behaviour*, (1), pp. 723–729.
- Blommaert, J. and Verschueren, J. (1998). The role of language in European nationalist ideologies. In: B.B. Schieffelin, K.A. Woolard, and P.V. Kroskrity, eds., *Language ideologies: Practice and theory*, 1st ed. New York: Oxford University Press, pp. 189–210.
- Bordal Steien, G. and Yakpo, K. (2020). Romancing with tone: On the outcomes of prosodic contact. *Language*, 96(1).
- Brooks, G.E. (1993). *Landlords and strangers: ecology, society, and trade in Western Africa, 1000–1630*. Boulder: West View Press.
- Caron, B. (2019). A corpus-based macro-syntactic study of Naija (Nigerian Pidgin). *NaijaSynCor*. [online] Available at: <http://naijasynkor.huma-num.fr/project.html> [Accessed 26 Aug. 2019].
- Carpenter, K. and Devonish, H. (2010). Swimming against the tide. In: B. Migge, I. Léglise, and A. Bartens, eds., *Creoles in education: An appraisal of current programs and projects*, 1st ed. Amsterdam: John Benjamins (Creole Language Library (CLL), 36), pp. 167–181.
- Carvalho, M.J.M. de (2018). Cidades escravistas. In: L.M. Schwarcz, and F. dos Santos Gomes, eds., *Dicionário da escravidão e liberdade: 50 textos críticos*, 1st ed. São Paulo: Companhia das Letras, pp. 156–162.
- Chakravorty, G. (1999). *Can the subaltern speak?* Harvard: Harvard University Press.
- Chaudenson, R. (1992). *Des îles, des hommes, des langues: essai sur la créolisation linguistique et culturelle*. Paris: L'Harmattan.
- Chomsky, N. (1997). A linguística como uma ciência natural. *Mana*, 3(2), pp. 183–198.
- Coetsem, F. van. (1988). *Loan phonology and the two transfer types in language contact*. Dordrecht: Foris Publications.
- Corum, M. (2015). *Substrate and adstrate: The origins of spatial semantics in West African Pidgincreoles*. Berlin: De Gruyter Mouton.
- DeCamp, D. (1971). *Toward a generative analysis of a post-creole continuum*. Cambridge: Cambridge University Press, pp. 349–370.
- DeGraff, M. (2001). Morphology in creole genesis: Linguistics and ideology. *Current Studies in Linguistics Series*, 36, pp. 53–122.
- DeGraff, M. (2005). Linguists' most dangerous myth: The fallacy of creole exceptionalism. *Language in Society*, 34(4), pp. 533–591.
- Demoule, J.-P. (2014). *Mais où sont passés les Indo-Européens? Le mythe d'origine de l'Occident*. Paris: Éditions du Seuil.
- Drechsel, E.J. (2014). *Language contact in the early colonial Pacific: Maritime Polynesian Pidgin before Pidgin English*. Cambridge: Cambridge University Press.
- Eberhard, D. (2019). Who texts what to whom and when? The patterning of minority language texting with evidence from Maguindanaon. In: *The 6th international conference on language documentation & conservation (ICLDC)*. Manoa, Hawai'i.
- Eckert, P. (2008). Variation and the indexical field. *Journal of Sociolinguistics*, 12(4), pp. 453–476.
- Ellison, T.M. and Miceli, L. (2017). Language monitoring in bilinguals as a mechanism for rapid lexical divergence. *Language*, 93(2), pp. 255–287.
- Epps, P. (2006). The Vaupés melting pot: Tukanoan influence on Hup. In: R.M.W. Dixon and A.Y. Aikhenvald, eds., *Grammars in contact. A cross-linguistic typology*. Oxford: Oxford University Press (Explorations in Linguistic Typology, 4), pp. 267–289.
- Eriksen, L. (2011). *Nature and culture in prehistoric Amazonia: Using G.I.S. to reconstruct ancient ethnogenetic processes from archaeology, linguistics, geography, and ethnohistory*. [online] Available at: <http://lup.lub.lu.se/record/1890748> [Accessed 26 Aug. 2019].
- Faraclas, N. et al. (2007). The complexity that really matters: The role of political economy in creole genesis. In: U. Ansaldo, S. Matthews, and L. Lim, eds., *Deconstructing creole*, 1st ed. Amsterdam: John Benjamins (Typological Studies in Language, 73), pp. 227–264.

- Faraclas, N., ed. (2012). *Agency in the emergence of creole languages: The role of women, renegades, and people of African and indigenous descent in the emergence of the colonial era creoles*. Amsterdam: John Benjamins.
- Ferguson, C.A. (1959). Diglossia. *Word*, 15(2), pp. 325–340.
- de Filippo, C. et al. (2012). Bringing together linguistic and genetic evidence to test the Bantu expansion. *Proceedings of the Royal Society B: Biological Sciences*, 279(1741), pp. 3256–3263.
- François, A. (2011). Social ecology and language history in the Northern Vanuatu linkage: A tale of divergence and convergence. *Journal of Historical Linguistics*, 1(2), pp. 175–246.
- Frota, S. and Vigário, M. (2000). Aspectos de prosódia comparada: ritmo e entoação no PE e no PB. In: *Actas do XV Encontro da Associação Portuguesa de Linguística*. Braga: APL, pp. 533–555.
- Gilles, P. (2009). Virtual convergence and dialect levelling in Luxembourgish. *Folia Linguistica*, 32(1–2), pp. 69–82.
- Green, M., Ayafor, M. and Ozón, G. (2016). *A spoken corpus of Cameroon Pidgin English: Pilot study*, University of Oxford Text Archive. [online] Available at: <http://ota.ox.ac.uk/desc/2563> [Accessed 14 Feb. 2019].
- Güldemann, T. (2018). Areal linguistics beyond contact, and linguistic areas of Afrabia. In: T. Güldemann, ed., *The languages and linguistics of Africa*. Berlin: De Gruyter Mouton, pp. 448–545.
- Gumperz, J.J. and Wilson, R. (1971). Convergence and creolization: A case from the Indo-Aryan/Draavidian border in India. In: D. Hymes, ed., *Pidginization and creolization of languages*, 1st ed. Cambridge: Cambridge University Press, pp. 151–167.
- Hammarström, H. and Güldemann, T. (2014). Quantifying geographical determinants of large-scale distributions of linguistic features. *Language Dynamics and Change*, 4(1), pp. 87–115.
- Hayward, R.J. (1991). A propos patterns of lexicalization in the Ethiopian language area. In: U. Claudi and D. Mendel, eds., *Ägypten im afro-orientalischen Kontext: Aufsätze zur Archäologie, Geschichte und Sprache eines unbegrenzten Raumes. Gedenkschrift Peter Behrens*. Köln: Rüdiger Köppe Verlag, pp. 139–156.
- Heyd, T. and Mair, C. (2014). From vernacular to digital ethnolinguistic repertoire: The case of Nigerian Pidgin. In: V. Lacoste, J. Leimgruber, and T. Breyer, eds., *Indexing authenticity: Sociolinguistic perspectives*. Berlin: De Gruyter, pp. 244–268.
- Holm, John. 2004. *Languages in Contact: The Partial Restructuring of Vernaculars*. Cambridge: Cambridge University Press.
- Hinskens, F. (2009). Dialect levelling: A two-dimensional process. *Folia Linguistica*, 32(1–2), pp. 35–52.
- Hualde, J.I. and Prieto, P. (2015). Intonational variation in Spanish: European and American varieties. In: S. Frota and P. Prieto, eds., *Intonational variation in Romance*. Oxford: Oxford University Press, pp. 350–391.
- Hualde, J.I. and Schwegler, A. (2008). Intonation in Palenquero. *Journal of Pidgin and Creole Languages*, 23(1), pp. 1–31.
- Huber, M. (1999). *Ghanaian Pidgin English in its West African context: A sociohistorical and structural analysis*. Amsterdam: John Benjamins (Varieties of English Around the World, G24).
- Huttar, G.L. (2019). Review of language contact in Africa and the African diaspora in the Americas: In honor of John V. Singler. In: C. Cutler, V. Zvezdana, and P. Angermeyer, eds., *Journal of Pidgin and Creole Languages*, 34(1), pp. 178–183.
- James, W. (2003). The role of tone and rhyme structure in the organisation of grammatical morphemes in Tobagonian. In: I. Plag, ed., *Phonology and morphology of creole languages*. Tübingen: Max Niemeyer Verlag, pp. 165–192.
- Johnson, A.W. and Earle, T.K. (2000). *The evolution of human societies: From foraging group to Agrarian state*. Stanford, CA: Stanford University Press.
- Kasraee, N. (2017). *Working towards a standard pidgin*. [online] Available at: www.bbc.co.uk/blogs/academy/entries/70f2a30c-40a5-463c-9480-1d63e7d5f44a [Accessed 26 Aug. 2019].
- Khan, A.J. (2006). *Urdu/Hindi: An artificial divide*. New York: Algora Publishing.
- Kivisild, T. et al. (2003). The genetic heritage of the earliest settlers persists both in Indian tribal and caste populations. *American Journal of Human Genetics*, 72(2), pp. 313–332.

- Klein, H.S. (2018). Demografia da escravidão. In: L.M. Schwarcz and F. dos Santos Gomes, eds., *Dicionário da escravidão e liberdade: 50 textos críticos*. São Paulo: Companhia das Letras, pp. 185–194.
- Klein, W. and Perdue, C. (1997). The basic variety (or: couldn't natural languages be much simpler?). *Second Language Research*, 13(4), pp. 301–347.
- Kopitar, J. (1829). Albanische, walachische und bulgarische Sprache. *Jahrbücher der Literatur*, 46, pp. 59–106.
- Koplenig, A. (2019). Language structure is influenced by the number of speakers but seemingly not by the proportion of non-native speakers. *Royal Society Open Science*, 6(181274).
- Kortmann, B. and Lunkenheimer, K., eds. (2011). *Electronic world atlas of varieties of English (eWAVE)*. Leipzig: Max Planck Institute for Evolutionary Anthropology. [online] Available at: <http://ewave-atlas.org/>.
- Kusters, W. (2003). *Linguistic complexity: The influence of social change on verbal inflection*. Utrecht: LOT.
- Labov, W. (2001). *Principles of linguistic change: Social factors*. Malden, MA: Blackwell.
- Laitin, D.D. (1992). *Language repertoires and state construction in Africa*. Cambridge: Cambridge University Press.
- Lefebvre, C. (1998). *Creole genesis and the acquisition of grammar: The case of Haitian Creole*. *Cambridge studies in linguistics*, vol. 88. Cambridge: Cambridge University Press.
- Lim, L. (2009). Revisiting English prosody: (Some) New Englishes as tone languages? *English World-Wide*, 30(2), pp. 218–239.
- Linebaugh, P. and Rediker, M. (2013). *The many-headed hydra: Sailors, slaves, commoners, and the hidden history of the revolutionary Atlantic*. Boston: Beacon Press.
- Litamahuputty, B. (2012). *Ternate Malay: Grammar and texts*. [PDF] Utrecht: LOT. Available at: www.lotpublications.nl/ternate-malay-ternate-malay-grammar-and-texts [Accessed 22 May 2020].
- Lüpke, F. (2016). Uncovering small-scale multilingualism. *Critical Multilingualism Studies*, 4(2), pp. 35–74.
- Lupyan, G. and Dale, R. (2010). Language structure is partly determined by social structure. *PLoS One*, 5(1), p. e8559.
- Makoni, S. and Pennycook, A. (2007). Disinventing and reconstituting languages. In: S. Makoni and A. Pennycook, eds., *Disinventing and reconstituting languages*, 1st ed. Clevedon: Multilingual Matters (Bilingual Education and Bilingualism, 62, pp. 1–41.
- Manfredi, S. et al. (2019). A first variationist approach of the NaijaSynCor corpus. Paper presented at the International Najá Symposium. University of Ibadan, Nigeria, 27–29 June.
- Masica, C.P. (1976). *Defining a linguistic area: South Asia*. Chicago: Chicago University Press.
- Matras, Y. and Sakel, J., eds. (2007). *Grammatical borrowing in cross-linguistic perspective*. Berlin: De Gruyter Mouton (Empirical approaches to language typology, 38).
- Mattheier, K.J. (1996). Varietätenkonvergenz: Überlegungen zu einem Baustein einer Theorie der Sprachvariation. In: P. Auer, F. Hinskens, and K.J. Mattheier, eds., *Sociolinguistica*, 10(1), pp. 31–52.
- Mattos, H. and Grinberg, K. (2018). Código penal escravista e estado. In: L.M. Schwarcz and F. dos Santos Gomes, eds., *Dicionário da escravidão e liberdade: 50 textos críticos*. São Paulo: Companhia das Letras, pp. 163–168.
- McWhorter, J.H. (1998). Identifying the creole prototype: Vindicating a typological class. *Language*, 74(4), pp. 788–818.
- Michael, M. (2008). Who is Amhara? *African Identities*, 6(4), pp. 393–404.
- Michaelis, S.M. et al., eds. (2013). *APiCS online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. Available at: <http://apics-online.info/>.
- Migge, B. (1998). *Substrate influence in the formation of the Surinamese plantation creole: A consideration of the sociohistorical data and linguistic data from Ndyuka and Gbe*. PhD. Ohio State University.
- Mintz, S.W. (1971). The sociohistorical background to pidginization and creolization. In: D. Hymes, ed., *Pidginization and creolization of languages*, 1st ed. Cambridge: Cambridge University Press, pp. 481–498.

- Morland, P. (2019). *The human tide: How population shaped the modern world*. New York, CO: PublicAffairs.
- Mufwene, S. (1996). The founder principle in creole genesis. *Diachronica*, 13, pp. 83–134.
- Mufwene, S.S. (2001). *The ecology of language evolution*. Cambridge University Press.
- Mufwene, S.S. and Vigouroux, C.B. (2008). Colonization, globalization, and language vitality in Africa: An introduction. In: C.B. Vigouroux and S.S. Mufwene, eds., *Globalization and language vitality: Perspectives from Africa*, 1st ed. London and New York: Continuum International Publishing, pp. 1–31.
- Muthwii, M.J. and Kioko, A.N., eds. (2004). *New language bearings in Africa: A fresh quest*. Buffalo, NY: Multilingual Matters.
- Muysken, P., ed. (2008a). *From linguistic areas to areal linguistics*. Amsterdam: John Benjamins.
- Muysken, P. (2008b). Introduction: Conceptual and methodological issues in areal linguistics. In: P. Muysken, ed., *From linguistic areas to areal linguistics*. Amsterdam: John Benjamins, pp. 1–23.
- Muysken, P. (2017). The transformation of a colonial language: Surinamese Dutch. In: K. Yakpo and P. Muysken, eds., *Boundaries and bridges: Language contact in multilingual ecologies*, 1st ed. Berlin: De Gruyter Mouton, pp. 283–308.
- Muysken, P. and Smith, N., eds. (2015). *Surviving the middle passage: The West Africa- Surinam Sprachbund*. Berlin: De Gruyter Mouton.
- Nassenstein, N., Hollington, A. and Storch, A. (2018). Disinventing and demystifying youth language: Critical perspectives. *The Mouth (Critical Studies on Language, Culture and Society)*, (3), pp. 9–28.
- Nordhoff, S. (2009). *A grammar of upcountry Sri Lanka Malay*. Utrecht: LOT.
- O’Connor, L. and Muysken, P. (2014). *The native languages of South America: Origins, development, typology*. Cambridge University Press.
- Osei-Tutu, K.O.A. (2016). Lexical borrowing in Ghanaian Student Pidgin (GSP) – The case of Akan loanwords and loan translations. In: N. Faraclas, et al., eds., *Celebrating multiple identities: Opting out of neocolonial monolingualism, monoculturalism and mono-identification in the greater Caribbean*. Willemstad: University of Curaçao and Fundashon pa Planifikashon di Idioma, pp. 47–54.
- Pakendorf, B., de Filippo, C. and Bostoen, K. (2011). Molecular perspectives on the Bantu expansion: A synthesis. *Language Dynamics and Change*, 1(1), pp. 50–88.
- Paris, D. (2009). “They’re in my culture, they speak the same way”: African American language in multiethnic high schools. *Harvard Educational Review*, 79(3), pp. 428–448.
- Paulston, C.B. and Paulston, R.G. (1980). Language and ethnic boundaries. *Language Sciences*, 2(1), pp. 69–101.
- Rao, R. and Sessarego, S. (2016). On the intonation of Afro-Bolivian Spanish declaratives: Implications for a theory of Afro-hispanic creole genesis. *Lingua*, 174, pp. 45–64.
- Rickford, J.R. (1987). *Dimensions of a creole continuum: History, texts, and linguistic analysis of Guyanese creole*. Stanford, CA: Stanford University Press.
- Rüsch, M. and Nassenstein, N. (2016). Ethno-regional ideologies and linguistic manipulation in the creation of the youth language Leb Pa Bulu. *Critical Multilingualism Studies*, 4(2), pp. 174–208.
- Salmons, J. (1992). *Accentual change and language contact: Comparative survey and a case study of early Northern Europe*. Stanford, CA: Stanford University Press.
- Schuchardt, H. and Gilbert, G.G. (1980). *Pidgin and creole languages: Selected essays by Hugo Schuchardt*. London: Cambridge University Press.
- Sessarego, S., Rao, R. and Butera, B. (2017). On the reconfiguration of prosodic systems: Aspects of Chinch Spanish intonation. In: *Society for Pidgin & Creole Linguistics (SPCL) summer conference*. University of Tampere.
- Singler, J.V. (1995). The demographics of creole genesis in the Caribbean: A comparison of Martinique and Haiti. In: Arends, J. ed., *The early stages of creolization*. Amsterdam: John Benjamins, pp. 203–232.
- Sinnemäki, K. and Di Garbo, F. (2018). Language structures may adapt to the sociolinguistic environment, but it matters what and how you count: A typological study of verbal and nominal complexity. *Frontiers in Psychology*, 9.

- van Sluijs, R., van den Berg, M. and Muysken, P. (2016). Exploring genealogical blends: The Surinamese Creole cluster and the Virgin Island Dutch Creole cluster. *Lingua*, (178), pp. 84–103.
- Smith, N. and Adamson, L. (2006). Tonal phenomena in Sranan. *Sprachtypologie und Universalienforschung*, 59(2), pp. 211–218.
- Sobolev, A.N. (2004). On the areal distribution of syntactic properties in the languages of the Balkans. *Balkan Syntax and Semantics*, 67, p. 59.
- Souag, L. (2014). *Berber and Arabic in Siwa, Egypt: A study in linguistic contact*. Cologne: Rüdiger Köppe Verlag.
- Stell, G. and Yakpo, K., eds. (2015). *Code-switching between structural and sociolinguistic perspectives*. Berlin: De Gruyter.
- Szeto, P.Y., Lai, J.Y. and Ansaldo, U. (2019). Creole typology is analytic typology. *Language Ecology*, 3(1), pp. 89–119.
- Thomason, S.G. and Kaufman, T. (1988). *Language contact, creolization, and genetic linguistics*. Berkeley: University of California Press.
- Thurston, W. (1987). *Processes of change in the languages of North-Western New Britain*. Canberra: Research School of Pacific and Asian Studies, Australian National University.
- Trudgill, P. (2010). Contact and sociolinguistic typology. In: R. Hickey, ed., *Handbook of language contact*. Malden, MA: Wiley-Blackwell, pp. 299–319.
- Trudgill, P. (2011). *Sociolinguistic typology: Social determinants of linguistic complexity*. Oxford: Oxford University Press.
- United Nations (2017). *World population prospects: The 2017 revision, key findings and advance tables. Working Paper ESA/P/WP/248*. [PDF] New York: United Nations Department of Economic and Social Affairs, Population Division.
- Vanhove, M. et al., eds. (2012). *Morphologies in contact*. Berlin: Akademie Verlag.
- Wa Thiong'o, N. (1992). *Decolonising the mind: The politics of language in African literature*. Nairobi: East African Publishers.
- Walicek, D.E. (2009). The founder principle and Anguilla's homestead society. In: R. Selbach, H.C. Cardoso, and M. van den Berg, eds., *Gradual creolization: Studies celebrating Jacques Arends*, 1st ed. Amsterdam: John Benjamins, pp. 349–372.
- Weinreich, U. (1953). *Languages in contact*. New York: Linguistic Circle of New York.
- Yakpo, K. (2009). Complexity revisited: Pichi (Equatorial Guinea) and Spanish in contact. In: N.G. Faraclas and T. Klein, eds., *Simplicity and complexity in creoles and pidgins*, 1st ed. London: Battlebridge, pp. 183–215.
- Yakpo, K. (2015). Code-switching and social change: Convergent language mixing in a multilingual society. In: G. Stell and K. Yakpo, eds., *Codeswitching between structural and sociolinguistic perspectives*, 1st ed. Berlin: De Gruyter, pp. 259–287.
- Yakpo, K. (2016a). O estatuto do pichi na Guiné Equatorial. *PLATÓ (Revista Digital do Instituto Internacional da Língua Portuguesa)*. (Glotopolítica na Guiné Equatorial), 3(6), pp. 20–40.
- Yakpo, K. (2016b). “The only language we speak really well”: The English creoles of Equatorial Guinea and West Africa at the intersection of language ideologies and language policies. In: S. Castillo-Rodríguez and L. Morgenthaler Garcia, eds., *International Journal of the Sociology of Language (Special issue: Exploring Glottopolitical Dynamics in Africa: The Spanish Colonial Past and Beyond)*, (239), pp. 211–233.
- Yakpo, K. (2017a). Creole in transition: Contact with Dutch and typological change in Sranan. In: K. Yakpo and P. Muysken, eds., *Boundaries and bridges: Language contact in multilingual ecologies*. Berlin: De Gruyter Mouton, pp. 57–85.
- Yakpo, K. (2017b). Half a billion speakers in 2100? Aspects & implications of the growth of the English creoles & pidgins of West Africa. In: *30th West African Languages Congress (WALC 2017) and 10th Linguistics Association of Ghana (LAG) conference*. Winneba, Ghana.
- Yakpo, K. (2017c). Towards a model of language contact and change in the English-lexifier creoles of Africa and the Caribbean. *English World-Wide*, 38(1), pp. 50–76.

- Yakpo, K. (2017d). Unity in diversity: The homogeneity of the substrate and the grammar of space in the African and Caribbean English-lexifier Creoles. In: C. Cutler, Z. Vrzic, and P. Angermeyer, eds., *Language contact in Africa and the African diaspora in the Americas. In honor of John V. Singler*, 1st ed. Amsterdam: John Benjamins, pp. 225–250.
- Yakpo, K. (2019). Inheritance, contact, convergence: Pronominal allomorphy in the African English-lexifier creoles. *English World-Wide*, 40(2), pp. 201–225.
- Yu, A.C. (2013). Individual differences in socio-cognitive processing and the actuation of sound change. In: A.C. Yu, ed., *Origins of sound change: Approaches to phonologization*, 1st ed. Oxford: Oxford University Press, pp. 201–227.

Language contact

Pragmatic factors

Peter Auer

1. Introduction and definitions

Language contact is usually seen as a result of social factors enabling, encouraging or forcing speakers of different languages to communicate with each other. The type and amount of linguistic contact appears to be conditioned by these social factors (cultural, political, or economic superiority and power, etc.), as well as the concomitant language ideologies. In addition, grammatical parameters (the linguistic resources available to the speakers, the amount of structural overlap between the grammars and vocabularies, the amount of variation within the languages, etc.) have been shown to impact on the quantity and quality of language contact (see Thomason and Kaufman, 1988 for an initial overview). It is often forgotten that in addition to (and sometimes underlying) these social and linguistic factors, the pragmatics of language use also play an important role. These are the topics of this chapter.

Before going into details, both the concept of language contact and that of pragmatics need some clarification. The idea of ‘languages’ being in ‘contact’ is part of the structuralist heritage of linguistics (even though the notion of ‘Sprachmischung’ has of course nineteenth-century predecessors); in this tradition, research on language contact has focused on the outcomes of processes due to which (at least) one linguistic system changes under the influence of another, without looking at these processes themselves (i.e., after the change was complete). The term ‘language contact’ is here short for ‘contact-induced language change.’ At this level of abstraction, the idea of a pragmatic, action-based approach to language contact makes little sense, for languages do not ‘act.’ The problem with such a structuralist approach to language contact is that its explanatory value is limited. Modern contact linguists have therefore moved beyond the idea of language systems influencing each other and consider instead the languages users as the real carriers of language contact. It is generally acknowledged today that language contact occurs in bilingual encounters (in the widest sense of ‘bilingual’ and ‘encounter,’ including, in addition to social encounters, also written communication) and in bilingual cognition, not (or only in a very metaphorical sense) between language systems, and it is there that any explanation must start (cf. Matras, 2009). There is therefore a large area of overlap between research on language contact and research on bilingualism and second language acquisition. Indeed, some researchers’ definition of language contact closely resembles a definition of

bilingualism. Already Weinreich starts his ground-breaking 1963 book on ‘Languages in Contact’ by equating the two: ‘Two or more languages will be said to be IN CONTACT if they are used alternately by the same persons’ (1963, p. 1). Thomason (2001, p. 1) echoes him when she defines ‘language contact [as] the use of more than one language in the same place at the same time.’ On the other hand, there is also general agreement that even at the level of the speakers, language contact presupposes not only bilingual language use, but also some kind of (unidirectional or mutual) influence of the speakers’ cognitive representations of their languages. One can of course question whether there can be bilingualism without such a mutual influence of the languages in the speakers’ mind (and often also in their linguistic behaviour): the normal and unmarked state for the bilingual is not the monolingual, but the bilingual mode in which both languages are activated (Grosjean, 2001). In this sense, bilingualism may always entail language contact in the individual. Nevertheless, there is a difference of analytical focus between studies of language contact and studies of bilingualism.

In a narrow sense, language contact requires, according to most researchers, that the unilateral or mutual influence of the two languages is more than momentary, and that it occurs in a similar way in more than one individual. It is therefore necessary to distinguish between the effects of the use of two languages on the cognitive representation of these languages in the individual speakers, or on their behaviour, which may be idiosyncratic and fleeting, and the larger-scale impact of bilingual communicative contacts at the community-level.

For instance, the acquisition of a second language regularly passes through phases of interferences with the first language. However, these interferences usually disappear in the course of the acquisition process or remain idiosyncratic. In some cases, however, interference features fossilize and occur so systematically in a group of learners that a new variety emerges due to language contact, which is perhaps even handed on to the next generation as an L1. Another example is code-switching. It is often the case that bilingual speakers switch between their languages within an encounter or even within a sentence, without this behaviour leaving a trace in the mental representation of their two language systems, let alone on the community level. Here again, we are on the level of individual actions, and no language contact (in the narrow sense) is involved *per se*. In other cases, however, frequent switching or mixing may lead to convergence of the contact varieties in the bilingual community; this is a lasting effect of bilingual behaviour which falls under the heading of language contact.

In sum, while the origins of language contact cannot be located anywhere but in bilingual language use and cognition, language contact means more than bilingual behaviour and language processing in the individual, or even in an aggregate of individuals in isolation. Like all language change, it is a social process. Conversely, contact-induced change also goes beyond bilingualism. The contact-induced innovations that originate in bilingual speakers may spread across the community and be adopted by monolingual speakers as well. The sociolinguistic mechanisms of such a spread are usually the same as in endogenous change, but they are part of the total picture of language contact.

In this chapter, the pragmatics of language contact will therefore *not* be equated with the pragmatics of bilingual (inter-)action; this would be a field that is impossible to summarize in a short article anyway. Rather, the pragmatics of language contact will be understood as dealing with those non-momentary and non-idiosyncratic contact phenomena that can be explained (at least in part) by the pragmatics of (bilingual) language use. Not much research has been done from such a perspective and some of the thoughts presented in the following will therefore have a slightly programmatic character.

The definition of pragmatics as a field within linguistics is notoriously difficult. Mostly, pragmatics is defined as the study of action, i.e., as the ways in which we can ‘do things with

words' (to quote Austin's felicitous phrase). Other writers define pragmatics as the study of how linguistic meaning is created in and by its context. The two views are intricately linked, as actions can only be recognized and thereby become meaningful in their context. In this chapter, both views will be pursued. But whatever definition of pragmatics is chosen, it is obvious that a pragmatic approach to language contact is more restrictive than, and hence not identical with, a speaker-centred approach. This is due to the fact that a large number of contact phenomena on the level of the individual speaker do not serve pragmatic ends. They are either not intentional and occur below the radar of action-formation or fail to contribute to meaning-making in context.

A pragmatic approach to language contact can be broken down again into three more specific research questions.

First, it is possible to look into the pragmatic needs of speakers in bilingual encounters in order to establish correlations with the ensuing contact-induced changes. The question here is: for which pragmatic reasons do the speakers use the language which, as a consequence, undergoes contact-induced change? Or, for which pragmatic purposes does a new contact variety emerge? This requires a study of the 'communicative household' (Luckmann, 1986) of a community, which (theoretically) comprises all the action and activity types that are relevant to it, together with their status in the community. We can then ask which of these actions and activities are linked to the language that is undergoing change (or the variety that emerges due to language contact), and which actions and activities are linked to the dominant (donor) language.

The aim here is to find correlations between specific pragmatic needs and specific types of language contact. To give a simple example: when the activities linked to a specific variety only serve the pragmatic needs of trade and commerce, limited to perfunctory and temporally limited exchanges, types of language contact will emerge (such as trade pidgins with simplified grammar, or *lingue franche* with distinctive substrate features of the speakers' L1) that are different from those in a social situation in which the contact language has to take over the entire communicative household (as in language shift).

The realm of pragmatic explanations for contact-induced changes increases substantially when acts of identity (in the sense of Le Page and Tabouret-Keller, 1985) are included. Social identity is a powerful motivation for copying elements from another language or variety into one's own (or for refraining from doing so), as such borrowing is a highly efficient strategy for expressing affiliation with a social group. The usual appeal to language attitudes as a 'predictor' of contact-induced change (Thomason, 2010, p. 36) can be reframed – at least in its intentional part – as a pragmatic wish to perform acts of identity. In fact, if attitudes are understood as dispositions for behaviour, an approach that stresses the performative, pragmatic character of acts of identity seems to be particularly adequate. Acts of identity take place in the social arena of interaction and may be more suited to explain the adoption and spread of an identity-related innovation in the community than mental dispositions.

While this first pragmatic approach to language contact is linked strongly to (macro-) social factors in language contact, the **second**, perhaps somewhat more controversial approach zooms in on the microstructure of conversational exchanges. Code-switching (the use of elements from various 'codes' in an interactional exchange) is a specific resource that is available to multilingual speakers for structuring these exchanges and in this sense obviously a pragmatic phenomenon. The research question here is: what kind of consequences does code-switching have on the languages involved? Frequent code-switching can make the 'codes' involved converge with each other over time. That convergence can go hand in hand with matter borrowing (Matras, 2009), in the simplest case of lexical material, in more advanced cases by grammatical fusion and the emergence of so-called mixed languages. But it can also be manifest

in pattern borrowing, which means that the grammatical systems of the languages-in-contact becoming more similar (which facilitates frequent clause-internal switching). Convergence here is a non-intentional effect of intentional switching between the languages for discourse-related reasons.

A **third** pragmatic approach in contact linguistics starts from the definition of pragmatics as the study of linguistic meaning-making in context, with a special focus on deictic and discourse-structuring grammatical means whose very job it is to situate language in context. These pragmatic elements of language play a special role in language contact. It is well known that discourse markers and other pragmatically rich elements are particularly prone to borrowing. But ‘pragmatic borrowing’ goes far beyond discourse markers. It can affect syntax (e.g., via word order) and morphology (e.g., via the borrowing of derivational morphology or modal and epistemic affixes) as well. Here, the main research question is to investigate which pragmatic needs of the speaker are so pressing that the relevant structures of one language are copied into the language that undergoes contact-induced change.

2. Historical overview: actions and activity types as drivers of language-contact

Explanations of language contact often refer to lists of generalized social/sociolinguistic scenarios. For instance, Myers-Scotton (2002, pp. 31–32) distinguishes between (1) military invasion and subsequent colonization, (2) living in a border area or an ethnolinguistic enclave, (3) migration for social and economic reasons, (4) education, (5) spread of international languages, and (6) ethnic awareness. While such listings are useful and valid for demonstrating the variety of sociolinguistic constellations under which bilingualism and language contact occur, there are reasons to believe that they are not specific enough to lead to hypotheses on correlations between phenomena of language contact and their sociolinguistic embedding. It seems more likely that such hypotheses can be formulated when generalized sociolinguistic scenarios (such as the ones mentioned earlier) are broken down into the typical and recurrent action and activity types that are required in each case.

Let us consider just one of these scenarios as an example: migration. Migration can have hugely different outcomes in terms of language contact depending on which parts of the communicative household require competence in the language of the receiving society, even when the same languages and the same societies are involved. For instance, adult work immigration from Turkey and other Mediterranean countries into Germany (as well as other central and northern Europe countries) in the 1960s and 1970s required the acquisition of some German, but the pragmatic needs it had to serve were limited. German was basically needed in order to understand instructions to do simple manual jobs in the workplace. Outside work, few opportunities to use German were offered by the receiving society (nor were they sought by the immigrants since few of them originally planned a more than temporary stay). The resulting pidgin-like variant of German (cf. for a summary Klein and Perdue, 1997) corresponded to these limited pragmatic opportunities. The huge numbers of later immigrants in the 1990s from Eastern Europe (mainly Russia and Poland) did not lead to any similar developments. One of the explanatory factors¹ surely is the much broader pragmatic profile connected to German which is needed today in order to succeed in a job market that requires much more sophisticated skills, together with the long-time perspective of living in Germany connected with these migratory movements which also requires communicative competences well beyond those needed by (and offered to) the ‘guest workers.’

That migration as such is too broad a factor when it comes to finding correlations with formal outcomes of language contact can also be demonstrated by looking at second or third generation immigrants today and comparing them to previous cases of labour immigration into Germany. An example is the substantial immigration of Polish-speaking workers into the mining sites of the Ruhr District in the late nineteenth century. This migration has left no traces; people of Polish-speaking descent shifted to German in the second generation. In contrast, present-day Turkish-German bilingualism is rather stable; again, a constellation of factors contributes to this situation (among them the continuous influx of new L1 speakers of Turkish), but one of them is pragmatic: Turkish is needed and useful not only for communication within the large Turkish-speaking community in Germany today, but also for uninterrupted family, friendship, and media contacts with Turkey. The linguistic household requires spoken Turkish for informal purposes, while competence in more formal (and particularly written) forms of the language is dwindling and characterized by grammatical interferences with German (cf. Şimşek and Schroeder, 2011). No specific Turkish contact-variety has consolidated itself, but for identity displays, a pan-ethnic German-based youth register (multi-ethnolect) may be used. In addition to various deviations from mainstream German grammar (cf. Siegel, 2018) it allows for the integration of a few Arab- or Turkish-origin interjections or address terms that are shared with other speakers of immigrant background.² Alternatively, the speakers may resort to code-switching and mixing.

The general aim of a pragmatic approach to language contact is to map the resources in the speakers' bi- or multilingual repertoire onto the structures of their communicative household. According to Luckmann (1986), this communicative household can be seen as a system of recurrent patterns of action that have undergone typification. Their relevance is socially and culturally determined. While in one society, certain action types may be particularly salient and of high social value, the same action types may not be of prime relevance in another. Sequences of actions may also become typified into communicative genres (Luckmann, 1986). For instance, western middle class cultures have typified certain forms of phatic communication under the folk concept of 'conversation' (as evidenced by the normative expectations that lead to an evaluation of good and bad conversations, and the codification of an 'art of conversation' that can be taught and acquired), while in others, there may be elaborate, formal rules for insulting or praising others that have developed into communicative genres of their own. Salutation routines may be of primary importance in one society while they may be reduced to a minimum in others. On a still larger level of abstraction, action and activity types can be grouped into 'domains of action,'³ based on society members' schematic knowledge about the actions and activities linked to them.

The program of a pragmatic approach to language contact yields the best results when the pragmatic needs that a language must serve are rather precisely defined. Here are some examples.

When interlanguage communication among adult speakers is restricted to a small and narrowly defined number of action types in only one domain of action, one of the recurrent linguistic solutions is the emergence of simplified contact varieties (Ferguson and DeBose, 1977). They present a minimal solution for problems of understanding in sociolinguistic contexts in which understanding is rather unlikely, given the non-overlapping linguistic repertoires of the speakers. A clear case is the domain of trade and commerce, and well-known examples are trade pidgins which developed along trade routes, for instance on the West African coast. Typically, these varieties are highly specialized in their pragmatic profile: they are only used for activities linked to buying and selling (making offers, bargaining, etc.). Another requirement

for trade pidgins is that they must be easy to acquire; hence, their structure typically shows reductions on all levels of linguistic structure. A typical example of such a trade pidgin is Russonorsk, as spoken in the arctic region in order to facilitate communication between Norwegians and Russians (as well as Sami etc.) in the fish trade between the mid-eighteenth century and the Russian revolution (cf. Jahr, 1997). The speakers were mainly fishermen who sold their fish to the Russian ships (the Norwegian middle classes apparently spoke a simplified version of Russian). The pidgin had a limited number of topically relevant words coming from both languages, almost no morphology (despite the rich morphology of both contact languages), rigid SOV word order (despite the pragmatic word order in both contact languages), and it shows overgeneralizations of the use of certain grammatical words from both languages (such as the Russian question word *kak* ‘how’ as a generalized conjunction, or the Norwegian preposition *på*/ Russian preposition *pa* ‘in’ as a generalized preposition). The overall strategy of Russonorsk (as well as other trade pidgins) is to (over-)exploit overlapping structures and drop the divergent structures in the grammar of both languages, which is a typical strategy for ensuring interaction in a situation of hardly overlapping repertoires when only a limited number of action types need to be served.

Consider, as an example for this strategy, the following extract (from Franziskus and Gilles, 2012, pp. 63–64). The recording was made in Luxemburg where a huge number of foreigners commute for work from the neighbouring countries Germany, France, and Belgium, usually on a daily basis. The two speakers, employees of a supermarket, share only few linguistic resources. Melanie is a native speaker of French and comes from Belgium; she has a very limited knowledge of Luxemburgish (a Germanic language closely related to the dialects spoken in neighbouring Germany) and no knowledge of German. Elena is a native speaker of German and has no knowledge of French or Luxemburgish. However, her own dialect allows her to understand some Luxemburgish. Melanie and Elena share a core inventory of words related to the (officially Luxemburgish-speaking) domain of work, some of which have cognates in their own language: *Etikett* (‘price tag’), *Klient* (‘costumer,’ with a slightly different meaning in German, and a different pronunciation in French and Luxemburgish: /kliã/), *storno* (‘cancellation,’ ‘reversing entry’), *retour* (‘return’), *präis* (Germ. /preis/, no cognate in French). The episode takes place while they are stocking shelves (French in italics, German underlined, Luxemburgish and dialect boldface, ambiguous passages recte; transcription according to GAT2):

- 01 mel: *et le* **prÄis** eh direct eh:
 ‘and the price ehm directly ehm’
- 02 (1.0)
- 03 etiKETT?
 ‘price tag?’
- 04 ele: <<in the back> ja=ja [ge] nau;>
 ‘yes yes exactly’
- 05 mel: [oui;]
- 06 (1.0)
- 07 ele: und eh (.) (*für*) **Oben**=ne?
 ‘and uhm (for) top=-eh?’
- 08 (1.0)
- 09 (*für*) **Oben**.
 ‘for top.’
- 10 wenn *kliEnt* **nEt** (.) **NET** [**kofen**=ne?]
 ‘when client not not buy eh?’

- 11 mel: [oui fri] GO.
‘yes fridge.’
- 12 (1.0)
- 13 ele: eh ma?
‘uhm do’
- 14 ja=aber ERST eh retour.=ne?
‘yes but first uhm retour. eh?’
- 15 STORno.
‘cancellation’
- 16 (1.0)
- 17 mel: ah oKE.
- 18 ele: ne?
‘right?’
- 19 machs jetzt hier normal dann sO: etiKE[TT?]
‘now you do it here normally then like this price label?’
- 20 mel: [oui]
- 21 ou kLEin etiKETT?=eh:-
‘or small price label=uhm:’
- 22 ele: (.) *non*
‘no’
- 23 machs hier=kanns hIer RIChtig etikett.
‘you put here = you can correct price tag here’
- 24 mel: *oui*
- 25 ele: (de) kIent [KAUFT ja.]
‘the client buys, yes.’
- 26 mel: [oKE]

In the pragmatically highly determined context of their work in the supermarket, the speakers have developed strategies to make themselves understood. Two of these strategies stand out, and it is easy to see their resemblance to the structures found in a trade pidgin such as Russonorsk. One is the use of key terms linked to the action domain in question and known to both of them. They provide a kind of skeleton for meaning-making. When these words are cognates in the contact languages, they are particularly useful. The second strategy is grammatical simplification; Elena’s German utterances such as *aber ERST eh retour* ‘but first return’ (14) or *kanns hIer RIChtig etikett* ‘you can correct price tag’ (23) come from a simplified register (‘xenolect’) in which full verbs are lacking. Melanie sometimes also speaks in fragments of German with highly simplified grammar (*kLEin etiKETT?*, line 21 instead of *ein kleines Etikett* ‘a small tag’). The same holds for Melanie’s French (cf. *oui friGO* ‘yes, fridge,’ line 11). A third strategy is to switch for minimal contributions in the other person’s language; cf. Elena’s use of the French negation particle *non* in line 22.

Russonorsk is typical but also special as the groups of speakers who used it to communicate were of more or less equal social status, which may have fostered the inclusion of Norwegian and Russian elements. In other cases of contact with a very narrow pragmatic profile, power asymmetries restrict the overtly borrowed components to one language (the one spoken by the more powerful group). The *Gastarbeiter* variety of German alluded to earlier is an example (as well as many other pidgins that have emerged in the work domain); the L1 of the speakers (Turkish, Italian, Greek, Spanish, etc.) does not play an overt role in it. So-called *Kitche Duits* spoken by African servants in German households is another example, although here more

indigenous lexicon seems to have been involved (cf. Deumert, 2009). Once again only one domain (that of work/service) is concerned, and the number of action types which the register has to serve is restricted.

In these examples, a minimum of mutual comprehension is at stake (and achieved); but there are also contact varieties of an equally narrow pragmatic profile which serve the opposite, i.e., non-comprehension. An example are cryptolects ('secret languages'). Here, the pragmatic motivation is to communicate with group-members in certain situations and for certain purposes in the presence of out-group participants without being understood (which may be advantageous, for instance while bargaining with a customer). For such a cryptolect to work, it is necessary to replace (parts of) the core referential lexicon of the language shared with the out-group with words taken from one or more languages that are not intelligible to them; syntax and inflectional morphology may not be affected. Matras (2009, pp. 291–297) discusses several cryptolects that used to be spoken in Germany, such as Yenish, which is in use among traveling families who live on changing service occupations, or Lekoudesch, which was used by Jewish cattle-traders. These languages can be called 'symbiotic' (Smith, 2000) in the sense of relying on another language (the host) for most of their grammar and lexicon, with only the relevant lexical items replaced by new (contact) vocabulary. The host language of Yenish is (dialectal) German, into which words from Rotwelsch (the German 'thieves' cant'), Hebrew and Romani were borrowed. The host language of Lekoudesch was a Judeo-German ethnolect to which Hebrew words were added.

Another example of a contact variety with a narrow pragmatic profile is language play. A very learned example is Macaronic Latin, which emerged as a literary genre of parody, and was popular from the fourteenth to the nineteenth century. It first developed in Italy and was then copied in almost all other European societies by elite writers. Macaronic Latin (cf. Demo, i.pr.) consisted of a mixture of (classical/vulgar/medieval/'kitchen') Latin and Italian/Italo-romance dialect. It started out as a parody of the 'degenerate' Latin spoken and written in Europe in the early renaissance (i.e., a Latin full of interferences with the vernacular languages). Quite different from the examples discussed before, the structure of Macaronic Latin involves no simplifications. Most of the grammatical elements are taken from Classical Latin, into which lexemes/roots of the vernacular language are inserted. The only pragmatic function in addition to parody and fun was to display the writer's sophisticated knowledge of Classical Latin and its literature (to which many allusions were made), which nicely corresponds with its complex structure.

So far, we have discussed examples in which a very specific and narrow pragmatic profile could be correlated with structural features of language contact. Once the pragmatic tasks to be served by the variety/register/language undergoing contact-induced change multiply, and become more complex, the relationship between pragmatic function and linguistic structure is more difficult to pinpoint. It is then often useful to compare two socially, linguistically, and culturally very close groups in which different processes of language contact set in due to diverging pragmatic profiles. Here is an example.

Adamou (2010) compares two languages spoken in Thrace (Greece) by Muslim minorities which have traditionally been in contact with Turkish: the Southern Slavonic language Pomak and the Balkan-Romani (Vlach) language spoken in Komotini. In both cases, the speakers are mostly trilingual today (with Turkish and Greek in addition to the minority language), and Turkish is more and more becoming the dominant language. However, Pomak is not undergoing much contact-induced change, although speakers code-switch between Pomak and Turkish; they do so while keeping the structures of the languages more or less separate. There are numerous lexical borrowings, which are always integrated into Pomak morphology; verbs are

rarely borrowed. In addition, the speakers use Turkish expressions for rituals (such as greetings) to express their Islamic identity (which of course contrasts with the orthodox Christian orientation of the Greek-speaking majority). Quite in contrast, Komotini Romani has developed into a ‘fused lect’ (Auer, 2014, i.e., a lect characterized by heavy structural borrowing), in which borrowed Turkish elements have become obligatory. This Turkish influence is not recent but must have impacted on Komotini Romani during the Ottoman empire (as similar forms of fusion can be found in Vlach communities in which Turkish is not spoken any longer). The most striking difference from Pomak is that Turkish verbs are borrowed together with the entire Turkish aspect, person, and tense morphology into Komotini Romani, leading to a partial compartmentalization of the language (verbal system = Turkish, nominal system = Romani). The impact of Turkish on the language has therefore been much stronger than in the case of Pomak – despite the fact that both minority groups have experienced the same sociolinguistic history with Turkish as the prestige language (and *lingua franca*) in the Balkans. As Adamou points out, the reasons are to be found in the language ecology of the two languages at the time when the present-day contact patterns originated, and the different pragmatic profiles that resulted from it. The Pomaks lived as (semi-sedentary) farmers and cattle-breeders in a rather isolated mountain region, and their communicative contact with Turkish-speaking people in Ottoman times was restricted to the religious domain (with the exception of (elite) members of the community who travelled). Consequently, the borrowings come from this domain; they served religious and cultural pragmatic needs (and therefore mainly extend to close kinship terms, greetings, and thanking expressions). The Roma in Ottoman times were mostly non-sedentary and always in interaction with Turkish-speaking people for instrumental reasons, as their existence was built on trading and providing services as craftsmen for the *gadze* (non-Roma). But at the same time, the Roma communities had strict rules on how to communicate with the outsiders thereby protecting their own group identity. This social-communicative situation is reflected in grammatical structure: Turkish elements are marked as such (by carrying along their Turkish grammatical markers) and not integrated into Romani structures, but they are also systematic and therefore compartmentalize the grammar. Hence the different language ecologies with their specific pragmatic profile resulting from different ways of interacting with mainstream society can be mapped on the existing language contact patterns. It might be added that a purely identity-related or attitudinal explanation would not be successful in accounting for the differences in this case: both groups are separate from Greek mainstream society (due, in the first place, to their religious orientation), but also from the Turks. If anything, it is the less Turkish-oriented community, the Roma, which has experienced more Turkish language contact.

But of course, the choice of a particular way of speaking one’s own language which is strongly or weakly (or not at all) influenced by an (outgroup) contact language, can also be an act of identity. Language-related acts of identity are superimposed on the pragmatics of language use, i.e., they are usually not performed for their own sake, but as a kind of metapragmatic comment on what is verbally done.

Germans in the late nineteenth century who increasingly avoided French loans and instead chose the (partly neologistic) German alternatives (*Fernsprecher* for *Telephon*, *Fernschreiben* instead of *Telegramm*, *Fahrkarte* instead of *Billet*, *Bahnsteig* instead of *Perron*, etc.) did so deliberately in order to position themselves against the ‘invasion’ of the German language by French and, by extension also against the French and France in general. Undoing the effects of a previous, long-standing language contact was an act of identity here (which was linguistically successful, in the sense that standard German as spoken today shows less French loanwords than Swiss standard German). The same acts of identity aiming at divergence can be

found today in nationalist independence movements within the European states. For instance, Corsican activists might prefer non-Italian (sounding) alternants to those close or identical to Italian: *avali* instead of *ora* ‘now,’ *ci vole à* & INF instead of *bisogna* & INF ‘it is necessary to . . .,’ etc. (cf. Kailuweit, 2014); Galician activists might deliberately choose Portuguese-sounding variants over Castilian ones, etc. All modern constructions of national standard languages are partly built on the creation of *Abstand* (in Kloss’ terms, 1967) and are therefore acts of identity.

Acts of identity are also involved when substrate features due to imperfect L2 learning are not avoided, although the speakers are well aware of them. The numerous national varieties of English with substrate interference from the local language(s) are good examples (see Rüdiger, 2017 for a recent description of Korean L2 English and its Korean substrate features, or the numerous studies on Singapore English and its Chinese substrate, e.g., Leimgruber, 2013). Phonology and phonetics are particularly well suited for identity displays of this kind, as denotational reasons for copying are lacking. A European example would be the persistent use of word-initial stress on bisyllabic words in the French taught and spoken as an L2 in Germanophone Switzerland (a feature that also occurs in traditional dialects of French-speaking Switzerland but is quickly disappearing there due to convergence with the Paris norm of French, resulting from long-standing contact with Germanic and its initial stress pattern in native trochees).

A well-known, though complex and partly disputed example of prestige-driven language contact not due to imperfect second language learning but to (elite) bilingualism is the spread of uvular /R/ in central Europe (see Trudgill, 1974; King and Beach, 1998). While it is unlikely that the entire area in which uvular /R/ is spoken today has borrowed this sound from the French (more exactly, Paris) pronunciation, there is evidence that on a more local level, such borrowing took place and was based on collective acts of identity. One example is the city of Ghent, where the local bourgeoisie copied the uvular rhotic from French, the prestige language in Belgium in the nineteenth century. From Ghent, it spread to the surroundings of the city (Rogier, 1994). In Brussels, too, contact borrowing from French into Flemish seems to have taken place (see van de Velde, Tops and van Hout, 2013 for the details of the spread of /R/ in Flanders, a process which is still under way).

3. Critical issues and topics

3.1 *Code-switching and convergence*

The second, quite different approach to the pragmatics of language contact takes as its starting point the pragmatics of bilingual interaction itself, i.e., the use of more than one language in an interactional episode in order to achieve certain pragmatic ends. In a terminology introduced in Auer (1999), I speak of code-switching here (as distinct from non-discourse-functional mixing). The details of how code-switching can function as a contextualization cue have been outlined elsewhere (cf. Auer, 1995) and will not be repeated here. It is important to underline that the discourse functions of code-switching are based on the participants’ perception and construction of the two codes as distinct. It is the very contrast of their juxtaposition that provides the potential of code-switching to contextualize an upcoming activity as different from the preceding, which, together with the sequential and other contextual embedding of the switch, makes it pragmatically meaningful. Syntactically speaking, code-switching can occur anywhere, but its preferred locus is the clause boundary (alternational switching) or a single word from which the switched-to language may ‘project’ to the phrase level (insertional

switching). In the latter case, we are dealing with a ‘nonce borrowing’ (but not all nonce borrowings have a function in discourse).

How can code-switching have an impact on the structure of the languages involved? There are several answers, all of which presuppose its gradual de-pragmatization and routinization. Paradoxically then, code-switching is more relevant for language contact the more its potential for assuming discourse functions diminishes, and above all, the more frequent it becomes.

Let us first consider the most obvious case, the way from single-word insertional code-switching to (established) loanwords (cf., e.g., Heath, 1989 with examples from Moroccan Arabic). Some of the more frequent discourse functions of insertional code-switching of single words include the wish to display cultural distinction, be it by claiming membership in the progressive avant-garde or the conservative arrière-garde, to exclude other participants, to pun and play with words, to circumvent and defuse taboo topics and delicate speech acts, to refer implicitly to another discourse or somebody else’s opinion (intertextual functions), etc. It stands to reason that this pragmatic strategy weakens the more it is repeated (and hence overexploited) and the more the juxtaposition of the two languages is the normal and unmarked way of speaking. Also, the contrast between the two codes on which the function of code-switching depends, diminishes the more the borrowed item is integrated into the sound pattern and/or in the grammatical frame of the receiving language. Integration and discourse-related function are therefore negatively correlated. A kind of firewall against integration is the preservation of the original morphological form of the borrowed word; hence the tendency for humanist code-switchers (like Martin Luther) to insert Latin words in their inflected form, a format preferred by educated speakers in Europe way into the twentieth century. For instance, Luther says *das auch **mortu+i** sollen herren warden* ‘that the dead will also be masters’ (with the plural nominative inflectional suffix – *i*), *so wurd er bald all **thesaur+os** wider zu sich bringen* ‘like this he will soon bring all the treasures back to him’ (accusative plural – *os*), *ob er nun **amor+em** verbirgt* ‘whether he hides (his) love’ (accusative singular – *em*), *so heyssen sie es **nov+am*** ‘so they call it new’ (accusative singular – *am*), etc. (in his *Tischreden*, cf. Stolt, 1964, pp. 59–75; the Latin words in boldface have the case and number endings required in this syntactic slot of the German sentence). The inverse does not hold, i.e., borrowings with a merely referential function may but need not be integrated. This may explain Poplack’s results of her diachronic study of English borrowings into Canadian French in which she finds no increase of integration over time (Poplack, 2018, pp. 127–140). Pragmatically motivated (intentional) nonce borrowing is of course not the only reason for borrowing, which may also be due to imperfect acquisition or language loss.

Once discourse-functional lexical insertions lose their contextualization value through more frequent use, they gradually become established loanwords. This is perhaps the most frequent effect of the pragmatics of bilingual discourse on the language system which is found practically in every language.

But lexical borrowing can have effects on the grammar of the receiving language as well. In particular, frequent borrowings of words showing the same derivational pattern may lead to the borrowing of this derivational pattern itself. It is a side-effect of the insertion of lexical materials. A well-known example is the extensive borrowing of derivational patterns from medieval French into English and German. Although most of these derivational affixes are today restricted to non-native stems, the restriction holds much less in the early history of their borrowing (cf. Hilpert, 2013, pp. 110–154 for the deverbal – *ment* suffix in English), and even where this restriction holds, the derivational pattern was/is highly productive in the receiving language, leading to new derivations above all with Latin/Greek stems. Some

derivational affixes copied from medieval French such as the German suffix *-ei* (from Old French *-ie*) as in *Bäcker-ei* ‘bakery,’ *Maler-ei* ‘painting (in its totality),’ have lost this restriction entirely.⁴

Whether the strategy to insert borrowed words together with their inflectional morphology (‘maximal insertion strategy’ as opposed to the ‘minimal insertion strategy’ of using only the stem/bare form; cf. Auer, 2014) can lead to the structural borrowing, is a more controversial question (see Matras, 2009, pp. 212–218 for a discussion). There is good evidence that so-called compartmentalized mixed languages result from this insertion strategy, in which a large part of the verbal system (lexemes and inflection) are taken from one language, and a large part of the nominal system from another (with many variants; see the preceding example of Komotini Romani).⁵

So far, I have discussed examples of ‘matter borrowing’; but code-switching can also make the languages in question converge in their structure (Thomason, 2001, pp. 131–136). One of the most radical examples often referred to in the literature comes from Gumperz’ and Wilson’s 1971 study of the Indian village of Kupwar.⁶ In this village in the district of Maharashtra, four languages have been in centennial contact: Kannada and Telugu (Dravidian languages) and Marathi and Hindi-Urdu (both Indo-Aryan). Multilingualism (particularly in the male population) and the permanent use of the languages in outgroup communication between the ethnic groups has not led to language shift, but rather to radical convergence, with Hindi-Urdu most, Kannada somewhat less and Marathi least affected. (Marathi is the official language of the state and used in particular as a *lingua franca*. Telugu has only few speakers and is not analyzed by Gumperz and Wilson any further.) When Gumperz and Wilson did their fieldwork (in the late 1960s), syntax, parts of morphology and phonetics had converged, while the lexicon and morphophonemics had remained separate. The entire process was additionally accompanied by simplification and reduction. As a result, the three languages as spoken in outgroup communication in the village could be translated into each other morpheme by morpheme. In particular, all syntactic (word order) differences were levelled out. Nevertheless, none of the three languages had been given up, as the existing ethnic divisions were a powerful social force keeping them separate. The overall picture was that of a balance between, on the one hand, a pragmatic motivation – the necessity of frequent contacts between the speakers of the three language communities in the village – leading to frequent code-switching and from there to convergence, and, on the other hand, a social motivation to index membership in the three ethnic communities based on caste, residence area, and religion. Gumperz and Wilson see these changes as ‘generated by code-switching’ (Gumperz and Wilson, 1971, p. 256).⁷

Less radical convergence can be found in many other bilingual communities in which switching and mixing are frequent. However, it is often difficult to decide whether switching has caused this convergence process (through which it becomes easier), or whether the convergence would have taken place anyway. Evidence for the active role of switching and mixing can be gathered from comparing monolingual and bilingual speech in the same sociolinguistic context. To give just one example, it is a well-established finding that Slavonic languages tend to lose gender distinctions when they are in close contact with Turkic languages (which do not have gender). Usually the masculine forms are overgeneralized into grammatical contexts in which the feminine is required. Muhamedova (2006) shows that in Kazakh/Russian mixing in urban Almaty, this loss of gender appears frequently in mixed utterances while it is much less prevalent in the monolingual speech of the same Russian/Kazakh bilinguals. This shows that switching/mixing at least favours the convergence of Russian to Kazakh, and perhaps even causes it.⁸

Code-switching and mixing can also (and regularly do) index social membership, i.e., switching and mixing styles often, if not always, are also acts of identity. Again, comparative studies with the same language pairs involved are telling. An example are the Arabic/Modern Hebrew contact patterns recently described by Kheir (2019 and MS) for the Arab and the Druze community in Israel. While the bilingual Arabs show the well-known discourse-functional patterns of code-switching and borrowing, Druze Arabic is so heavily mixed with Hebrew that Kheir analyzes it as being on the way to a matrix-language turn-over (in the sense of Myers-Scotton, 1998). In the speech of the Israeli Arabs, language contact with Hebrew also exists, but it never endangers the status of Arabic as the matrix language, while in the speech of the Israeli Druze, it is no longer possible to establish such a matrix language, since both languages equally contribute to the grammar of the Hebrew/Arabic mixed utterances. Kheir shows that the mixed Arabic/Hebrew is linked to the Druze community's distinctly more positive attitudes towards Israel and Modern Hebrew when compared to those of the Israeli (Muslim) Arabs; it can be followed that their mixing behaviour is an act of identity affiliating the speakers with mainstream Israeli society.

3.2 *Pragmatic borrowing*

With the third pragmatic approach to language contact, we enter a very different field. The question here is whether and under which conditions pragmatic features of a language can be borrowed into another. Interestingly, some of these elements are high on the borrowability scale (Matras, 2007), such as discourse markers and connectors, while others are low (e.g., person inflection). The class of linguistic resources that serve to embed a linguistic utterance into its situational context is therefore certainly not a 'natural' class with regard to borrowability.

Here, I only give a rough sketch with several examples from different layers of linguistic structure, obviously without any claim to completeness.

One of the most frequently borrowed elements in language contact is discourse markers. Discourse markers are 'part of an apparatus with which the speaker monitors and directs the hearer's processing of propositional content' (Matras, 2009, p. 95; see the overview in Matras, 2009, pp. 137–145). They are therefore pragmatic operators. Maschler (1998, pp. 130–131) further differentiates between interpersonal markers (for expressing agreement, or epistemic status), referential markers (for instance used to connect utterances by marking contrast or cause), structural markers (which organize the order of actions or introducing side remarks) and cognitive markers (for instance to mark hesitations). Various reasons are mentioned in the literature for which these markers may be particularly prone to borrowing. Syntactic non-integration into the grammatical structure of the sentence is probably mentioned most often. In some cases, the discourse markers that are borrowed allow the speaker to manage interaction in a way which is not possible on the basis of the resources of the receiving language alone, or the new borrowed and the equivalent old marker take on different functions and thereby provide a richer system of pragmatic markers (Oesch-Serra, 1998). In this case, the borrowing has functional reasons and enriches the grammatical system of the receiving language. However, in other cases, this explanation does not hold. Matras (1998, 1999) suggests a cognitive explanation in hypothesizing that speakers fail to keep their languages apart when using discourse markers because their use requires a special cognitive effort. This makes them fall back on the more entrenched options in their dominant language ('selection malfunction'). For him, discourse markers are just one element of a larger system of 'utterance modifiers,' which also includes connectors, interjections, tag questions, focus markers, modal particles, and certain adverbials. It seems, however, that utterance modifiers which do not occur at the clause

periphery, such as focus and modal particles, are less susceptible to borrowing.⁹ In the case of interjections, identity-related motivations may be more important than monitoring and directing the hearer's understanding. They seem to behave more like borrowed turn-final endearment terms, which also primarily have the function of indicating social affiliations. A point in case is the pan-European use of the Arabic/Turkish interjection *vallah* together with Arabic/Turkish address terms such as *lân*, *abi*, *kardesh/kardash*, *moruk/morok*, etc. in the urban vernacular youth vernaculars which have emerged on the basis of multi-ethnolects in the middle and north European countries, while discourse markers and particularly connectors have not been reported often to be borrowed into this register (see the contributions in Kern and Selting, 2011; Quist and Svendsen, 2010).

In derivational morphology, one of the frequently borrowed pragmatic affixes are diminutives. They clearly serve a pragmatic function, as they express affection and intensification more often than simple smallness. As an example, Weinreich (1963, p. 34) mentions the borrowing of the Yiddish diminutive – (*al*)(*e*) (identical with the (Southern) German suffix) into Modern Hebrew as an 'endearing form.' Another example is the Spanish diminutive – *ito*, which is borrowed into various Mesoamerican languages (Chamoreau, 2012). The distributional patterns suggest that suffix borrowing is preceded by a phase of lexical borrowings (i.e., stem, usually a proper name, and suffix were borrowed together) and that the suffix generalizes to different degrees to non-borrowed stems (a process known from other derivational borrowings).

Deictic elements for referring to place, time and person are rarely borrowed. However, if deictic adverbials take on functions typical of 'utterance modifiers,' their borrowability increases. For instance, originally temporal adverbials that mark temporal succession are often used as connectors in narrative utterances ('and then'), and the originally temporal adverbials of the discourse marker family *nu/nâ* (referring to the time of speaking: 'now') are famous for having developed a variety of discourse functions in many Germanic and Slavonic languages. They were borrowed between the European and into various non-European languages, among them Modern Hebrew (cf. Auer and Maschler, 2016 for examples). Matras (2009, p. 141) quotes the following example from a Volga-German speaker (from Anders, 1993, Russian in italics):

Potom (1.0) *nu* die Wolgarepublik, *potom/nu* so is weiter und weiter/ (1.0) die ganze *istorija!*
(*'then* (1.0) *well*, the Voga Republic, *then/well* it went on and on like this/ (1.0) the whole *story!*)

Demonstratives also tend to take on discourse marker functions and are then easily borrowed. For instance, Muhamedova (2006) shows that the Russian particles *eto* (originally a demonstrative adjective, but often used as a hesitation marker) and *vot* (originally a deictic manner adverb, but also used as a discourse marker for turn and topic closure) are borrowed into Kazakh and into Uighur (2006, pp. 185–187, 210–212).

Personal pronouns are rarely borrowed, and if they are, some other function in addition to person reference is usually involved (such as honorification; cf. the overview in Matras, 2009, pp. 203–208). Pattern borrowing seems to be more frequent. An example is the Highest Alemannic dialect spoken in the Northern Italian language enclave of Issime in the Aosta Valley (Zürner, 1999, pp. 214–218). Here, the full pronouns of the plural have developed composite forms modelled after the (non-standard) plural in the contact languages Italian, French and Francoprovençal. Hence, the simple nominative plural forms of the first, second and third person – *wir*, *ir* and *dschi* – alternate with the composite forms *wir+endri*, *ir+endri/ar+endri*, *dschi+endri* (and accordingly for the other cases); cf. Italian *noi altri*, *voi altri*, etc.¹⁰

Although deictic elements are rarely borrowed, the deictic system of a language can change under the influence of another language. For instance, Rehbein (2012) investigates the use of the Turkish demonstrative article *o* ('that') under the influence of German. He finds that bilingual children in Germany occasionally use *o* not for a deictic shift as in monolingual Turkish (drawing the attention of the hearer to a new referent), but as an anaphoric marker which is perhaps on its way to becoming a definite determiner modelled after the German anaphoric determiner *dies-*, referring back to an already established discourse referent. Pro drop languages have also been claimed to lose that feature in contact with languages such as German or English (Rehbein, 2012 for Turkish, but see the counterevidence for English/Spanish contact in Torres Cacoullos and Travis, 2018).

Pattern borrowing is also quite frequent in another deictic domain, that of tense and aspect. For instance, there is evidence that the analytic future and past tense patterns were borrowed from German into Czech in the twelfth century (Berger, 2014). Also well-known are cases of semantic borrowing. An example is the Spanish spoken by Quechua/Spanish bilinguals (Andean Spanish), where the Std. Spanish past perfect tense (*había* + past participle) has developed into a mirative marker, modelled on the Quechua marker – *sqa* for non-experienced past (cf. Babel and Pfänder, 2014).

Finally, it should not be forgotten that syntax also has pragmatic functions via word order, ellipsis, and various syntactic operations such as dislocations. Syntactic borrowing may therefore also be a case of pragmatic borrowing. As an example, consider again the Highest Alemannic dialects in the Aosta valley in Italy (Zürcher, 1999, pp. 340–349). These dialects are in the process of converging syntactically with Italian, the dominant language for most speakers, particularly the younger ones. Italian can topicalize the subject by fronting:

Nostra figlia fra un anno entra al collegio
lit.: 'our daughter, in one year will go to boarding-school'

instead of the unmarked word order:

Fra un anno nostra figlia entra al collegio or
Nostra figlia entra al collegio fra un anno.

In German (including its dialects) it is not possible to front the subject to clause-initial position if another constituent already occupies the position before the verb:

**Unsere Tochter in einem Jahr geht ins Internat.*¹¹

In order to come close to the Italian word order, the dialects of Gressoney and Issime therefore overexploit the pattern of left dislocation as it is available in German:

Ündsch töchter z joar gait=dsch den al collegio

'our daughter in-one year goes-she then to-the boardingschool'

Left dislocations in German (and its dialects) require the resumptive (demonstrative) pronoun to follow the left-dislocated NP directly:

Unsere Tochter, die geht in einem Jahr auf's Internat.
Lit. 'our daughter, she goes in one year to the boarding-school.'

In Gressoney/Issime, however, the anaphoric/resumptive pronoun *dsch* is a clitic attached to the finite verb, a highly non-canonical construction in German.

In order to downgrade a topic, Italian can move the subject to the right periphery of the clause:

È di Torino quell'avvocato
'(He) is from Turin, this lawyer.'

As Italian is a subject-drop language, *è di Torino* is a full clause, and *quell'avvocato* has the status of a right dislocation.

In colloquial German, which is a topic-drop language, the pre-verbal constituent can also be omitted if it is known from context:

Ist aus Turin, der Rechtsanwalt.

The dialects of Gressoney and Issime, however, do not drop the topic pronoun but – just like in the previous case of a left dislocation – attaches its clitic form to the finite verb, resulting in VS/S word order:

Ischt=er von turiin, dee a^avokat
'Is he of Turin, the lawyer'

This construction is used for a declarative sentence in the dialect but would only be acceptable as a question in German (including the Alemannic dialects).

The two contact-induced syntactic changes show that the dialects have converged with the Italian model almost completely. They are ready¹² to reanalyze the clitic pronouns as inflectional affixes, which would make the convergence perfect:

| | | | | | | | | | | |
|-----------------|--|-------------------|--|-------------------------------|--|---------------|--|-------------------|--|--------------------|
| <i>Nostra</i> | | <i>figlia</i> | | <i>fra un anno</i> | | <i>entr+</i> | | <i>a</i> | | <i>al collegio</i> |
| <i>Ündscht</i> | | <i>töchter</i> | | <i>z joar</i> | | <i>gait +</i> | | <i>dsch (den)</i> | | <i>al collegio</i> |
| <i>È</i> | | <i>di Torino</i> | | <i>quell'avvocato</i> | | | | | | |
| <i>Ischt=er</i> | | <i>von turiin</i> | | <i>dee a^avokat</i> | | | | | | |

4. Future directions

In this chapter, I have sketched three pragmatic perspectives on language contact. The topic can be approached by looking at the borrowing of pragmatic elements of language, i.e., those that embed linguistic utterances into a context of situation, which also the metapragmatics of framing utterances and managing speaker-hearer interaction (section 3.2). Even a cursory look at these pragmatic elements, which range from discourse markers to word order demonstratives, shows that they behave in a heterogeneous way with regard to borrowing. It seems therefore unlikely that an integrated approach can be developed in this field.

The second way of approaching language contact from a pragmatic angle is to focus on the effects of bilingual interaction on the languages involved (section 3.1). There are good reasons to believe that code-switching (including ad hoc borrowing) can have an impact on language

structure, starting from the very simple observation that ad hoc borrowings can develop into established loanwords. Some so-called mixed languages may also be the outcome of habitualized code-switching. In addition, there is evidence that code-switching can lead to the structural convergence of the switched languages. In all of these cases, language change is the unintentional effect of a particular verbal action, i.e., code-switching.

The third pragmatic approach to language contact is also the most ambitious program, although it has the longest tradition (section 2). It starts from the pragmatic profiles attached to the languages in contact, i.e., the actions, activities, and domains of activities that are associated with the linguistic resources in the repertoire of speakers or communities. Although this approach has always enjoyed common-sense appeal, its theoretical underpinnings and empirical relevance still need to be elaborated. The general assumption is that specific pragmatic profiles (and hence pragmatic needs linked to specific action types) correlate with specific structural outcomes of language contact. Such a pragmatic approach will be able to provide explanations for language contact phenomena and may contribute one factor to the complex of interacting parameters that are responsible for the effects bilingualism can have on language structure.

5. Further reading

Matras, Y. (2007). The borrowability of grammatical categories. In: Y. Matras and J. Sakel, eds., *Grammatical borrowing in cross-linguistic perspective*, 1st ed. Berlin: De Gruyter Mouton, pp. 31–74.

Important insights into the borrowability of pragmatic categories.

Matras, Y. (2009). *Language contact*. Cambridge: University Press.

Recommended as a thoroughly pragmatic orientation in contact linguistics.

Myers-Scotton, C. (2002). *Contact linguistics. Bilingual encounters and grammatical outcomes*. Oxford: Oxford University Press.

A standard introduction with a strong focus on sociolinguistics conditions on language contact.

6. Related topics

Code-switching and bilinguals' grammars, borrowing, social factors

Abbreviations

| | |
|-----|-----------------------|
| INF | infinitive |
| L1 | first language |
| L2 | second language |
| NP | noun phrase |
| SOV | subject, object, verb |

Notes

- 1 Surely, there are other reasons as well; for instance, the infrastructure for formal German language teaching has greatly improved.
- 2 This register is used by some monolingual Germans as well in the meantime; cf. Dirim and Auer, 2004; Freywald et al. 2011, among others.
- 3 'Domains' were introduced to bilingual studies by Fishman (1964), with reference to previous work by Schmidt-Rohr (1932, pp. 182–184). He conceives them in terms of topics, locales and role

relationships, although he sometimes also speaks of ‘domains of social interaction’ and even of ‘institutionalized spheres of activity’ (Fishman, Cooper, and Ma, 1971), which implies a pragmatic understanding. While for Fishman, domains are useful for describing patterns of language choice (certain domains are associated with certain languages), the idea pursued here is that actions, activities (genres) and domains (as spheres of actions) imply specific pragmatic demands and obligations on the speaker, which can explain (some of the) structural outcomes of language contact.

- 4 The suffix is still productive in German to form abstract nouns with a slightly negative connotation, cf. *Liebel+ei* ‘love affair,’ *Prügel+ei* ‘brawl,’ etc.
- 5 See Auer (2014) for details and more literature on the emergence of ‘mixed languages’ (fusion) from code-switching and mixing.
- 6 See Kulkarni-Joshi (2016) for a recent appraisal of the linguistic facts as described by Gumperz and Wilson and for an empirical study on the developments in Kupvar over the last 50 years.
- 7 The study does not give details on how code-switching is defined, and which types of switching can be found in the community. The analysis suggests that code-switching results from every person using his/her own language when communicating with a member of another ethnic group.
- 8 The issue of whether and to which degree code-mixing can lead to convergence is not uncontroversial. For instance, Torres Cacoulios and Travis (2018, pp. 160–173) find no convergence between Spanish and English – here, the overt subject pronoun in Spanish as compared to subject-drop – in the vicinity of code-mixing/code-switching as compared to subjects further away from the switch. Apparently, the activation of a bilingual mode is not enough to lead to convergence throughout the data in this particular case. Also see Backus (2005) for a discussion.
- 9 Instead, modal particles such as the German *Abtönungspartikeln* are simply lost in language contact (cf. Salmons, 1990). See Matras (2007, p. 57) for the borrowability hierarchy discourse markers > particles.
- 10 Note that the composite third person plural form has no obvious model in either of the contact languages, i.e., the system has been extended to the third person by analogy. In the second person, the simple form is only used as the polite address form (VOS) today, i.e., functional differentiation has set in.
- 11 The variant *In einem Jahr, unsere Tochter geht ins Internat* is marginally possible.
- 12 Of course, the process is not yet completed, as the enclitic pronouns still inflect for gender and number.

References

- Adamou, E. (2010). Bilingual speech and language ecology in Greek Thrace: Romani and Pomak in contact with Turkish. *Language in Society*, 39(2), pp. 147–171.
- Anders, K. (1993). *Einflüsse der russischen Sprache bei deutschsprachigen Aussiedlern*. Hamburg: Germanisches Seminar.
- Auer, P. (1995). The pragmatics of code-switching: A sequential approach. In: L. Milroy and P. Muysken, eds., *One speaker, two languages*, 1st ed. Cambridge: University Press, pp. 115–135. (Reprinted in: Wei, L., ed. (2007). *The bilingualism reader*, 2nd ed. Milton Park: Routledge, pp. 123–138.)
- Auer, P. (1999). From code-switching via language mixing to fused lects: Toward a dynamic typology of bilingual speech. *International Journal of Bilingualism*, 3(4), pp. 309–332.
- Auer, P. (2014). Language mixing and language fusion: When bilingual talk becomes monolingual. In: J. Besters-Dilger, C. Dermarkar, S. Pfänder, and A. Rabus, eds., *Congruence in contact-induced language change*, 1st ed. Berlin: de Gruyter, pp. 294–336.
- Auer, P. and Maschler, Y., eds. (2016). *NU/NA: A family of discourse markers across the languages of Europe and beyond*. Berlin and Boston: de Gruyter.
- Babel, A. and Pfänder, S. (2014). Doing copying: Why typology doesn’t matter to language speakers. In: J. Besters-Dilger, C. Dermarkar, S. Pfänder, and A. Rabus, eds., *Congruence in contact-induced language change*, 1st ed. Berlin: de Gruyter, pp. 239–257.
- Backus, A. (2005). Codeswitching and language change: One thing leads to another? *International Journal of Bilingualism*, 9(3), pp. 307–340.

- Berger, T. (2014). The convergence of Czech and German between the years 900 and 1500. In: J. Besters-Dilger, C. Dermakar, S. Pfänder, and A. Rabus, eds., *Congruence in contact-induced language change*. Berlin: de Gruyter, pp. 184–200.
- Chamoreau, C. (2012). The Spanish diminutive markers – *ito/-ita* in Mesoamerican languages: A challenge for acceptance of gender distinctions. In: M. Vanhove, Th. Stolz, A. Urdze, and H. Otsuka, eds., *Morphologies in contact*. Berlin: Akademie Verlag, pp. 71–92.
- Demo, Š. (To appear). Artificial fusion: The curious case of Macaronic Latin. In: *International Journal of Bilingualism* (Special Issue on *Language Fusion*, ed., by P. Auer and N. Hakimov).
- Deumert, A. (2009). Namibian Kiche Duits: The making (and decline) of a neo-African language. *Journal of Germanic Linguistics*, 21(4), pp. 349–417.
- Dirim, I. and Auer, P. (2004). *Türkisch sprechen nicht nur die Türken. Über die Unschärfebeziehung zwischen Sprache und Ethnie in Deutschland*. Berlin: de Gruyter.
- Ferguson, Ch.A. and DeBose, Ch.E. (1977). Simplified registers, broken language, and pidginization. In: A. Valdman, ed., *Pidgin and creole linguistics*, 1st ed. Bloomington, IN and London: Indiana University Press, pp. 99–125.
- Fishman, J. (1964). Language maintenance and language shift as a field of inquiry: A definition of the field and suggestions for its further development. *Linguistics*, 2(9), pp. 32–70.
- Fishman, J., Cooper, R.L. and Ma, R. (1971). *Bilingualism in the barrio*. Bloomington, IN: Indiana University Press.
- Franziskus, A. and Gilles, P. (2012). ‘Et le präs direct etikett?’ Non-overlapping repertoires in workplace communication in Luxembourg. *Sociolinguistica: Internationales Jahrbuch für Europäische Soziolinguistik*, 26, pp. 58–71.
- Freywald, U., Mayr, K., Özçelik, T. and Wiese, H. (2011). Kiezdeutsch as a multiethnolect. In: F. Kern and M. Selting, eds., *Ethnic styles of speaking in European metropolitan cities*, 1st ed. Amsterdam: John Benjamins, pp. 45–72.
- Grosjean, F. (2001). The bilingual’s language modes. In: J.L. Nicol, ed., *One mind, two languages: Bilingual language processing*, 1st ed. Malden, MA: Blackwell, pp. 1–22.
- Gumperz, J.J. and Wilson, R. (1971). Convergence and creolization: A case from the Indo-Aryan/Dravidian border. In: D. Hymes, ed., *Pidginization and creolization of languages*, 1st ed. Cambridge: Cambridge University Press, pp. 151–168. Reprinted in: Gumperz J.J. (1971). *Language in social groups. Essays selected and introduced by A. S. Dil*. Stanford, CA: Stanford University Press.
- Heath, J. (1989). *From code-switching to borrowing. A case study of Moroccan Arabic*. London: Kegan Paul International.
- Hilpert, M. (2013). *Constructional change in English*. Cambridge: Cambridge University Press.
- Jahr, E.H. (1997). Language contacts in Northern Europe. In: H. Goebel, P.H. Nelde, Z. Stary, and W. Wölck, eds., *Kontaktlinguistik. Ein internationales Handbuch zeitgenössischer Forschung*, vol. 2, 1st ed. Berlin: de Gruyter, pp. 937–948.
- Kailuweit, R. (2014). Avoiding typological affinity: Negative borrowing as a strategy of Corsican norm finding. In: J. Besters-Dilger, C. Dermakar, S. Pfänder, and A. Rabus, eds., *Congruence in contact-induced language change*, 1st ed. Berlin: de Gruyter, pp. 368–389.
- Kern, F. and Selting, M., eds. (2011). *Ethnic styles of speaking in European metropolitan areas*. Amsterdam: John Benjamins.
- Kheir, A.E. (2018). [manuscript] *To codeswitch or not to codeswitch? Codeswitching and sociopolitical identity among the Druze and Arabs in Israel*.
- Kheir, A.E. (2019). The matrix language turnover hypothesis: The case of the Druze language in Israel. *Journal of Language Contact*, 12(2), pp. 479–512.
- King, R.D. and Beach, S.A. (1998). On the origins of German uvular [R]: The Yiddish evidence. *American Journal of Germanic Linguistics and Literatures*, 10(2), pp. 279–290.
- Klein, W. and Perdue, C. (1997). The basic variety (or: couldn’t natural languages be much simpler?) *Second Language Research*, 13(4), pp. 201–347.
- Kloss, H. (1967). Abstand languages and ausbau languages. *Anthropological Linguistics*, 9, pp. 29–41.

- Kulkarni-Joshi, S. (2016). Forty years of language contact and change in Kupwar: A critical assessment of the intertranslatability model. *Journal of South Asian Languages and Linguistics*, 3(2), pp. 147–174.
- Le Page, R. and Tabouret-Keller, A. (1985). *Acts of identity: Creole-based approaches to language and ethnicity*. Cambridge: University Press.
- Leimgruber, J. (2013). *Singapore English: Structure, variation, and usage*. Cambridge: University Press.
- Luckmann, T. (1986). Grundformen der gesellschaftlichen Vermittlung des Wissens: Kommunikative Gattungen. In: F. Neidhardt, M.R. Lepsius, and J. Weiss, eds., *Kultur und Gesellschaft*, 1st ed. Opladen: Westdeutscher Verlag, pp. 191–211. (Partial English translation in: J. Angermüller, D. Maingueneau, and R. Wodak, eds., *The discourse studies reader. Main currents in theory and analysis*, 1st ed. Amsterdam: John Benjamins, pp. 351–356.
- Maschler, Y. (1998). On the transition from code-switching to a mixed code. In: P. Auer, ed., *Code-switching in conversation. Language, interaction and identity*, 1st ed. London: Routledge, pp. 125–150.
- Matras, Y. (1998). Utterance modifiers and universals of grammatical borrowing. *Linguistics*, 36, pp. 281–331.
- Matras, Y. (1999). Fusion and the cognitive basis for bilingual discourse markers. *International Journal of Bilingualism*, 4, pp. 505–528.
- Matras, Y. (2007). The borrowability of grammatical categories. In: Y. Matras and J. Sakel, eds., *Grammatical borrowing in cross-linguistic perspective*, 1st ed. Berlin: De Gruyter Mouton, pp. 31–74.
- Matras, Y. (2009). *Language contact*. Cambridge: Cambridge University Press.
- Muhamedova, R. (2006). *Untersuchungen zum kasachisch-russischen Code-mixing (mit Ausblick auf den uigurisch-russischen Sprachkontakt)*. München: Lincom.
- Myers-Scotton, C. (1998). A way to dusty death: The matrix language turnover hypothesis. In: L.A. Grenoble and L.J. Whaley, eds., *Endangered languages: Language loss and community response*, 1st ed. Cambridge: Cambridge University Press, pp. 289–316.
- Myers-Scotton, C. (2002). *Contact linguistics. Bilingual encounters and grammatical outcomes*. Oxford: Oxford University Press.
- Oesch-Serra, C. (1998). Discourse connectives in bilingual conversation: The case of an emerging Italian-French mixed code. In: P. Auer, ed., *Code-switching in conversation. Language, interaction and identity*, 1st ed. London: Routledge, pp. 101–124.
- Poplack, S. (2018). *Borrowing – loanwords in the speech community and in the grammar*. New York: Oxford University Press.
- Quist, P. and Svendsen, B.A., eds. (2010). *Multilingual urban Scandinavia*. Bristol, etc.: Multilingual Matters.
- Rehbein, J. (2012). Turkish deixis and its contact-induced change. In: E. Kincses-Nagy and M. Biacsi, eds., *The Szeged conference: Proceedings of the 15th international conference on Turkish linguistics*. Szeged: Department of Altaic Studies, pp. 459–468.
- Rogier, D. (1994). De verspreiding van een social hoogwaardige taalkenmerk: de huig-R rond Gent. *Taal en Tongval*, Special Issue 7, pp. 45–53.
- Rüdiger, S. (2017). *Spoken Korean English*. PhD. University of Bayreuth.
- Schmidt-Rohr, G. (1932). *Die Sprache als Bildnerin der Völker*. Jena: Eugen Diederichs Verlag.
- Siegel, V. (2018). *Multiethnolektale Syntax: Artikel, Präpositionen und Pronomen in der Jugendsprache*. Heidelberg: Winter.
- Şimşek, Y. and Schroeder, C. (2011). Migration und Sprache in Deutschland – am Beispiel der Migranten aus der Türkei und ihrer Kinder und Kindeskinde. In: Ş. Ozil, M. Hofmann und Y. Dayıođlu-Yücel, eds., *50 Jahre türkische Arbeitsmigration in Deutschland*, 1st ed. Göttingen: V & R unipress, pp. 205–228.
- Smith, N. (2000). Symbiotic mixed languages: A question of terminology. *Bilingualism*, 3(2), pp. 122–123.
- Stolt, B. (1964). *Die Sprachmischung in Luthers Tischreden*. Stockholm: Almqvist & Wiksell.
- Thomason, S.G. (2001). *Language contact. An introduction*. Washington: Georgetown University Press.
- Thomason, S.G. (2010). Contact explanations in linguistics. In: R. Hickey, ed., *The handbook of language contact*, 1st ed. Malden, MA: Wiley-Blackwell, pp. 31–47.

- Thomason, S.G. and Kaufmann, T. (1988). *Language contact, creolization, and genetic linguistics*. Berkeley: University of California Press.
- Torres Cacoullos, R. and Travis, C.E. (2018). *Bilingualism in the community. Code-switching and grammars in contact*. Cambridge: Cambridge University Press.
- Trudgill, P. (1974). Linguistic change and diffusion: Description and explanation in sociolinguistic dialect geography. *Language in Society*, 3(2), pp. 215–246.
- Van de Velde, H., Tops, E. and van Hout, R. (2013). The spreading of uvular [R] in Flanders. In: L. Spreafico and A. Vietti, eds., *Rhotics. New data and perspectives*, 1st ed. Bozen and Bolzano: Bolzano University Press, pp. 227–248.
- Weinreich, U. (1963). *Languages in contact*. The Hague: Mouton.
- Zürner, P. (1999). *Sprachinseldialekte: Walserdeutsch im Aostatal*. Aarau: Sauerländer.

Cognitive factors of language contact

Kees de Bot and Lars Bülwö

1. Introduction – cognitive factors and their limits of contribution

Most linguists would agree that language contact can serve as an explanation for linguistic development on the individual level, as well as for language variation and change on the community level. As social beings we are, of course, always in contact with each other and thus with different idiolects, which include different languages, varieties, and styles. Humans only learn and develop languages in social settings, i.e., in contact situations. Therefore, some researchers have argued that language contact is the most important reason for individual language development as well as for language variation and change on the community level. One common consequence of language contact is bilingualism. There is, however, still a need to link the process of language contact with research outcomes on bilingualism at the individual level (Li Wei, 2013, p. 31). This deficit is highlighted by Romaine (2005, p. 49):

Linguists studying language contact often seek to describe changes at the level of linguistic systems in isolation and abstraction from speakers, thus losing sight of the fact that the [...] individual is the ultimate locus of contact.

Although the desideratum is evident, there are relatively few studies in contact linguistics that explicitly deal with cognitive factors (see chapters in Zenner, Backus and Winter-Froemel, 2019; Matras, 2009; Myers-Scotton, 2002; Winford, 2009; Muysken, 2010).

There are two types of cognitive factors that may play a role in language contact, which is defined here as speakers of language (variety) A communicating with speakers of language (variety) B, so contact is seen primarily as an inter-individual process. These types primarily reveal contact as an inter-individual process. The first type relates to cognitive factors that have an impact on language use and interaction more generally, such as aural perception, pragmatics, intentionality, and the general processes of language production and perception. The second type concerns the specific cognitive factors in bilingual and multilingual processing. There is a vast literature on both types; we will deal with only a part of that in our contribution.

This chapter aims to provide an overview of the role of cognitive processing in bilinguals for language contact. More precisely, we try to outline current trends in studying cognitive

factors in bilingualism. Today, bilingualism is considered to be a process rather than a fixed state (De Bot and Houtzager, 2018). From a Complex Dynamic Systems Theory perspective, Larsen-Freeman (2007) has argued that language use and language development cannot be separated because a language can only be developed when it is used. Thus, language use is language change, and using one language will affect other languages that an individual has incorporated. More use of the L1 may lead to a decline of skills in the L2 and *vice versa*. In essence, contact with one language in terms of use will have an impact on the whole language system. Due to the fact that linguistic input is ubiquitous, e.g., watching movies, reading blogs, listening to music, among others, the language system may either adopt new input as part of the system or adapt due to the new input. Contact between languages in an individual speaker can lead to all sorts of processes: language development, but also decline or stagnation (see Section 4).

2. Historical overview

In this part of the chapter, we provide a brief historical overview of cognitive factors in language contact research. The origins of contact linguistics can be traced to the nineteenth century, as testified by the influential works of Hugo Schuchardt (1882) and William Dwight Whitney (1881). Schuchardt and Whitney are known as the founding fathers of Creole and Pidgin Studies, and they intensively studied language mixing in numerous and diverse contexts. Schuchardt associated the problem of language mixing with individual bilingualism. He saw bilingualism as ‘a rather complex problem that can only be clarified in a psychological way’ (Schuchardt, 1882, p. 868). Paul ([1880] 1891) referred to the idea of language mixing (*Sprachmischung*) in his *Prinzipien der Sprachgeschichte*. He pointed out that language mixing is one of the most important questions in linguistics to understand language change and that this question is related to the role of language contact in individuals: ‘The strongest predisposition to mixture is given when there are individuals who speak more than one language; possibly several languages indifferently, but, at least, one other besides their mother-tongue’ (Paul [1880] 1891, p. 467). It is important to note that linguists in the nineteenth century already started to link the phenomena of language contact with the cognitive effects of bilingualism.

The structuralism of the early twentieth century brought forth a mentalist turn in linguistics, and the focus shifted from the individual language user to abstract hypostatized linguistic systems. Studies on language mixing at this time focused mainly on vocabulary, i.e., on lexical borrowing and the existence of loanwords in the language system (e.g., Mencken, 1919). The idea that new linguistic forms emerge when people from different linguistic backgrounds come into contact were not completely forgotten, but it would take almost half a century before the individual bilingual speaker was again the centre of interest in language contact research. Haugen’s (1950, 1953) and Weinreich’s (1953) pioneering studies are most notable in this context in that they marked the beginning of modern contact linguistics as we conceptualize it today (Matras, 2009; Onysko, 2019; Winford, 2003).

Weinreich’s and Haugen’s approaches show a shift in focus – away from the attempt to seek universal explanations in abstract linguistic systems and towards the analysis of multilingual speakers who are embedded in multilingual societies, whose idiolects are characterized by his or her social environment and cognitive prerequisites. Weinreich (1953, p. 1), for example, begins his influential book *Languages in Contact* with the observation that the ‘locus of contact’ is the bilingual individual. Haugen (1953, p. 4) considered the bilingual mind as ‘a veritable laboratory for the interaction of competing linguistic patterns.’ Weinreich and Haugen

were convinced by the need to understand the cognitive effects of bilingualism to gain better insights into language contact. Thus, the study of cognitive factors in language contact was, until today, mainly related to research on cognitive factors in bilingualism.

Weinreich's and Haugen's multilingual backgrounds are certainly crucial, concerning their rather broad approach to bilingualism. Weinreich (1953, p. 1) states: 'it is immaterial whether the two systems are "languages," "dialects of the same language," or "varieties of the same dialect."' Bilinguals, in this sense, use two or more languages and/or varieties alternatively. Thus, Weinreich and Haugen still had a structuralist perspective from which languages were defined as separate entities in both the individual cognitive system and as a communicative tool.

Various studies on cognitive processes in bilinguals in the 1970/80s led to the question: are two or more languages localized in the same or different areas of the brain? More precisely: do different languages inhabit separate entities in the brain, comprising discrete lexicons and grammars, or do bilingual individuals operate on the basis of one unified language system? The assumption that the brains of bilinguals organize languages into separate systems is challenged by recent findings from psycholinguistic studies (see Section 3), but to answer this question, it might be necessary first to ask: what exactly is a language (variety)? Where does a language (variety) start, and where does it end? A full discussion of this issue is beyond the scope of the present chapter, though before moving on it is essential to note that we will limit ourselves to some quotes by Hopper (1998) who argues with respect to the emergence of language that the perceived systematicity in languages is an illusion:

There is no natural fixed structure to language. Rather, speakers borrow heavily from their previous experiences of communication in similar circumstances, on similar topics, and with similar interlocutors. Systematicity, in this view, is an illusion produced by the partial settling or sedimentation of frequently used forms into temporary subsystems.

(Hopper, 1998, pp. 157–158)

Therefore, we also have to reconceive the task of language learning:

Learning a language is not a question of acquiring grammatical structure but of expanding a repertoire of communicative contexts. Consequently, there is no date or age at which the learning of language can be said to be complete. New contexts, and new occasions of negotiation of meaning, occur constantly. A language is not a circumscribed object but a loose confederation of available and overlapping social experiences.

(Hopper, 1998, p. 171)

The question remains how bilinguals can manage to keep their languages separate in the process of speaking and understanding. For a better understanding of these processes, Section 3 addresses one of the leading language processing models that is widely used in research on language production. Levelt's (1989) SPEAKING model, which was launched 30 years ago, is still the leading model and continues to be cited frequently. The model shows how complex the language production process is and how information must be retrieved from various sources to make the production process flow.

Up to this point, we have argued that contact linguistics must account for, rather than simply acknowledge, the cognitive dimension of language contact and that it must be integrated into models of language use and linguistic knowledge. This is easier said than done, as though our methods have improved, we still do not have direct access to the bilingual brain. Possible

answers must be inferred from observable phenomena like language use and experimental data (see Section 5). The fact that we are becoming increasingly aware of how dynamically the brain works and how everything is interconnected adds further complexity to the problem. Insights from cognitive linguistics allow us to identify general cognitive mechanisms; the way they apply in specific contexts, however, remains largely unpredictable (Schmid, 2016, p. 553). This might be one of the main reasons why they have not played a prominent role in empirical studies in contact linguistics so far and why the field has not yet managed to reach its full potential. Luckily, the last two decades have witnessed a wealth of psycholinguistic research regarding language processing and the production of multiple languages in the bilingual mind. Therefore, the following section focuses on the major directions and findings in psycholinguistic research on bilingualism.

3. Critical issues and topics

Research on cognitive factors in language contact is, since the studies of Weinreich and Haugen, empirically connected to research into cognitive factors in bilingualism (Matras, 2009; Muysken, 2010; Paradis, 2004). In what follows, we review key issues in the study of cognitive factors in this field. Bear in mind that bilingualism is a difficult variable to study empirically; it is neither clear to quantify nor does it comprise distinct categories (Barac, Moreno and Bialystok, 2016, p. 1287). Furthermore, the sheer amount of literature makes it next to impossible to provide a complete account of all topics and issues. However, in selecting important strands, focus can be given to (1) how to distinguish different forms of individual bilingualism and (2) how bilingualism affects language production and processing.

Forms of individual bilingualism

Even though Weinreich and Haugen gave a rather broad definition of bilingualism, the term is often associated with balanced bilingualism, i.e., ‘the native-like control of two languages’ (Bloomfield, 1935, p. 56). Most linguists in this field would, however, agree ‘that bilinguals are not simply the sum of two monolinguals in one’ (Matras, 2009, p. 61). Bilinguals may use their languages differently across different contexts. It is well established that many bilinguals have certain preferences of one language over another in particular contexts. Some bilingual children, for example, prefer to speak in one language with their father and in the other with their mother (De Houwer, 2007), or one language might be used at home while a different language is used at school. Furthermore, bilinguals can use two or more languages in one conversation. This practice is not only a strategy to fill gaps in the vocabulary of one language; it can also have a socio-symbolic function. The (covert-)prestige of multilingual practices is well attested for urban multi-ethnolects (Quist, 2008; Rampton, 2011).

The term ‘bilingual,’ then, should not be limited to speakers who have mastered two or more languages to a native level. Instead, it should refer to users who possess multiple competencies in more than one language (see chapters in Cook and Li Wei, 2016). As individuals become bilingual in rather different ways and times in their lives, it is necessary to distinguish groups of bilinguals. One way to separate them would be by their age of onset: researchers in the field differentiate between *early bilinguals* and *late bilinguals* (McLaughlin, 1984). Early bilinguals learn a second language within a period during which children develop their basic linguistic knowledge, beginning just after birth and ending around puberty. Within the group of early bilinguals, we can distinguish between *simultaneous* and *successive* bilingualism. Simultaneous bilinguals acquire two or more languages from birth, while successive bilinguals start

to learn a second language later than their first language. Those who learn a second language after the period of gathering basic grammar and phonology in the first language has passed are *late bilinguals* or *second-language learners* (L2 learning).

The concepts of early and late bilingualism are often associated with the critical period hypothesis (CPH) (Lenneberg, 1967). The CPH, which dates back to the 1950s, claims an ideal window to learn languages is during infancy. This claim is linked to the assertion that the brain loses its plasticity during the process of maturation. At a time when research into cognitive factors in language contact was rare, Penfield and Roberts (1959, p. 236) wrote: ‘for the purpose of learning languages, the human brain becomes increasingly stiff and rigid after the age of nine.’ Following the CPH, many scholars believed that the loss of cerebral plasticity through the process of ageing negatively affects the ability to learn languages (L1 and L2) to a native level after puberty (Lenneberg, 1967, p. 176). This view was supported by early studies on immigrants’ language development which indicated that, in a naturalistic setting, children seem to acquire languages with greater ease than adults, who seem to have greater difficulty in achieving the same level of fluency or accuracy in L2 learning (Asher and Garcia, 1969).

Over time, however, several major problems have emerged regarding the validity and empirical verification of the CPH (Bialystok and Kroll, 2018; Singleton and Pfenninger, 2018). The first problem is that there is no consensus about the onset of the critical period (see Singleton, 2005 and studies in Singleton and Lengyel, 1995), and there seem to be different windows of opportunity for different language domains (DeKeyser, Alfi-Shabtay and Ravid, 2010). Therefore, some linguists prefer to use the term *sensitive period*, i.e., ‘the period during which a child can acquire language easily, rapidly, perfectly and without instruction’ (Richards and Schmidt, 2002, p. 145). The framing of this terminology, however, does not help to define the ideal period. The issue of definition is related to a second problem, which is that the human brain is constantly adapting to new experience and retains plasticity throughout its lifespan (Bialystok and Kroll, 2018; Gutchess, 2014); ‘Even more importantly, many neurobiological studies essentially fail to relate differences in brain activation patterns to differences in target language proficiency’ (Singleton and Pfenninger, 2018, p. 256). There is much evidence to disprove the idea that late acquirers cannot attain native-like L2 mastery. Many behavioural studies show that learning a L2 can be carried out successfully after puberty (e.g., Marinova-Todd, 2003; Kinsella and Singleton, 2014). These studies strongly indicate that socio-affective factors and a range of individual cognitive differences confound the maturational factor (Singleton and Pfenninger, 2018, p. 268).

Therefore, a common way of distinguishing late bilinguals is by the context in which they acquire a L2 or L3. They can learn a language either in an *instructional setting* (e.g., in a classroom) or in a *non-instructional* environment (e.g., at home). The current list of settings in which additional languages can be learned is certainly not exhaustive; for example, it matters whether the L2 is the dominant (and mostly prestigious) language or the non-dominant language in the community. Thus, it is also important to take into account the social setting in which languages are learned and used, to understand the multiple outcomes of language contact (Hoffmann, 2001; Thomason and Kaufman, 1988). These examples illustrate that bilinguals ‘are not the sum of two complete or incomplete monolinguals; rather, he or she has a unique and specific linguistic configuration’ (Grosjean, 1989, p. 259).

A second issue we like to focus on in this section is of how the bi- or multilingual mind manages to process and produce several languages. Considering the plethora of psycholinguistic research into multilingual speech processing and production in the last two decades, we focus on a few selected strands. The most important questions are how bilinguals are able to keep their languages apart in production and perception, how code-switching takes place and

how different languages are organized in the brain. Given the focus of the present handbook, we will limit ourselves to the processing issues.

Language processing in bilinguals

From a cognitive linguistic point of view, two or more languages are not organized in separate systems in the brain (Bybee, 2010; Onysko, 2019, p. 31). Language production and processing are dependent on activity in neuronal networks which have no *a priori* language-specific modules or boundaries; ‘instead, language must be viewed as recruiting processes from the general cognitive system’ (Bialystok et al., 2009, p. 98).

There is no robust evidence that bilinguals vary from monolinguals in neurological processes (Berken et al., 2016; Fabbro, 1999). In dealing with cognitive processing in bilingual speakers there are, however, two major aspects that have to be accounted for (De Bot, 2010, p. 338): (1) how do bilinguals keep their languages apart, and (2) how do they implement language choice? From a psycholinguistic perspective, keeping languages separate and code-mixing are different sides of the same coin.

As to the first question, many proposals have been made to clarify how bilingual speakers keep their languages separate. A vast majority of studies have focused on the cognitive ability (or inability) of early bilinguals to separate their languages (Paradis, 2001). The question is whether bilingual children have one linguistic system (one-system hypothesis), which they gradually separate into two or more systems (Redlinger and Park, 1980), or whether these children have from an early age (by the age of two or earlier) acquired competency in keeping two linguistic systems separate (two-system hypothesis) (Lanza, 1997, p. 1). Early studies support the one-system hypothesis, observing that children lack translation equivalents and struggle to distinguish their languages at an early stage (Redlinger and Park, 1980; McLaughlin, 1984). Today, the two-system hypothesis is supported by many linguists in the field. Deuchar and Quay (2000), for example, argue that mixing primarily occurs due to gaps in children’s vocabulary, and they do not consider the need to keep languages separate a general problem for children. Any gaps in bilingual children’s vocabulary can be explained by the convincing evidence that they command fewer words in each language than their monolingual peers (Bialystok et al., 2009, p. 91). Children also have separate phonological and syntactical systems (Paradis, 2001; Meisel, 1989). In general, bilingual children can keep their languages apart, ‘but the child’s choices might still be influenced by the complexity of changing interlocutors and contexts, resulting in lapses in reacting appropriately to relevant contextual constraints’ (Matras, 2009, p. 66). Children not only pick up two or more linguistic systems but also a pragmatic and communicative competence (Köppe, 1996). ‘For the child growing up bilingually, learning when and when not to code-switch is an important aspect of *language socialization*’ (Lanza, 1997, p. 7).

When two or more systems are established in the mature bilingual brain, the question remains as to how bilinguals keep them apart in language processing. Earlier proposals assume input and output switches. These proposals have been abandoned for suggestions based on activation spreading. Paradis’s (2004) subsystems hypothesis has been particularly influential in this regard. According to Paradis (2004, pp. 210–219), each language in the bilingual mind forms a subset of the larger language neurofunctional system. The linguistic units (words as well as syntactic rules or phonemes) of a given language, therefore, form a subset of the total inventory. The idea is that each subset can be activated independently. Therefore, some subsets (e.g., from typologically related languages as Dutch and German) may have considerable overlap in the form of cognate words. The subsets are maintained and

activated by the use of words in specific settings or contexts. In most contexts, words from one language (variety or style) will be used together, but in settings where code-mixing is expected, speakers may develop a subset in which words from more than one language can be used jointly. An advantage of the subsystems hypothesis is that it does not postulate a mechanism specific to bilinguals that inhibits one of their languages (Paradis, 2004, p. 211). Rather, the activation of a particular subset reduces the set of linguistic elements from which a selection is made.

The subsystems hypothesis is highly compatible with basic ideas from cognitive linguistics and connectionism. The key principle of connectionism is that mental phenomena as language processing can be described by interconnected neuronal networks of uniform units (Roelofs, 1992). One aspect of connectionist models is activation. Activating a subset based on a conversational setting should activate a particular language, but it can also result in enabling a dialect, register, or style. A given bilingual setting, therefore, may also activate a related bilingual subset that allows code-mixing. All subsets can be activated bottom-up (e.g., when a sentence in a language primes a specific subset) and/or top-down (when a speaker selects one or more languages to communicate).

It is important to bear in mind that it is not necessarily a linguistic unit that activates a subset. Bilinguals choose their languages according to a variety of factors, including the type of person being addressed, the location or social setting of the conversation, and the subject matter. Even more complex are the cases in which a bilingual person talks to another bilingual. Code-mixing is a widespread practice among bilinguals and can take many forms (Li Wei, 2013, p. 33). In these settings, many factors unite to define the most appropriate styles of speaking.

Another important aspect is that the vast majority of behavioural and imaging studies show that the languages of a multilingual speaker are always co-activated (Kroll and de Groot, 1997; Martin et al., 2009; Kootstra and Muysken, 2017). Thus, the multilingual speaker always has to inhibit the language resource that is not the target language (Bialystok et al., 2009). We will further elaborate on this in Section 4.

Focus has shifted within the last decades from concerns for whether bilinguals can keep two or more linguistic systems apart towards how bilingual resources influence each other. Growing evidence from psycholinguistic studies underlines the complex nature of interlingual representation and processing in bilinguals (Paradis, 2001; Kroll and de Groot, 1997). Interlingual processing, or rather language contact, might lead to cross-linguistic interaction on all linguistic levels.

All this leads us to the question of how bilinguals implement language choice. To answer this question, we want to apply Levelt's SPEAKING model, the most important language processing model available, and to link it with intra-individual variation. A full treatment of the model is beyond the scope of this chapter, so we will focus on the parts of the model (Levelt (1989) prefers the term 'blueprint') where variation may be found. The model is essentially lexically based, which means that communicative intention is expressed through lexical concepts, which include syntactic information and phonological specification. Lexical concepts that are pre-verbal, in the sense that they are not formulated in normal language, connect information from the conceptualizer to the formulator, thereby matching lexical concepts and lemmas in the lexicon. In production, the lexical concept activates one, and only one lemma (Levelt, p.c. 9/1/2019). All information that is relevant for sentence formation and phonological spell-out has to be coded in such a way that relevant procedures, activated by specifications in the lemma, will incrementally lead to utterances.

Our interest is in the crucial step in the process: the activation of lemmas. In the most rigorous interpretation, matching lexical concepts with lemmas, as well as lemmas with lexemes and other procedures later on in the production process, will activate one specific element. Whether the selected item is the one aimed at is another issue. The spreading activation from one level to another is probably not noise-free. Non-target features may get activated to such an extent that even a small increase in activation may make an element dominant which, according to the adapted version of the SPEAKING model by Levelt, Roelofs and Meyer (1999), is reached when the level of activation of one element is greater than the total activation of all other lexical candidates. However, mis-selection is rare, and speech errors at the lexical level are infrequent compared to phonological ones.

In the research on intrapersonal variation in sociolinguistics and applied linguistics, the fundamental assumption is that it is possible to say the same thing in different ways. There are various reasons why this assumption is wrong. In the previous section, we argued that spreading activation mechanisms leads to the selection of a single candidate. If multiple candidates present themselves and are processed in parallel, the model runs into trouble: what happens to the alternative candidates? Are they processed in parallel, how and when are they siphoned off, and do they continue to have an impact in later stages of production? For the model, the safest option is to assume the selection of just one candidate.

Why is all of this relevant for the current discussion? Variation is at the core of language contact research. Through variation, children can acquire information about what is static and what is variable in language use. Similarly, ‘old-school’ variationist language contact research merely reflects the contribution of various factors mentioned previously as the transmission of a language across generations. However, the confusion that is created by assuming that the same intention can be expressed in different ways is particularly problematic for L2 learners of a language or variety.

The Levelt model presented is essentially a steady-state model, as it can deal with different levels of proficiency in the modelling but provides no acquisitional components or learning mechanisms (Lowie and Verspoor, 2011). There are now numerous models based on Levelt’s model that try to explain these phenomena. De Bot (2010, p. 343), however, concludes:

Within the tradition of which such models are a part, these characteristics may be unproblematic, but in recent years new perspectives on cognition have developed that lead to a different view. The most important development is the emergence of a dynamic perspective on cognition in general and on language processing in particular.

A dynamic conception of cognition would also correspond to observations that the linguistic knowledge of an individual can change over a subject’s lifespan. Taking a dynamic and non-modular approach to language production and processing in bi- or multilinguals assumes that mastering more than one language should also affect other nonverbal cognitive abilities, as language processing is embedded within a more general cognitive system (Bialystok et al., 2009, p. 98). A great deal of recent research shows the cognitive benefits of having multilingual competencies. The next section will discuss these findings.

4. Current contributions and research

Bi- or multilingualism has been long seen as a problem for children, in particular in the educational system. It was a popular belief that learning more than one language would overtax

children as they would have to master a further grammar and vocabulary additional to their monolingual peers. This thinking was not only questioned by Weinreich's and Haugen's studies as outlined in Section 2 but also by results from early and current psycholinguistic experiments.

The bilingual advantage hypothesis

Peal and Lambert (1962) used a battery of intelligence tests (including tests of nonverbal intelligence) to examine bilingual children (in this case, French and English) in Montreal. They found that bilingual children outperformed monolingual children on most of the tests. Bilingual children had a particular advantage on tests requiring symbol manipulation and reorganization. Peal and Lambert's (1962) study, as Bialystok et al. (2009, p. 98) summarize, 'was important in showing both that bilingualism in children might help rather than hinder the development of other abilities and also that language learning may influence nonverbal cognitive processes supporting the view that language is *not* a separate and independent module of mind.' In the last two decades, the quantity of research into bilingual advantage has risen dramatically. In what follows, we briefly summarize how bilingualism affects cognitive function and development.

While there is consensus in the field that bilinguals show evidence of increased metalinguistic awareness (Jessner, 2008), less agreement can be found regarding enhanced executive control for bilinguals when compared to monolinguals. Many studies show that bilinguals outperform monolinguals on several executive control tasks (see review in Bialystok et al., 2009), but some studies have not found such effects (Paap and Greenberg, 2013). Therefore, more research is necessary to fully understand the conditions and mechanisms under which the consequences for bilinguals appear (Barac, Moreno and Bialystok, 2016). Bialystok (2015, p. 118) underlines the importance of research in this field: 'An enhancement of executive function is not trivial: Executive function is a major predictor of academic success and academic success predicts long-term health and well-being.'

Bialystok (2015, p. 119) argues that an enhanced executive control function in bilinguals occurs due to their handling the perpetual conflict of jointly activated languages. This involves a complex process, demanding a variety of components such as inhibition, cognitive flexibility, monitoring, working memory, and bilingual experience. ERP-studies demonstrate that even short-term bilingual exposure in early childhood can alter neural correlates (Conboy and Mills, 2006; Barac, Moreno and Bialystok, 2016). However, handling competition between two or more languages throughout a person's lifespan does not seem to be the overall mechanism that leads to enhanced executive control.

Further insights are supported by two studies on the visual stimulus processing of infants. These studies indicate that general bilingual experience might affect the mechanism that places attention on the language environment (Sebastián-Gallés et al., 2012; Weikum et al., 2007). The infants in these studies were shown mute videos of a person who read a text first in one language and then in another. The monolingual babies did not recognize when the person in the video switched to the other language, but the bilingual infants did. More interestingly, the bilingual infants even noticed a switch when the two languages being spoken were completely different from those they knew (Sebastián-Gallés et al., 2012). The assumption is that the bilingual infants used a more general mechanism for discrimination than simply registering facial features associated with their known languages.

Regarding this finding, Bialystok (2015, p. 120) argues that increased attention and different representational structures seem to be decisive for enhanced executive control in bilinguals.

Bilinguals seem to invest more effort in processing than monolinguals because they must develop two or more representational structures. ‘Once two representational structures are established, executive function is recruited to maintain attention to the target language’ (Bialystok, 2015, p. 120). Barac, Moreno, and Bialystok (2016, p. 1286) conclude ‘that bilingualism provides a powerful form of “brain training” that improves children’s development of executive control.’

Bilingualism and ageing

Although in the last decades much research has been carried out into the effects of bilingualism for children’s linguistic and cognitive abilities, still little is known on the effects of bilingualism for cognitive and linguistic development across a person’s lifespan. In particular, the question of how bilingualism or L2 acquisition affects older adults has been neglected. Because applied linguists have brought attention to the topic in recent years, we will now summarize some of these previous findings.

As outlined in Section 3, the brain retains plasticity across its lifespan while responding to incoming experience. Bilingualism is, as Sullivan and Bialystok (2017, pp. 1–2) point out, ‘an intense experience, possibly more intense than other activities in which we routinely engage, and language representations are deeply intertwined with all knowledge, making them integral to cognition.’ It is, therefore, very likely that learning more than one language, even after puberty, affects cognitive development throughout a person’s life.

What are these effects? Naturally, cognitive decline is unavoidable with ageing. The rate of that decline is, however, not a linear progress; rather, it is highly variable and hard to predict because of the complex interplay of many factors over time.

We have already outlined the effects of bilingualism on children (for a summary see Bialystok et al., 2009). Even brief bilingual exposure can affect neural development in early childhood (Conboy and Mills, 2006). The findings for young adults, however, are less clear. Whereas reaction time studies show no consistent differences between bi- and monolinguals (Paap and Greenberg, 2013), neuroimaging studies reveal significant differences (Bialystok, 2015, p. 4). To date, there have been only a few behavioural and neuroimaging studies of older adults. Gold et al. (2013), for instance, find an improvement in behavioural performance for older bilinguals than for monolinguals (with no effect for young adults). Furthermore, they found evidence for more efficient brain functions in the bilingual group. These findings are in line with results from Bialystok et al. (2004), who report better executive control for older bilinguals compared to older monolinguals. When it comes to age-related effects in verbal tests, behavioural studies are less clear. Mackey and Sachs (2012), for example, examined the impact of inter-individual differences in cognitive capacities on L2 learning success in older adults. Their results indicate that participants’ L1 listening-span capabilities were more predictive of their L2 progress than age.

A growing body of literature has documented that bilingualism helps to maintain cognitive functioning with ageing. These findings point to a concept called ‘cognitive reserve’ (Stern, 2002). It is argued that the use of multiple languages builds the cognitive reserve of older individuals, which protects them from cognitive decline. Many studies point to findings that indicate that bilinguals show a delay in the onset of dementia and Alzheimer’s disease (Woumans et al., 2015; Alladi et al., 2013). Pfenninger and Singleton (2019) conclude that the ‘long-term active use of two languages seems to be neuroprotective’; they insist on filling in this empirical gap with robust results. Few studies examine the interdependency between L2 learning in older adults and the effects on their cognition. Pfenninger

and Singleton (2019) ask for ‘denser time serial measurements in order to focus on micro-development and idiodynamic trajectories, capturing and capitalizing on individuality in the L2 learning process, homing in on variability, non-linear development, and progression in iterative steps.’ This approach demands a focus on the process of intra-individual development.

Something must also be said about how bilingualism influences language usage across a person’s lifespan. There is a consensus that acquiring another language, at whatever time in a person’s life, will affect his or her native language(s). There is no doubt that a large degree of variation and intra-individual change in the L1 can be attributed to learning and using the L2 (Bice and Kroll, 2015; Chang, 2013). This is, however, not a one-way street; exposure in the L2 interacts in a complex manner with the L1 and *vice versa*.

Assumptions and conceptions about how bilingual experience affects cognitive development have thus fundamentally changed over the last decades. Being bilingual seems to give a person a cognitive advantage in old age. Nevertheless, more robust empirical results are needed to validate and further develop our knowledge of the cognitive effects of bilingualism in language contact.

5. Main research methods

To develop the CDST perspective on language contact further, we need longitudinal and dense data on different time scales. Historical linguistics can provide data on change in the language system over decades and centuries. Research on L2 development as mentioned earlier has used longitudinal data from specific elements of the target language. Examples are Chan, Lowie and de Bot (2014) and Lowie, Plat and de Bot (2014). Chan, Lowie and de Bot (2014) used data on a year-long set of weekly written and spoken texts in Chinese and English by bilingual twins in Taiwan. They showed that despite the fact these were monozygotic twins growing up in exactly the same environment, their individual paths of development were remarkably dissimilar. There were no discernible sociolinguistic factors that could explain this variation, which seems to result from the interaction of variables over time and leads to complex patterns in development.

Lowie, Plat and de Bot (2014) present data from a four-year longitudinal study on L1 and L2 development using psycholinguistic methods, in particular, word naming in English and Dutch. The participant carried out a word naming task of 200 words in the two languages, once in the morning and once in the late afternoon. Exactly the same task with the same stimuli and the same equipment was carried out 44 times over a four-year period, with irregular intervals. The data were analyzed using a ‘pink noise’ approach, which refers to changes in the degree of autonomy and control in such a highly controlled task. The study showed a systematic difference between morning and afternoon sessions, suggesting an early chrono-type of the consultant, a systematic advantage for the L1 even after 44 repetitions of the same stimuli, as well as an effect on the noise pattern of more or less contact with the two languages.

Research on language contact has typically been carried out in settings in which individuals with different languages come into contact. So, what are in contact are therefore not so much languages as systems, but the individuals who are using them. Therefore, case studies of individuals in which languages meet are essential. The manner in which changes in the system spread among community members is mainly a sociolinguistic matter, to which cognitive approaches as yet have little to contribute.

6. Future directions

One of the main new lines of research on cognitive factors of language contact is how a given language, i.e., a dynamic system, develops over time. Dynamic systems typically adapt due to incorporating and adopting new elements and due to internal reorganization. As mentioned before, we know little about which parts of the input are processed psycholinguistically, or why and how language changes due to internal reorganization. Contact between languages through L2 learning can lead to all sorts of outcomes. Language development is one, but we can also better understand language decline and stagnation (otherwise known as language attrition) through language contact. Language attrition refers to the loss or decrease of a first- or later-learned language, and it can happen at any point in an individual's lifespan (Schmid, Köpke and de Bot, 2013). Reasons for attrition might be a lack of contact with speakers of the language over a longer period of time or cognitive decline in elderly bilinguals (Li Wei, 2013, p. 34). The assumption is that lack of contact with speakers of a given language over a longer period of time leads to a decline in linguistic skills. However, recent research on language attrition as presented in the *Handbook of Language Attrition* (Schmid and Köpke, 2019) casts doubt on the adage 'Use it or lose it.' Language attrition in the elderly is another matter. The factors cognitive ageing, change on environment, and type, and amount of input and decline of peripheral skills interact over time, and their contribution is hard to measure.

Another issue is, to what extent does historical data provide the data needed for a dynamic systems approach, for which data at different time scales are essential? Variation is an essential element of all dynamically developing systems, but variation has been shown to be variable itself. From a sociolinguistic perspective, an interesting point of analysis is how changes can move from the individual level to the population level (Tagliamonte, 2012, p. 247).

Language contact research is typically done with groups and is, therefore, based on averages. In the research on variation in the CDST tradition, interest primarily concerns individual developmental paths. The question is, to what extent do group studies allow for generalization from the sample to the population and for ergodicity which focuses on the relation from the group to the individual.

Finally, much of the work on language development in the CDST line tries to model the development of individual trajectories based on assumed or proven contributions of specific factors in the equations that describe and predict distinct developmental trajectories (Verspoor, de Bot and Lowie, 2011).

7. Further reading

There is a vast literature on the cognitive aspects of multilingualism that borders on research on language contact from a psycholinguistic perspective. In recent years several handbooks on bilingualism and multilingualism have been published. These include de Groot (2011), Kroll and de Groot (2005), de Houwer and Ortega (2018), and Altarriba and Heredia (2018). In handbooks numerous articles are relevant for the topics dealt with in the current chapter.

Aalberse, S. and Muysken, P. (2018). Bilingualism and language contact. In A. de Houwer and L. Ortega, eds., *The Cambridge handbook of bilingualism*, 1st ed. Cambridge: Cambridge University Press, pp. 524–543.

This chapter focuses on the notion of language ecology which means that 'the setting in which languages are spoken affects the type of changes we observe in them' (Aalberse and Muysken, 2018, p. 542). The authors present a number of examples from Dutch and Mandarin Chinese to make their point. There clearly is a need to consider linguistic factors in addition to psycho- and sociolinguistic ones.

Higby, E. and Obler, L. (2015). Losing a first language to a second language. In: J. Schwieter, ed., *The Cambridge handbook of bilingual processing*, 1st ed. Cambridge: Cambridge University Press, pp. 645–664.

In research on language contact, language attrition is now becoming a recognized part of the field. In this article the authors discuss the large area of first language loss and its replacement by a second language. This refers to an ongoing discussion on what has been called the balance effect: does the growth of one language go at the expense of the other languages of a multilingual. There appear to be a number of studies that tackle this topic, but the analyses show that it is hard to get real evidence to support the balance hypothesis. It would be interesting to see how the individual factors the authors mention would be applicable to language change at the community level.

Spivey, M. and Cardon, C. (2015). Methods for studying adult bilingualism. In: J. Schwieter, ed., *The Cambridge handbook of bilingual processing*, 1st ed. Cambridge: Cambridge University Press, pp. 108–132.

This article is a passionate call for the use of recently developed methods of data gathering and analysis in cognitive science to study bilinguals and their language use. The authors argue that most linguists stick to traditional approaches of language data, while new and highly relevant methods are becoming available. Their suggestions are also relevant for the study of the impact of psycholinguistic and neurolinguistics data on language contact theories.

8. Related topics

Usage-based approaches, bilingual language acquisition, L1 attrition in the twenty-first century, language contact in the lab, and code-switching

Abbreviations

| | |
|------|--------------------------------|
| CDST | complex dynamic systems theory |
| CPH | critical period hypothesis |
| L1 | first language |
| L2 | second language |
| L3 | third language |

References

- Alladi, S., Bak, T.H., Duggirala, V., Surampudi, B., Shailaja, M., Shukla, A.K., Chaudhuri, J.R. and Kaul, S. (2013). Bilingualism delays age at onset of dementia, independent of education and immigration status. *Neurology*, 81(22), pp. 1938–1944.
- Altarriba, J. and Heredia, R.R., eds. (2018). *An introduction to bilingualism: Principles and processes*. New York: Routledge.
- Asher, J.J. and Garcia, R. (1969). The optimal age to learn a foreign language. *The Modern Language Journal*, 53(5), pp. 334–341.
- Barac, R., Moreno, S. and Bialystok, E. (2016). Behavioral and electrophysiological differences in executive control between monolingual and bilingual children. *Child Development*, 87(4), pp. 1277–1290.
- Berken, J.A., Chai, X., Chen, J.-K., Gracco, V.L. and Klein, D. (2016). Effects of early and late bilingualism on resting-state functional connectivity. *Journal of Neuroscience*, 36(4), pp. 1165–1172.
- Bialystok, E. (2015). Bilingualism and the development of executive function: The role of attention. *Child Development Perspectives*, 9(2), pp. 117–121.
- Bialystok, E., Craik, F.I.M., Green, D.W. and Gollan, T.H. (2009). Bilingual minds. *Psychological Science in the Public Interest*, 10(3), pp. 89–129.

- Bialystok, E., Craik, F.I.M., Klein, R. and Viswanathan, M. (2004). Bilingualism, aging, and cognitive control: Evidence from the Simon task. *Psychology and Aging*, 19(2), pp. 290–303.
- Bialystok, E. and Kroll, J.F. (2018). Can the critical period be saved? A bilingual perspective. *Bilingualism: Language and Cognition*, 21(5), pp. 908–910.
- Bice, K. and Kroll, J.F. (2015). Native language change during early stages of second language learning. *NeuroReport*, 26(16), pp. 966–971.
- Bloomfield, L. (1935). *Language*. London: George Allen and Unwin.
- Bybee, J. (2010). *Language, usage and cognition*. Cambridge: Cambridge University Press.
- Chan, H-P., Lowie, W. and de Bot, K. (2014). A case study of lexical development of writing and speaking in identical twins. In: J. Gao, V. Allegranzi, M. Rivière, L. Sauwala, C. Simon, and L. Xue, eds., *Actes des 16èmes Rencontres Jeunes Chercheurs en Sciences du Langage: Modèles et Modélisation dans les Sciences du Langage*, 1st ed. Paris: HAL-SHS, pp. 54–65.
- Chang, C.B. (2013). A novelty effect in phonetic drift of the native language. *Journal of Phonetics*, 41(6), pp. 520–533.
- Conboy, B.T. and Mills, D.L. (2006). Two languages, one developing brain: Event-related potentials to words in bilingual toddlers. *Developmental Science*, 9(1), F1–F12.
- Cook, V. and Li Wei., eds. (2016). *The Cambridge handbook of linguistic multi-competence*. Cambridge: Cambridge University Press.
- De Bot, K. (2010). Cognitive processing in bilinguals: From static to dynamic models. In: R.B. Kaplan, ed., *The Oxford handbook of applied linguistics*, 2nd ed. Oxford: Oxford University Press, pp. 348–335.
- De Bot, K. and Houtzager, N. (2018). Multilingualism processing and aging. In: J. Altarriba and R.R. Heredia, eds., *An introduction to bilingualism: Principles and processes*. New York: Routledge, pp. 93–106.
- De Groot, A.M.B. (2011). *Language and cognition in bilinguals and multilinguals: An introduction*. New York: Psychology Press.
- De Houwer, A. (2007). Parental language input patterns and children’s bilingual use. *Applied Psycholinguistics*, 28, pp. 411–424.
- De Houwer, A. and Ortega, L., eds. (2018). *The Cambridge handbook of bilingualism*. Cambridge: Cambridge University Press.
- DeKeyser, R.M., Alfi-Shabtay, I. and Ravid, D. (2010). Cross-linguistic evidence for the nature of age effects in second language acquisition. *Applied Psycholinguistics*, 31(3), pp. 413–438.
- Deuchar, M. and Quay, S. (2000). *Bilingual acquisition: Theoretical implications of a case study*. Oxford: Oxford University Press.
- Fabbro, F. (1999). *Neurolinguistics of bilingualism: An introduction*. Hove, New York: Psychology Press.
- Gold, B.T., Kim, C., Johnson, N.F., Kryscio, R.J. and Smith, C.D. (2013). Lifelong bilingualism maintains neural efficiency for cognitive control in aging. *Journal of Neuroscience*, 33(2), pp. 387–396.
- Grosjean, F. (1989). Neurolinguists, beware! The bilingual is not two monolinguals in one person. *Brain and Language*, 36(1), pp. 3–15.
- Gutchess, A. (2014). Plasticity of the aging brain: New directions in cognitive neuroscience. *Science*, 346(6209), pp. 579–582.
- Haugen, E. (1950). The analysis of linguistic borrowing. *Language*, 26(2), pp. 210–231.
- Haugen, E. (1953). *The Norwegian language in America: A study in bilingual behavior. vol. I. The bilingual community*. Philadelphia: University of Pennsylvania Press.
- Hoffmann, C. (2001). Towards a description of trilingual competence. *The International Journal of Bilingualism*, 5(1), pp. 1–17.
- Hopper, P. (1998). Emergent grammar. In: M. Tomasello, ed., *The new psychology of language: Cognitive and functional approaches to linguistic structure*, vol. I, 1st ed. Mahwah, NJ: Erlbaum, pp. 155–175.
- Jessner, U. (2008). Teaching third languages: Findings, trends, and challenges. *Language Teaching*, 41(1), pp. 15–56.
- Kinsella, C. and Singleton, D. (2014). Much more than age. *Applied Linguistics*, 35(4), pp. 441–462.

- Kootstra, G.J. and Muysken, P. (2017). Cross-linguistic priming in bilinguals: Multidisciplinary perspectives on language processing, acquisition, and change. *Bilingualism: Language and Cognition*, 20(2), pp. 215–218.
- Köppe, R. (1996). Language differentiation in bilingual children: The development of grammatical and pragmatic competence. *Linguistics*, 34(5), pp. 471–504.
- Kroll, J.F. and De Groot, A.M.B. (1997). Lexical and conceptual memory in the bilingual: Mapping form to meaning in two languages. In: A.M.B. de Groot and J.F. Kroll, eds., *Tutorials in bilingualism: Psycholinguistic perspectives*, 1st ed. Mahwah, NJ: Erlbaum, pp. 169–199.
- Kroll, J.F. and De Groot, A.M.B., eds. (2005). *Handbook of bilingualism: Psycholinguistic approaches*. Oxford: Oxford University Press.
- Lanza, E. (1997). *Language mixing in infant bilingualism: A sociolinguistic perspective*. Oxford: Clarendon Press.
- Larsen-Freeman, D. (2007). On the complementarity of chaos/complexity theory and dynamic systems theory in understanding the second language acquisition process. *Bilingualism: Language and Cognition*, 10(1), pp. 35–37.
- Lenneberg, E.H. (1967). *Biological foundations of language*. New York: Wiley.
- Levelt, W.J.M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
- Levelt, W.J.M. (2019). Interview. Conducted by Kees de Bot, Amsterdam 9/1/2019.
- Levelt, W.J.M., Roelofs, A. and Meyer, A. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences*, 22(1), pp. 1–75.
- Li Wei. (2013). Conceptual and methodological issues in bilingualism and multilingualism research. In: T.K. Bhatia and W.C. Ritchie, eds., *The handbook of bilingualism and multilingualism*, 2nd ed. Malden, MA: Wiley-Blackwell, pp. 26–51.
- Lowie, W., Plat, R. and de Bot, K. (2014). Pink noise in language production: A nonlinear approach to the multilingual lexicon. *Ecological Psychology*, 2(3), pp. 216–228.
- Lowie, W. and Verspoor, M. (2011). The dynamics of multilingualism: Levelt’s speaking model revisited. In: M. Schmid and W. Lowie, eds., *Modeling bilingualism: From structure to chaos*, 1st ed. Amsterdam and Philadelphia: John Benjamins, pp. 267–288.
- Mackey, A. and Sachs, R. (2012). Older learners in SLA research: A first look at working memory, feedback, and L2 development. *Language Learning*, 62(3), pp. 704–740.
- Marinova-Todd, S.H. (2003). *Comprehensive analysis of ultimate attainment in adult second language acquisition*. PhD. Harvard University.
- Martin, C.D., Dering, B., Thomas, E.M. and Thierry, G. (2009). Brain potentials reveal semantic priming in both the ‘active’ and the ‘non-attended’ language of early bilinguals. *NeuroImage*, 47(1), pp. 326–333.
- Matras, Y. (2009). *Language contact*. Cambridge: Cambridge University Press.
- McLaughlin, B. (1984). *Second language acquisition in childhood. Volume I. Preschool children*, 2nd ed. Hillsdale, NJ: Erlbaum.
- Meisel, J.M. (1989). Early differentiation of languages in bilingual children. In: K. Hyltenstam and L.K. Obler, eds., *Bilingualism across the lifespan: Aspects of acquisition, maturity and loss*, 1st ed. Cambridge: Cambridge University Press, pp. 13–40.
- Mencken, H.L. (1919). *The American language: An inquiry into the development of English in the United States*. New York: Alfred A. Knopf.
- Muysken, P. (2010). Scenarios for language contact. In: R. Hickey, ed., *The handbook of language contact*. Oxford: Blackwell, pp. 265–281.
- Myers-Scotton, C. (2002). *Contact linguistics: Bilingual encounters and grammatical outcomes*. Oxford: Oxford University Press.
- Onysko, A. (2019). Reconceptualizing language contact phenomena as cognitive processes. In: E. Zenner, A. Backus and E. Winter-Froemel, eds., *Cognitive contact linguistics: Placing usage, meaning and mind at the core of contact-induced variation and change*, 1st ed. Berlin: de Gruyter, pp. 23–50.
- Paap, K.R. and Greenberg, Z.I. (2013). There is no coherent evidence for a bilingual advantage in executive processing. *Cognitive Psychology*, 66(2), pp. 232–258.

- Paradis, J. (2001). Do bilingual two-year-olds have separate phonological systems? *International Journal of Bilingualism*, 5(1), pp. 19–38.
- Paradis, M. (2004). *A neurolinguistics theory in bilingualism*. Amsterdam: John Benjamins.
- Paul, H. ([1880] 1891). *Principles of the history of language*. London: Longmans.
- Peal, E. and Lambert, W. (1962). The relation of bilingualism to intelligence. *Psychological Monographs: General and Applied*, 76(27), pp. 1–23.
- Penfield, W. and Roberts, L. (1959). *Speech and brain mechanisms*. Princeton, NJ: Princeton University Press.
- Pfenninger, S. and Singleton, D. (2019). A critical review of research relating to the learning, use and effects of additional and multiple languages in later life. *Language Teaching*, 52(4), pp. 419–449.
- Quist, P. (2008). Sociolinguistic approaches to multiethnolect: Language variety and stylistic practice. *International Journal of Bilingualism*, 12(1–2), pp. 43–61.
- Rampton, B. (2011). From multi-ethnic adolescent ‘heteroglossia’ to ‘contemporary urban vernaculars’. *Language & Communication*, 31(4), pp. 276–294.
- Redlinger, W.E. and Park, T.-Z. (1980). Language mixing in young bilingual children. *Journal of Child Language*, 7(2), pp. 337–352.
- Richards, J.C. and Schmidt, R. (2002). *Longman dictionary of language teaching and applied linguistics*. 3rd ed. Harlow: Pearson.
- Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition*, 42(1–3), pp. 107–142.
- Romaine, S. (2005). Language-contact studies. In: U. Ammon, N. Dittmar, K. Mattheier and P. Trudgill, eds., *Sociolinguistics/Soziolinguistik. An international handbook of the science of language and society*, 1st ed. Berlin: de Gruyter, pp. 49–58.
- Schmid, H.-J. (2016). Why cognitive linguistics must embrace the social and pragmatic dimensions of language and how it could do so more seriously. *Cognitive Linguistics*, 27(4), pp. 543–557.
- Schmid, M.S. and Köpke, B., eds. (2019). *The Oxford handbook of language attrition*. Oxford: Oxford University Press.
- Schmid, M.S., Köpke, B. and de Bot, K. (2013). Language attrition as a complex, non-linear development. *International Journal of Bilingualism*, 17(6), pp. 675–682.
- Schuchardt, H. (1882). Zur afrikanischen Sprachmischung. *Das Ausland. Wochenschrift für Länder- und Völkerkunde*, 55, pp. 867–869.
- Sebastián-Gallés, N., Albareda-Castellot, B., Weikum, W.M. and Werker, J.F. (2012). A bilingual advantage in visual language discrimination in infancy. *Psychological Science*, 23(9), pp. 994–999.
- Singleton, D. (2005). The critical period hypothesis: A coat of many colours. *International Review of Applied Linguistics*, 43(4), pp. 269–285.
- Singleton, D. and Lengyel, Z., eds. (1995). *The age factor in second language acquisition*. Clevedon: Multilingual Matters.
- Singleton, D. and Pfenninger, S. (2018). L2 acquisition in childhood, adulthood and old age: Misreported and under-researched dimensions of the age factor. *Journal of Second Language Studies*, 1(2), pp. 254–275.
- Stern, Y. (2002). What is cognitive reserve? Theory and research application of the reserve concept. *Journal of the International Neuropsychological Society*, 8(3), pp. 448–460.
- Sullivan, M.D. and Bialystok, E. (2017). The importance of bilingualism for the aging brain. In: E. Bialystok and M.D. Sullivan, eds., *Growing old with two languages: Effects of bilingualism on cognitive aging*, 1st ed. Amsterdam: John Benjamins, pp. 1–8.
- Tagliamonte, S.A. (2012). *Variationist sociolinguistics: Change, observation, interpretation*. Malden, MA: Wiley-Blackwell.
- Thomason, S.G. and Kaufman, T. (1988). *Language contact, creolization and genetic linguistics*. Berkeley: University of California Press.
- Verspoor, M., de Bot, K. and Lowie, W. (2011). *A dynamic approach to second language development. Methods and techniques*. Amsterdam: John Benjamins.
- Weikum, W.M., Vouloumanos, A., Navarra, J., Soto-Faraco, S., Sebastián-Gallés, N. and Werker, J.F. (2007). Visual language discrimination in infancy. *Science*, 316(5828), p. 1159.

- Weinreich, U. (1953). *Languages in contact*. The Hague: Mouton.
- Whitney, W.D. (1881). On mixture in language. *Transactions of the American Philological Association*, 12, pp. 5–26.
- Winford, D. (2003). *An introduction to contact linguistics*. Malden, MA: Wiley-Blackwell.
- Winford, D. (2009). On the unity of contact phenomena and their underlying mechanisms. In: L. Isurin, D. Winford, and K. de Bot, eds., *Multidisciplinary approaches to code switching*, 1st ed. Amsterdam: John Benjamins, pp. 279–305.
- Woumans, E., Santens, P., Sieben, A., Versijpt, J., Stevens, M. and Duyck, W. (2015). Bilingualism delays clinical manifestation of Alzheimer’s disease. *Bilingualism: Language and Cognition*, 18(3), pp. 568–574.
- Zenner, E., Backus, A. and Winter-Froemel, E., eds. (2019). *Cognitive contact linguistics: Placing usage, meaning and mind at the core of contact-induced variation and change*. Berlin: de Gruyter.

Typological factors

Felicity Meakins

1. Introduction

Typological factors have been argued to provide more information about the shape of language contact outcomes than social theories, which only tend to offer explanations for the broader motivations of code-switching, borrowing, and the formation of mixed languages, such as reasons for the practice itself and the choice of matrix language (Backus, 2003; Meakins, 2011b; Pfaff, 1979). The structural profile of morphemes and the typological match of languages in contact are considered crucial factors in the resultant patterns of language mixing. This idea has its roots in Weinreich (1974 [1953]) who considered the borrowability of morphemes to be dependent on their inherent formal properties (e.g., bound/free, lexical/grammatical, agglutinative/polysynthetic/isolating) and typological equivalences between languages in contact. Areas where typological equivalences are relevant include word and morpheme order, the existence of a particular linguistic or functional category in both languages and matches in terms of the expression of that category e.g., whether negation is expressed by a particle, auxiliary, clitic, or morpheme in both languages.

Since Weinreich, the typological profile of morphemes has been shown to be important in predicting which morphemes will transfer¹ or be resistant to transfer in situations of borrowing, code-switching, and mixed language formation, for example Poplack's (1980) Equivalence Constraint, Field's (2002) Principle of System Incompatibility, and Myers-Scotton's (2002) 4-M model within the Matrix Language Frame (MLF) model of code-switching. Typological congruence between languages has also been shown to be important in theories of language contact, for example Myers-Scotton's (1993a) Blocking Hypothesis, Sebba's (1998) Principle of Categorial Congruence and McConvell's (2008) Centre of Gravity Hypothesis. This chapter will review the literature on the role of typology in borrowing, code-switching, and mixed language genesis (Section 2 and Section 3); and the role of contact in typological change and theory (Section 4), with suggestions for future directions (Section 5).

2. Historical overview

The role of typology in borrowing

A dominant theory of borrowing has been Thomason and Kaufman's (1988) social motivations approach which claims that social factors play a fundamental and determinate role in the linguistic outcomes of language contact, and that, given the right level of social disruption, substantial structural borrowing is not unusual. Under Thomason and Kaufman's model, two social features are necessary for extensive borrowing – time and a level of bilingualism. Thus, extensive and prolonged community bilingualism is considered a necessary condition for borrowing structural elements of a language, such as inflectional morphology, which are considered highly resistant to transfer.

While Thomason and Kaufman's theory accounts for why borrowing occurs in the first place, it does not account for many long observed cross-linguistic tendencies of languages in contact, such as the fact that nouns are more likely to be borrowed than verbs, uninflected words transfer more readily than morphology, and borrowed inflectional morphology is vanishingly rare (Whitney, 1881; Haugen, 1950; Moravcsik, 1978; Muysken, 1981; Singh, 1982; Aikhenvald and Dixon, 2007; Matras and Sakel, 2007; Matras, 2007; Weinreich, 1974 [1953]; Winford, 2003; Gardani, Arkadiev and Amiridze, 2015). This section outlines cross-linguistic patterns of borrowing and presents the typological explanations which are given for these cross-linguistic patterns in the borrowing literature.

Most cross-linguistic patterns of transferability have produced borrowing hierarchies which largely replicate and provide more detail to the early findings of Whitney (1881). Whitney observed that nouns were the most susceptible to borrowing, followed by other parts of speech, then suffixes, inflections, and finally sounds. Haugen (1950, p. 224) conducted a study of borrowing in Norwegian and Swedish in the United States and found that nouns were the least resistant to borrowing, followed by verbs, adjectives, and interjections. He did not include morphology on this scale, however he concluded that 'the more structural a feature is, the less likely it is to be borrowed' (Haugen, 1950, p. 225). Singh's (1982) study of English borrowings into Hindi also produced a similar hierarchy: nouns > adjectives > verbs > prepositions, as did Muysken's (1981) study of Spanish borrowings into Quechua. Muysken's scale is more detailed, but reflects the patterns of previous scales, with lexical elements dominating the heavily borrowed end of the scale: nouns > adjectives > verbs > prepositions > coordinating conjunctions > quantifiers > determiners > free pronouns > clitic pronouns > subordinating conjunctions.

Thomason and Kaufman's (1988, pp. 74–75) borrowability scale is similar, but also includes what has since been called by Matras and Sakel (2007) PAT(tern) borrowing, i.e., changes to distributional patterns in a clause. As with other scales, Thomason and Kaufman also include word and morpheme forms, referred to as MAT(ter) borrowings by Matras and Sakel (2007). Thus, Thomason and Kaufman propose the following scale: non-basic vocab > basic > functional vocab, e.g., conjunctions and adverbs > pre/postpositions, derivational affixes, inflectional affixes > word order, borrowing postpositions in a prepositional language > extensive word order change, inflectional affixes (e.g., case) > typological disruption, changes in word structure (e.g., adding prefixes in suffixing language), change from flexional to agglutinative morphology. Finally, other studies provide more detail for sections of borrowability scales. Matras (2011, p. 216) has refined the borrowability scale in terms of conjunctions, observing the following hierarchy: 'but' > 'or' > 'and, only' > 'too.' Matras (2009, 1998, 2007) also

contains several hierarchies that are linked through a cognitive trigger hypothesis rather than a structural-typological one.

Other borrowing scales explicitly posit an implicational relationship between categories where one linguistic category can only be borrowed if another has already been borrowed. An example of this type of scale comes from Moravcsik (1978, pp. 110–112) who suggests that non-lexical items will not be borrowed unless some lexical items have already been borrowed (lexical > functional), borrowed lexical items such as verbs will only be observed in a language if borrowed nouns are already present (nouns > other lexical items), and that ‘no inflectional affixes can belong to the set of properties borrowed from a language unless at least one derivational affix also belongs to the set’ (derivational > inflectional) (Moravcsik, 1978, p. 112). Nonetheless a number of other case studies of inflectional borrowings can be found (Gardani, 2008; Gardani, Arkadiev and Amiridze, 2015; Meakins, 2011a; Seifart, 2013).

As well as the differential behaviour of morphemes in borrowing, structural constraints are said to affect the ability of morphemes to transfer from one language to another. These constraints are based on the typology of the languages in contact or on the degree of syntactic involvement a morpheme has in the clause. Early work by Weinreich (1974 [1953]) provides the first typological model for borrowing, which has mechanisms whereby morphemes may be transferred. Weinreich (1974 [1953], p. 35) frames borrowing scales in terms of the morpheme’s degree of structural integration, but also goes further by considering borrowing not merely as a function of the inherent ‘borrowability’ of a morpheme, but as the compatibility of both languages in the borrowing relationship.

(T)he transferability of morphemes is considered as a correlate of their grammatical function in the source language and the resistance of the recipient language.

(Weinreich, 1974 [1953], p. 31)

Weinreich presents a number of factors which affect the likelihood of the transfer of functional items. Firstly, he suggests that if the structures of the source and recipient languages are congruent, then transfer is strongly facilitated (Weinreich, 1974 [1953], pp. 32–33). He also suggests that overt morphemes tend to replace zero morphemes (Weinreich, 1974 [1953], p. 33), and the ‘relatively unbound morpheme is most likely to replace its counterpart in another language if the latter is more bound and is involved in a greater variation of alternates’ (Weinreich, 1974 [1953], p. 34). The last two suggestions relate to previous observations which have led to borrowing scales. Morphemes with complex functions are less likely to be borrowed than those which have simpler and more transparent functions (Weinreich, 1974 [1953], p. 34). This means that, due to their opaque nature, functional morphemes are less likely to be borrowed than morphemes with lexical content. Moreover, Weinreich suggests that ‘the fuller the integration of the morpheme, the less likelihood of its transfer’ (Weinreich, 1974 [1953], p. 35). In other words, the more salient the morpheme boundary, the easier it is to borrow.

Weinreich’s model excludes the possibility of borrowing inflectional morphology, yet cases do occur. Indeed Heath’s (1978, pp. 105–107) observations of case suffix transfer in northern Australian prompted him to suggest some alternative factors which affect the ‘borrowability’ of inflectional morphology.

- 1 Morpheme syllabicity (morphemes that are independently pronounceable);
- 2 Sharpness of boundaries between morphemes;
- 3 Unifunctionality of morphemes (e.g., not portmanteau morphs);

- 4 Categorical clarity of morphemes (broader environment is not required to discern function);
- 5 Analogical freedom from other morphemic systems in the same language.

Winford (2003, p. 92) reconciles Weinreich and Heath's positions by suggesting that they can be subsumed into three general categories where constraints for borrowing are based on:

- 1 Congruency of morphological structures;
- 2 Transparency/markedness;
- 3 Functional considerations.

The first constraint follows Weinreich's proposal that borrowing is a function of the relationship between the source and recipient languages. If the structures of the two languages correspond typologically, then this similarity will facilitate a direct mapping of morphemes from one language to the other (Winford, 2003, p. 93). Conversely, typological distance inhibits the transfer of inflectional morphology. Field (2002, pp. 41–44) frames this constraint as the Principle of System In/compatibility and it is not dissimilar to Sebba's (1998) notion of Categorical Congruence in the code-switching literature. On the other hand, Gardani (2008) suggests that the degree of syntactic integration is the relevant factor in determining the borrowability of inflectional morphology – more so than the typological match of the languages in contact. This proposal relates to Heath's (1978) fifth factor 'analogical freedom from other morphemic systems in the same language.' Gardani (2008) hypothesizes that the degree to which an inflectional morpheme interacts with other parts of the grammar may hold some clues as to the differences in transferability, with morphemes less reliant on the syntax such as gender and TAM marking, more easily borrowed (see also Section 3). Myers-Scotton has a similar intuition in her 4-M model of code-switching, as will be discussed next. The proposal is largely supported by the data, however a very few cases of contextual inflection transfer have been observed, for example structural case-marking as will be discussed later. Finally, Field (2002) also suggests that semantic congruence is an important factor in borrowability.

The role of typology in code-switching

Code-switching is another area of language contact where different types of morphemes exhibit different patterns of distribution. Code-switching, as a general term, refers to both mixing between and within sentences. Here the focus is on *intra-sentential* code-switching where the grammatical systems of two or more languages come into contact and interact. Like the borrowing literature, two main approaches to this type of code-switching exist: social motivations accounts (Fishman, 1972, 1965, 1964; Gumperz, 1982; Auer, 1998) and structural theories. Structural accounts are either formally (e.g., Disciullo, Muysken and Singh, 1986; MacSwan, 1999) or typologically orientated. Of interest here are typological accounts which explain the shape of code-switching and the restrictions on mixing grammars, in particular Poplack's (1980) Equivalence Constraint, Myers-Scotton's (2002) Matrix Language Frame (MLF) model of code-switching, and Sebba's (1998) application of typological congruence to code-switched clauses.

The earliest constraints-based theories of code-switching focused on linear ordering, specifically the juxtaposition of two elements, and whether switching is possible between them. In a study of Spanish-English code-switching of Mexican Americans, Pfaff (1979, p. 314) suggests that a switch may occur where the surface structures of the languages map onto

each other. This constraint was later formalized by Poplack and Sankoff as the Equivalence Constraint in further studies of Spanish-English code-switching (Poplack, 1980; Poplack and Meechan, 1995; Sankoff and Poplack, 1981). In this approach, the linear equivalence of elements within sentential phrases such as VPs, DPs, and PPs is considered the determining factor of whether switches occur or not. Thus, switches only occur at points where the surface structures of the languages coincide, for example determiner-adjective-noun order must match between languages in contact for transfer to occur. Myers-Scotton (2002) has a similar constraint which she calls the Morpheme Order Principle and her Blocking Hypothesis also evokes the importance of the typological match between languages, suggesting that insertions from the source language can only occur where the thematic role, morphological structure, and discursive function of the equivalent category are congruent with the recipient language.

Myers-Scotton and Jake's notion of the Matrix Language Frame (MLF) model and the 4-M model has also been influential in constraint-based theories of code-switching (Myers-Scotton, 2002; Myers-Scotton and Jake, 2000, 2017). The MLF model is based on two oppositions: the matrix language versus the embedded language, and content versus system morphemes. The matrix language is the dominant language which sets the grammatical frame for the code-switching, and the embedded language contributes content morphemes within this frame (Myers-Scotton, 1998, p. 291). Myers-Scotton classifies morphemes according to her own 4-M model. She divides them into *content* and *system* morphemes, with system morphemes further divided into *early* and *late* system morphemes. Late system morphemes are of two types: *bridge* and *outsider* morphemes. These are represented schematically in Figure 10.1.

Content morphemes participate in the thematic grid of the utterance. They assign or receive thematic roles, where system morphemes do not (Myers-Scotton, 1993b, pp. 98–99). Prototypical examples are nouns and verbs. On the other hand, *system* morphemes are more functional in nature. This category includes inflectional morphology amongst other morphemes. *Early* system morphemes do not assign or receive thematic roles; however, they pattern with the content, adding extra meaning to the head of a phrase. These morphemes also depend on the head (a content morpheme) of their maximal projection for their syntactic role (Myers-Scotton and Jake, 2000, p. 1063). Examples of early system morphemes in English include the determiner and the plural marker (Myers-Scotton, 2003, p. 77). *Late* system morphemes do not convey conceptual information; rather grammatical information is contained in these morphemes. There are two different types of late system morphemes: *bridge* system morphemes and *outsider* system morphemes. The difference between these two morphemes lies in where they receive their assignment. Bridge system morphemes depend on information from

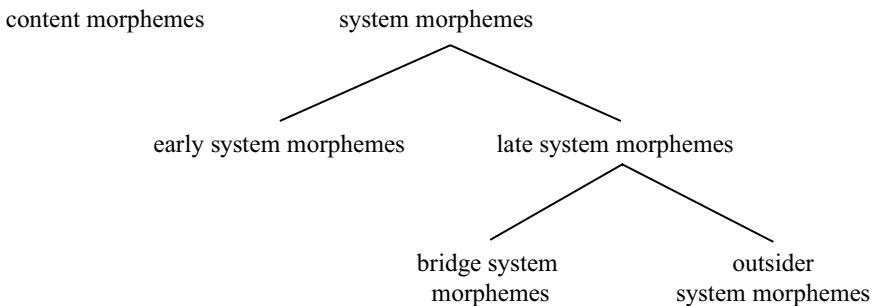


Figure 10.1 Myers-Scotton's 4-M model

within their phrasal node, whereas outsider system morphemes rely on a source *outside* of their immediate phrasal node (Myers-Scotton, 2003, pp. 78–79). Outsider system morphemes include what Booij (1994, 1996) calls contextual inflection, for example subject-verb agreement markers and case morphology (Myers-Scotton and Jake, 2000, pp. 1065–1066).

Within this morphological framework, Myers-Scotton (1993a, p. 83) predicts that, in code-switching, all system morphemes are blocked from participating in code-switching because they rely on the larger syntactic context. This prediction is called the System Morpheme Principle, and is not unlike Gardani’s (2008) claim for contextual inflection and borrowing (see Section 4). These predictions about the behaviour of inflectional morphology in code-switching is largely upheld by data. In her own work on Swahili-English code-switching Myers-Scotton finds that, where Swahili is the matrix language, only English content words are inserted into a grammatical frame which consists of Swahili inflectional affixes. Muysken (2000, pp. 155–156) observes similar patterns of mixing between Dutch and various languages including Malay, Sranan, Chinese, and Turkish which act as matrix languages. On the whole, Dutch only contributes content words to these code-switching combinations. Nonetheless some counter-examples exist, for example the insertion of Turkish inflected verbs in a Romani matrix clause by the Romani community in Thrace (Greece) (Adamou and Granqvist, 2015), and the use of Gurindji case-marking in a Kriol matrix clause, which was the mechanism by which Gurindji case-marking was absorbed into the developing mixed language (Meakins, 2014).

The degree to which languages match typologically also informs Sebba’s (1998) notion of Categorical Congruence. Categorical Congruence is the structural and typological compatibility or ‘match’ of functionally equivalent elements from interacting languages. Although it plays a role in a number of theories of code-switching (and borrowing), Sebba goes further suggesting that it is in fact the basis for an overarching principle of a syntax of code-switching:

An element of language L_1 (morpheme, word or phrase) may be replaced by a congruent element from the other language, L_2 if one exists.

(Sebba, 1998, p. 8)

The congruence between particular categories across languages can have one of four effects on code-switching: blocking, harmonization, neutralization, and compromise, with the first two accounting for a majority of the code-switching data. Firstly, the notion of *blocking* is essentially based on the same idea as the Equivalence Constraint, Morpheme Order Principle, and Blocking Hypothesis where switches are limited by structural features of the interacting languages (Sebba, 1998, p. 13). Constraints arise from structural incompatibility between the code-switching languages. On the other hand, Categorical Congruence may promote code-switching if speakers treat categories in different languages as the same. Sebba (1998, p. 9) calls this effect of code-switching, *harmonization*. Finally note that Sebba does not consider congruence to be an absolute feature of languages in contact, but rather ‘(b)ilinguals “create” congruence categories by finding common ground between the languages concerned’ (1998, pp. 7–8).

The role of typology in mixed language formation

Mixed languages ‘are a type of contact language that arises as the result of the fusion of two languages, normally in situations of bilingualism’ (Meakins, 2018). Although there is some debate on their origins (e.g., Bakker, 2003), they are generally considered to form from unmarked code-switching practices or extensive borrowing (which are probably also

related) (Auer, 1999; McConvell and Meakins, 2005; Meakins, 2011b; Myers-Scotton, 2003; Maschler, 1998). In this respect, the typology of the two source languages and congruence between functionally equivalent elements has been shown to have a large effect on the outcome of the shape of mixed languages.

Myers-Scotton (2003, p. 91) uses her MLF framework to provide an account of how mixed languages form. She considers mixed languages as the fossilization of code-switching and suggests they ‘represent turnovers [of the matrix language] that do not go to completion, but “stop along the way”’ (Myers-Scotton, 2002, p. 249). Mixed languages stop at different places along the scale, which explains why they surface in different forms and with the split in different places. G(rammar)-L(exicon) mixed languages, the most common type, have a morpho-syntactic frame from one language and large amounts of lexicon from another. For example, *Media Lengua* is syntactically Quechua with large amounts of Spanish lexicon. The structure of this mix nonetheless obeys the constraints of the MLF framework. (Note though that Myers-Scotton does not consider these true mixed languages.)

In other cases, mixed languages are the result of a shift in dominance of the interacting languages. The matrix language and embedded language begin the process of swapping roles. The embedded language gains strength and begins to contribute a significant amount of material in the form of system morphemes to the morpho-syntactic frame of the mix. However, this turnover fossilizes half-way, resulting in ‘composite matrix language.’ For Myers-Scotton, this composite matrix language is crucial to a true mixed language (but most of the literature casts the net wider than these structurally mixed types, see (Bakker, 2015; Meakins, 2013, 2018). Myers-Scotton calls this process the Matrix Language Turnover hypothesis. The outcome is a language which contains late system morphemes such as inflectional morphology from the weaker language, i.e., the language which was the embedded language in the code-switching (Myers-Scotton, 2003, p. 92). Myers-Scotton (2003, p. 91) also suggests that the loss of late system morphemes from the more dominant language (the prior matrix language) or the reanalysis of morphemes from the weaker language to function in syntactic roles is also evidence of a fossilized turnover. Under this theory, very few languages which have been identified as mixed languages actually qualify. These languages are Michif, Mednyj Aleut, Light Warlpiri, and Gurindji Kriol, which are a subtype of mixed language referred to as V(erb)-N(oun) mixed languages or structural mixes elsewhere in the mixed language literature.

Morphology from both French and Cree is present in Michif, which is spoken in Canada. Verbal inflections are derived from Cree and the nominal system preserves both French plural morphology and adjectival agreement (Bakker, 1997). Mednyj Aleut, spoken on Mednyj Island in the Bering Strait, consists of many Aleut nominal inflections, including two case distinctions (absolute and relative) and various derivational suffixes (agentive, instrumental, locative, detransitive, inchoative markers, etc.), as well as finite verbal inflectional morphology from Russian, including portmanteau morphemes which express tense, number, person markers (Golokovo, 1994; Thomason, 1997). Light Warlpiri and Gurindji Kriol, spoken in northern Australia, exhibit a split between the nominal and the verbal systems. The nominal structure including case marking (ergative, dative, locative, allative and ablative) is from the heritage language, Warlpiri or Gurindji, with Kriol providing the verb structure including TAM auxiliaries and some suffixes (Meakins, 2011b; O’Shannessy, 2013).

Myers-Scotton’s Matrix Language Turnover hypothesis makes little in the way of predictions about which elements come from which language in Michif, Mednyj Aleut, Light Warlpiri, and Gurindji Kriol. She considers this a largely socio-historical process which relates to power and political imbalances associated with these languages. On the other hand, McConvell’s (2002, 2008) Centre of Gravity Hypothesis uses typology to predict which

language will provide the verb structure and which will provide the noun structure. He relies on the distinction between head and dependent marking languages (cf. Nichols, 1986). McConvell suggests that the typology of the ancestral language may have something to do with the nature of the split, and the way in which this ancestral language shifts to an introduced language via an intermediary period of code-switching and perhaps more structural language mixing. McConvell observes that head-marking and dependent-marking languages have different centres of gravity. The typology of the ancestral language affects the path of the ‘turnover’ to the new language by producing different results in the intermediate language stages (see also Matras (2000, 2003)). McConvell suggests that head-marking languages have a tendency to retain the verbal grammar after the nominal grammar has ‘turned over’ to the introduced language. For example, Michif has retained the Cree verbal grammar whilst adopting French nominal grammar. Conversely dependent-marking languages keep the nominal structure of the ancestral language whilst turning over to the VP structure of the new language (McConvell, 2002, p. 345). McConvell suggests this is the case with Gurindji Kriol. The turnover to Kriol has frozen midway maintaining the nominal grammar of Gurindji.

Although McConvell’s model is heavily based on typology, it focuses on the typology of the ancestral language. McConvell does not discuss the typology of the introduced language, which may be relevant to the resultant structural mix. It is also incorrect to assume that Cree was the ancestral language in the case of Michif, because Michif is the result of mixed marriages between Cree-speaking women and French-speaking men (Bakker, 1997, p. 208). In this respect, neither language can be assumed to be the ancestral or introduced language. McConvell’s analysis works better for Gurindji Kriol, where the original language of the population was clearly Gurindji. Thus, it is more reasonable to suggest that Gurindji was the ancestral language and speakers stopped midway in a shift towards Kriol. However, the typological analysis of Gurindji is problematic given that it is neither strongly head- nor dependent-marking. Gurindji contains both case-marked nouns and cross-referencing pronominal clitics which, though not always marked directly on the verb, are a part of the verb complex.

The role of typology in mixed languages has also been applied to subsystems of languages. For example, Gurindji Kriol and Light Warlpiri show remarkable structural similarity with each other. These structural similarities are not surprising given that the mixed languages are derived from typologically similar languages, Gurindji and Warlpiri (Ngumpin-Yapa, Pama-Nyungan), and share the Kriol component. Nonetheless, one of the more striking differences between the languages is the source of verbs. One-third of the verbs in Gurindji Kriol are derived from Gurindji, whereas few verbs in Light Warlpiri are of Warlpiri origin. Additionally, verbs of Gurindji origin in Gurindji Kriol are derived from coverbs, whereas the Warlpiri verbs in Light Warlpiri come from inflecting verbs. Meakins and O’Shannessy (2012) claim that the relative morphological boundedness of the coverbs (unbound in Gurindji, bound in Warlpiri) and the presence of a typologically equivalent category in Kriol (an uninflected free-form main verb) predict the presence of Gurindji or Warlpiri coverbs in the mixed languages.

3. Critical issues and topics

When languages come into contact, functionally equivalent elements from both languages come into competition. What determines whether one elements ‘wins’ or disappears, or is functionally re-purposed, or exists alongside the equivalent element from the other language to double-mark a function? In all of these language scenarios – borrowing, code-switching, and mixed language formation – the resultant shape of languages in contact is largely a combination of socio-historical and typological pressures. In terms of the typological literature, two

major themes emerge from the literature. Constraints on the ability of words or morphemes to transfer from one language to another are attributed to either (1) the absolute transferability of a morpheme, or (2) the typological congruence between the languages in contact. In the first case, transfer constraints are attributed to the nature of the word or morpheme itself. For example, borrowing hierarchies claim that nouns are the easiest to transfer and inflectional morphology is the hardest, and similar claims have been made in the code-switching literature. These patterns have most often been attributed to ‘boundedness,’ either morphological boundedness, i.e., the degree to which morphemes are free or bound, for example Poplack’s (1980) Free Morpheme Constraint; or syntactic boundedness, i.e., the degree to which morphemes interact with the rest of the syntax of a sentence, for example Gardani’s (2008) observation that contextual inflections are more difficult to borrow than inherent inflections. The typological match of languages in contact is also considered a crucial factor in the transferability of particular morphemes. At issue here is not the independent ability of individual morphemes transfer, but how well these morphemes map onto structures in the recipient or embedded language. For example, Weinreich (1974 [1953], p. 31) considered the borrowability of morphemes to be dependent on typological equivalences between the source and recipient languages. This idea has been reformulated in a number of other works in the borrowing and code-switching including Poplack (1980), Thomason and Kaufman (1988, pp. 72–74), Myers-Scotton (1993a), Sebba (1998), and Field (2002, pp. 41–44).

4. Current contributions and research

More recently, the literature on typology and language contact has begun examining how languages in contact affect the overall typological shape of languages. Another important recent contribution is the use of contact phenomena as a source of evidence for typological theory.

Contact-induced typological change

Language contact can also induce wholesale typological change in languages. On an areal scale, contact-induced typological change may result in Linguistic Areas or Sprachbunds where neighbouring languages share grammatical structures, for example affixal, word or phrasal order, but retain their native morphological and lexical forms (Muysken, 2008; Thomason and Kaufman, 1988). The Balkan Sprachbund is one such example (see chapter in this volume). Ross (1996, p. 95) calls this contact process ‘metatypy,’ whereby the syntax of one language is restructured on the model of the syntax of another language. Metatypy can be thought of as wide-scale PAT(tern) borrowing (Sakel, 2007). Contact-induced typological change can also affect pairs of languages, rather than larger areas of neighbouring language. For example, Kannada (Dravidian) has been re-modelled on the structure of Marathi (Indo-Aryan) (Gumperz and Wilson, 1971), although this claim has been challenged by Kulkarni-Joshi (2016) through extensive research on free speech. This phenomenon has been well described for a number of decades.

More recently, the extreme end of the metatypy scale has been identified. These cases are referred to as ‘converted’ mixed languages (cf., Bakker, 2003). For example, Sri Lanka Malay is a Malay/Indonesian (Austronesian) variety heavily restructured under the influence of Tamil (Dravidian) and more recently Sinhala (Indo-Aryan). Structurally, Sri Lanka Malay developed from an isolating language to an agglutinating language under the influence of Tamil. It has also acquired SOV word order, pre-nominal determiners and adjectives, and, most significantly, case-marking, due to this contact (Ansaldo, 2011; Nordhoff, 2012; Smith, Paauw and

Hussainmiya, 2004). Similarly, Takia, a language spoken on Karkar Island off the north coast of Papua New Guinea (PNG) has an Austronesian lexicon (and is therefore sometimes classified as Austronesian (Dunn et al., 2005)), however the language has undergone extensive restructuring on the model of Waskia (Trans New Guinea) which is also spoken on Karkar Island. In terms of morphology, Takia has developed TAM enclitics as a result of contact with Papuan languages, where Austronesian languages use preverbal auxiliary forms. (Ross, 2001, 2006).

Language contact as a source of evidence for typological theory

Turning now to the impact of contact linguistics on typological research, it is apparent that the behaviour of linguistic structures in situations of language contact can be used as evidence for theoretical perspectives, although evidence against the role of typology in shaping code-switching and language change also exists (Torres Cacoullos and Travis, 2018). In the extreme is a theory of morphology based on the behaviour of morphemes in situations of language contact, i.e., Myers-Scotton and Jakes's (2000, 2017) 4-M model which is used to make predictions about switch sites in insertional code-switching.

Borrowing has also been used to make arguments for categories of morphology. For example, Gardani (2008, 2018) provides evidence for Booij's (1994, 1996) distinction between contextual and inherent morphology within the Strong Lexicalist framework, using the borrowing behaviour of morphemes. Gardani notes that inherent inflections, e.g., TAM, gender, and number affixes, are more likely to be borrowed than contextual inflection, e.g., case affixes and agreement marking. Contextual morphology is determined by the syntactic contexts in which it occurs. Its role is to mark the relationship between a head and a dependent in a syntactic relationship of either government or agreement. Structural case is a good example. Where a language marks case, the verb assigns case to its arguments (subject, object, and indirect object). Another type of case, semantic case, belongs to the category of inherent inflection. This type of inflection is determined by the information a speaker wishes to convey and is exemplified by local case markers such as the locative, allative, and ablative. Gardani (2008) observes that inherent morphology is more readily borrowed than contextual morphology due to its dependence on other parts of the grammar.

Code-switching has also been used as evidence for the typological structure of languages. Meakins (2014) uses code-switching to argue for the status of nominals in terms of argument relations in non-configurational languages. She observes that generativists have argued that nominals in non-configurational languages have the status of adjuncts or secondary predicates rather than arguments, although several arguments against this approach have come from the LFG (Lexical Functional Grammar) literature. Meakins provides new evidence for the claim that nominals are not arguments in non-configurational languages from Gurindji-Kriol code-switching where Gurindji-derived case-marked nominals only occur cross-referenced with a pronoun. She argues that they may be thought of as *added* structures (e.g., adjuncts or secondary predicates) rather than arguments which are *inserted* into the Kriol matrix clause, and suggests that this adjunct-like structure must have been available in the source language, Gurindji.

5. Future directions

As with many disciplines in linguistics, studies of language contact are now using evidence which is 'discoverable and replicable,' in the scientific sense. For this, corpora of contact languages and processes are being developed and made accessible to test theoretical claims

(Adamou, 2016; Meakins, Green and Turpin, 2018). In the case of mixed languages, corpora have been created for Angloromani (Matras, 2010; Matras et al., 2007), Sri Lanka Malay (Ansaldò, Matthews and Lim, 2007; Nordhoff, 2012) and Gurindji Kriol (Meakins, 2011b); and this is also the case for code-switching, for example the Greek-Thrace Romani corpus (Adamou, 2016), French-English corpus (Poplack, 2015) and Spanish-English corpus (Torres Cacoullos and Travis, 2015). Large corpora of individual language contact situations means being able to move beyond making broad claims based on individual linguistic constructions for a limited number of speakers, which is particularly problematic since contact often induces variation and change. For example, a recent study of Gurindji Kriol which utilized data from 78 Gurindji people across three generations using 120 variables (with 292 variants) found that morphological simplification was not inevitable in language contact and suggested that taking a large sample of language elements rather than focusing on individual features allows a broader assessment of the typological processes underlying contact language development (Meakins et al., 2019).

Large surveys now also exist *across* as well as *within* languages. These studies are also challenging the received wisdom of the typological contact literature. For example, cross-linguistic surveys have found instances of inflectional borrowings where they had previously been considered not possible (Gardani, 2008; Gardani, Arkadiev and Amiridze, 2015; Meakins, 2011a; Seifart, 2013). At the most extreme end of the scale, a number of instances of contextual borrowings have been observed. Meakins (2015) notes that Australia has a high number of case borrowings (e.g., Heath, 1978), which is attributed to the non-argument status of nouns in non-configurational languages (Meakins, 2011a).

Case studies such as these speak to the use of quantitative methods in contact linguistics. For example, although case studies of inflectional transfer are often held up as counter-examples to the generalization that inflectional morphology is considered the most resistant to borrowing (e.g., Thomason, 2015), borrowing scales and code-switching rules should not be treated as absolute. They are probabilistic and are based on cross-linguistic tendencies. In this respect, they do not exclude borrowings or switches of any kind, but rather suggest that some transfers are rarer or more common than others. Pioneers in this field were Poplack, Sankoff and their collaborators (Poplack, Pousada and Sankoff, 1982), but the uptake has been slow. Probabilistic approaches require large data sets, which are often difficult to build. These data sets can involve either individual contact situations, for example Meakins et al. (2019), or a diverse range of language pairs for making cross-linguistic comparisons, for example Matras and Sakel (2007) use 27 pairs of languages and AfBo consists of descriptions of 101 cases of affix borrowing (Seifart, 2013). Their use will provide a more informed approach to making generalizations about the role of typology in language contact.

6. Further reading

Gardani, F. (2018). On morphological borrowing. *Language and Linguistics Compass*, 12, pp. 1–17.

This paper shows that the behaviour of morphology in situations of language contact can be used as a source of evidence for the theory of grammar, for example in reflecting fine-grained distinctions between subcomponents of morphology.

Matras, Y. (2007). The borrowability of structural categories. In: Y. Matras and J. Sakel, eds., *Grammatical borrowing in cross-linguistic perspective*, 1st ed. Berlin: De Gruyter Mouton.

This chapter provides an overview of data chapters in the Matras and Sakel book about the likelihood of borrowing different structural categories including morphology. They survey data from 27 pairs of languages in contact.

Meakins, F. (2018). Mixed languages. In: M. Aronoff, ed., *Oxford research encyclopedia of linguistics*, 1st ed. Oxford: Oxford University Press.

This paper gives a comprehensive overview of languages which have been labelled ‘mixed,’ including their sociohistorical background and typological characteristics. Their morphological makeup constitutes a part of their classification as mixed languages.

Myers-Scotton, C. and Jake, J. L. (2017). Revisiting the 4-M model: Codeswitching and morpheme election at the abstract level. *International Journal of Bilingualism*, 21, pp. 340–366.

This paper is the most up-to-date instantiation of the 4-M model and principles and predications that follow from Myers-Scotton and Jake’s classification of morphemes. The paper is essentially a theory of morphology, based on their behaviour in situations of code-switching.

Weinreich, U. (1974 [1953]). *Languages in contact: Findings and problems*. The Hague: Mouton.

This book is a must-read classic of language contact. It provides the basis for most modern predictions and observations about the behaviour of morphology in situations of language contact.

7. Related topics

Creoles and pidgins, mixed languages, borrowing

Abbreviations

| | |
|-----|----------------------------|
| DP | determiner phrase |
| G | grammar |
| L | lexicon |
| LFG | Lexical Functional Grammar |
| MAT | matter |
| MLF | Matrix Language Frame |
| N | noun |
| O | object |
| PAT | pattern |
| PNG | Papua New Guinea |
| PP | prepositional phrase |
| S | subject |
| TAM | tense, aspect, mood |
| V | verb |
| VP | verb phrase |

Note

1 I use the term ‘transfer’ to include phenomena which other people have termed ‘borrowing,’ ‘replication,’ ‘copying,’ and ‘diffusion.’ We also include ‘insertional code-switching’ (cf. Muysken, 2000) which occurs intra-sententially and differs from borrowing only by degree.

References

- Adamou, E. (2016). *A corpus-driven approach to language contact: Endangered languages in a comparative perspective*. Boston and Berlin: De Gruyter Mouton.
- Adamou, E. and Granqvist, K. (2015). Unevenly mixed Romani languages. *International Journal of Bilingualism*, 19, pp. 525–547.

- Aikhenvald, A.Y. and Dixon, R.M.W. (2007). *Grammars in contact: A cross-linguistic typology*. Oxford: Oxford University Press.
- Ansaldo, U. (2011). Metatypy in Sri Lanka Malay. In: G. Sharma and R. Singh, eds., *Annual review of South Asian languages and linguistics*, 1st ed. Berlin: Walter de Gruyter.
- Ansaldo, U., Matthews, S. and Lim, L., eds. (2007). *Deconstructing creole*. Amsterdam: John Benjamins.
- Auer, P. (1998). *Code-switching in conversation: Language, interaction and identity*. London: Routledge.
- Auer, P. (1999). From codeswitching via language mixing to fused lects: Toward a dynamic typology of bilingual speech. *Journal of Bilingualism*, 3, pp. 309–332.
- Backus, A. (2003). Can a mixed language be conventionalized alternational codeswitching? In: Y. Matras and P. Bakker, eds., *The mixed language debate: Theoretical and empirical advances*, 1st ed. Berlin: De Gruyter Mouton.
- Bakker, P. (1997). *A language of our own: The genesis of Michif, the mixed Cree-French language of the Canadian Métis*. New York: Oxford University Press.
- Bakker, P. (2003). Mixed languages as autonomous systems. In: Y. Matras and P. Bakker, eds., *The mixed language debate: Theoretical and empirical advances*, 1st ed. Berlin: De Gruyter Mouton.
- Bakker, P. (2015). Typology of mixed languages. In: A.Y. Aikhenvald and R.M.W. Dixon, eds., *The Cambridge handbook of linguistic typology*, 1st ed. Cambridge: Cambridge University Press.
- Booij, G. (1994). Against split morphology. *Yearbook of Morphology 1993*, pp. 27–50.
- Booij, G. (1996). Inherent versus contextual inflection and the split morphology hypothesis. *Yearbook of Morphology 1995*, pp. 1–16.
- Disciullo, A.-M., Muysken, P. and Singh, R. (1986). Government and code-mixing. *Journal of Linguistics*, 22, pp. 1–24.
- Dunn, M., Terrill, A., Reesink, G., Foley, R.A. and Levinson, S.C. (2005). Structural phylogenetics and the reconstruction of ancient language history. *Science*, 309, pp. 2072–2075.
- Field, F. (2002). *Linguistic borrowing in bilingual contexts*. Amsterdam: John Benjamins.
- Fishman, J. (1964). Language maintenance and language shift as a field of inquiry. *Linguistics*, 9, pp. 32–70.
- Fishman, J. (1965). Who speaks what language to whom and when. *La Linguistique*, 2, pp. 67–68.
- Fishman, J. (1972). Domains and the relationship between micro- and macro- sociolinguistics. In: J. Gumperz and D. Hymes, eds., *Directions in sociolinguistics*, 1st ed. New York: Holt, Rinehart and Winston.
- Gardani, F. (2008). *Borrowing of inflectional morphemes in language contact*. Frankfurt: Peter Lang.
- Gardani, F. (2018). On morphological borrowing. *Language and Linguistics Compass*, 12, pp. 1–17.
- Gardani, F., Arkadiev, P. and Amiridze, N., eds. (2015). *Borrowed morphology*. Berlin: De Gruyter Mouton.
- Golokovo, E.V. (1994). Mednyj Aleut or Copper Island Aleut: An Aleut-Russian mixed language. In: P. Bakker and M. Mous, eds., *Mixed languages: 15 case studies in language intertwining*, 1st ed. Amsterdam: Uitgave IFOTT.
- Gumperz, J. (1982). *Discourse strategies*. Cambridge: Cambridge University Press.
- Gumperz, J. and Wilson, R. (1971). Convergence and creolization. In: D. Hymes, ed., *Pidginization and creolization of languages*, 1st ed. Cambridge: Cambridge University Press.
- Haugen, E. (1950). The analysis of linguistic borrowing. *Language*, 26, pp. 210–231.
- Heath, J. (1978). *Linguistic diffusion in Arnhem land*. Canberra: Australian Institute of Aboriginal Studies.
- Kulkarni-Joshi, S. (2016). Forty years of language contact and change in Kupwar: A critical assessment of the intertranslatability model. *Journal of South Asian Languages and Linguistics*, 32, pp. 147–174.
- MacSwan, J. (1999). *A minimalist approach to intrasentential codeswitching*. New York: Garland.
- Maschler, Y. (1998). On the transition from code-switching to a mixed code. In: P. Auer, ed., *Code-switching in conversation: Language, interaction and identity*, 1st ed. London: Routledge.
- Matras, Y. (1998). Utterance modifiers and universals of grammatical borrowing. *Linguistics*, 36, pp. 281–331.

- Matras, Y. (2000). Mixed languages: A functional-communicative approach. *Bilingualism: Language and Cognition*, 3, pp. 79–99.
- Matras, Y. (2003). Mixed languages: Re-examining the structural prototype. In: Y. Matras and P. Bakker, eds., *The mixed language debate: Theoretical and empirical advances*, 1st ed. Berlin: De Gruyter Mouton.
- Matras, Y. (2007). The borrowability of structural categories. In: Y. Matras and J. Sakel, eds., *Grammatical borrowing in cross-linguistic perspective*, 1st ed. Berlin: De Gruyter Mouton.
- Matras, Y. (2009). *Language contact*. Cambridge: Cambridge University Press.
- Matras, Y. (2010). *Romani in Britain. The afterlife of a language*. Edinburgh: Edinburgh University Press.
- Matras, Y. (2011). Universals of structural borrowing. In: P. Siemund, ed., *Linguistic universals and language variation*, 1st ed. Berlin: De Gruyter Mouton.
- Matras, Y., Gardner, H., Jones, C. and Schulmann, V. (2007). Angloromani: A different kind of language? *Anthropological Linguistics*, 49, pp. 142–184.
- Matras, Y. and Sakel, J., eds. (2007). *Grammatical borrowing in cross-linguistic perspective*. Berlin: De Gruyter Mouton.
- McConvell, P. (2002). Mix-im-up speech and emergent mixed languages in indigenous Australia. *Texas linguistic forum (Proceedings from the 9th annual symposium about language and society)*, 44, pp. 328–349.
- McConvell, P. (2008). Mixed languages as outcomes of code-switching: Recent examples from Australia and their implications. *Journal of Language Contact*, 2, pp. 187–212.
- McConvell, P. and Meakins, F. (2005). Gurindji Kriol: A mixed language emerges from code-switching. *Australian Journal of Linguistics*, 25, pp. 9–30.
- Meakins, F. (2011a). Borrowing contextual inflection: Evidence from northern Australia. *Morphology*, 21, pp. 57–87.
- Meakins, F. (2011b). *Case marking in contact: The development and function of case morphology in Gurindji Kriol*. Amsterdam: John Benjamins.
- Meakins, F. (2013). Mixed languages. In: Y. Matras and P. Bakker, eds., *Contact languages: A comprehensive guide*, 1st ed. Berlin: De Gruyter Mouton.
- Meakins, F. (2014). Nominals as adjuncts or arguments: Further evidence from language mixing. In: R. Pensalfini, M. Turpin, and D. Guilleman, eds., *Language description informed by theory*, 1st ed. Amsterdam: John Benjamins.
- Meakins, F. (2015). From absolutely optional to only nominally ergative: The life cycle of the Gurindji Kriol ergative suffix. In: F. Gardani, P. Arkadiev, and N. Amiridze, eds., *Borrowed morphology*, 1st ed. Berlin: De Gruyter Mouton.
- Meakins, F. (2018). Mixed languages. In: M. Aronoff, ed., *Oxford research encyclopedia of linguistics*, 1st ed. Oxford: Oxford University Press.
- Meakins, F., Green, J. and Turpin, M. (2018). *Understanding linguistic fieldwork*. London: Routledge.
- Meakins, F., Hua, X., Algy, C. and Bromham, L. (2019). The birth of a new language does not favour simplification. *Language*, 95, pp. 294–332.
- Meakins, F. and O’Shannessy, C. (2012). Typological constraints on verb integration in two Australian mixed languages. *Journal of Language Contact*, 5, pp. 216–246.
- Moravcsik, E. (1978). Universals of language contact. In: J. Greenberg, ed., *Universals of human language: Vol. 1 method and theory*, 1st ed. Stanford, CA: Stanford University Press.
- Muysken, P. (1981). Halfway between Quechua and Spanish: The case for relexification. In: A. Highfield and A. Valdman, eds., *Historicity and variation in creole studies*, 1st ed. Ann Arbor: Karoma.
- Muysken, P. (2000). *Bilingual speech: A typology of code-mixing*. Cambridge: Cambridge University Press.
- Muysken, P., ed. (2008). *From linguistic areas to areal linguistics*. Amsterdam: John Benjamins.
- Myers-Scotton, C. (1993a). *Duelling languages: Grammatical structure in codeswitching*. Oxford: Clarendon Press.

- Myers-Scotton, C. (1993b). *Social motivations for codeswitching: Evidence from Africa*. Oxford: Clarendon Press.
- Myers-Scotton, C. (1998). A way to dusty death: The Matrix Language turnover hypothesis. In: L. Grenoble and L.J. Whaley, eds., *Endangered languages: Language loss and community response*, 1st ed. Cambridge: Cambridge University Press.
- Myers-Scotton, C. (2002). *Contact linguistics: Bilingual encounters and grammatical outcomes*. Oxford and New York: Oxford University Press.
- Myers-Scotton, C. (2003). What lies beneath: Split (mixed) languages as contact phenomena. In: Y. Matras and P. Bakker, eds., *The mixed language debate: Theoretical and empirical advances*. Berlin: De Gruyter Mouton.
- Myers-Scotton, C. and Jake, J.L. (2000). Four types of morpheme: Evidence from aphasia, code switching, and second-language acquisition. *Linguistics*, 38, pp. 1053–1100.
- Myers-Scotton, C. and Jake, J.L. (2017). Revisiting the 4-M model: Codeswitching and morpheme election at the abstract level. *International Journal of Bilingualism*, 21, pp. 340–366.
- Nichols, J. (1986). Head-marking and dependent-marking grammar. *Language*, 62(1), pp. 56–119.
- Nordhoff, S., ed. (2012). *The genesis of Sri Lanka Malay: A case of extreme language contact*. Leiden: Brill.
- O'Shannessy, C. (2013). The role of multiple sources in the formation of an innovative auxiliary category in light Warlpiri, a new Australian mixed language. *Language*, 89, pp. 328–353.
- Pfaff, C.W. (1979). Constraints on language mixing: Intrasentential code-switching and borrowing in Spanish/English. *Language*, 55, pp. 291–318.
- Poplack, S. (1980). Sometimes I'll start a sentence in Spanish *y termino en español*: Towards a typology of code-switching. *Linguistics*, 18, pp. 581–618.
- Poplack, S. (2015). Code-switching (linguistic). In: J. Wright, ed., *International encyclopedia of the social and behavioral sciences*, 1st ed. United States: Elsevier.
- Poplack, S. and Meechan, M. (1995). Patterns of language mixture: Nominal structure in Wolof-French and Fongbe-French bilingual discourse. In: L. Milroy and P. Muysken, eds., *One speaker, two languages: Cross-disciplinary perspectives on code-switching*. Cambridge: Cambridge University Press.
- Poplack, S., Pousada, A. and Sankoff, D. (1982). Competing influences on gender assignment: Variable process, stable outcome. *Lingua*, 57, pp. 1–28.
- Ross, M. (1996). Contact-induced change and the comparative method: Cases from Papua New Guinea. In: M. Durie and M. Ross, eds., *The comparative method reviewed: Regularity and irregularity in language change*, 1st ed. New York: Oxford University Press.
- Ross, M. (2001). Contact-induced change in Oceanic languages in north-west Melanesia. In: A.Y. Aikenvald and R.M.W. Dixon, eds., *Areal diffusion and genetic inheritance: Problems in comparative linguistics*, 1st ed. Oxford: Oxford University Press.
- Ross, M. (2006). Metatypy. In: K. Brown, ed., *Encyclopedia of language and linguistics*, 1st ed. Oxford: Elsevier.
- Sakel, J. (2007). Types of loans: Matter and pattern. In: Y. Matras and J. Sakel, eds., *Grammatical borrowing in cross-linguistic perspective*, 1st ed. Berlin: Mouton.
- Sankoff, D. and Poplack, S. (1981). A formal grammar for code switching. *Papers in Linguistics*, 14, pp. 3–46.
- Sebba, M. (1998). A congruence approach to the syntax of code-switching. *International Journal of Bilingualism*, 2, pp. 1–20.
- Seifart, F. (2013). AfBo: A world-wide survey of affix borrowing. *Max Planck institute for evolutionary anthropology*. [Online] Available at: <http://afbo.info/> [Accessed 21 April 2020].
- Singh, R. (1982). On some 'redundant compounds' in modern Hindi. *Lingua*, 56, pp. 345–351.
- Smith, I., Paauw, S. and Hussainmiya, B.A. (2004). Sri Lankan Malay: The state of the art. In: R. Singh, ed., *Yearbook of South Asian languages 2004*, 1st ed. Berlin: De Gruyter Mouton.
- Thomason, S.G. (1997). Mednyj Aleut. In: S.G. Thomason, ed., *Contact languages: A wider perspective*, 1st ed. Amsterdam: John Benjamins.

- Thomason, S.G. (2015). When is the diffusion of inflectional morphology not dispreferred? In: F. Gardani, P. Arkadiev and N. Amiridze, eds., *Borrowed morphology*, 1st ed. Berlin: De Gruyter Mouton.
- Thomason, S.G. and Kaufman, T. (1988). *Language contact, creolization, and genetic linguistics*. Berkeley: University of California Press.
- Torres Cacoullos, R. and Travis, C. (2015). Two languages, one effect: Structural priming in spontaneous code-switching. *International Journal of Bilingualism*, 19, pp. 733–753.
- Torres Cacoullos, R. and Travis, C. (2018). *Bilingualism in the community: Code-switching and grammars in contact*. Cambridge: Cambridge University Press.
- Weinreich, U. (1974 [1953]). *Languages in contact: Findings and problems*. The Hague: Mouton.
- Whitney, W.D. (1881). On mixture in language. *Transactions of the American Philological Association*, 12, pp. 5–26.
- Winford, D. (2003). *An introduction to contact linguistics*. Malden, MA: Blackwell.

Cross-language contact in the developing grammars of bilingual children

Jennifer Austin

1. Introduction

This chapter provides an overview of language contact in the speech of young bilingual children, focusing particularly on the interaction between their developing grammars. In the following sections, I will contrast past analyses of language mixing in the grammars of young children with the current understanding of why these mixes occur and what they indicate about the children's grammatical representations.

Young bilingual children frequently mix elements from their two languages in combinations that would be unacceptable for bilingual adults. Examples of these non-adult mixes can be seen in (1)–(3). In example (1), a child has mixed an Italian NP and determiner with a German adjective, a type of mixing within a syntactic phrase that is not usually acceptable to adult bilinguals.

1. Lexical mixing (mixed words in an adjective phrase):
ein pannolino klein (Carlotta, It-Ger, 2;9, years old, Arnaus Gil et al., 2012)
 a diaper (Italian) small (German)

In examples (2) a, b we see that the children are blending words from two languages such as 'sockatines' (English 'socks' plus Spanish *calcetines* 'socks') or combining a bound morpheme from one language with an unbound morpheme from the other language, as in *tácale* (English 'talk' plus Spanish *le* 'to him or her').

2. Morphological mixing (mixing of bound and unbound morphemes from different languages, or blending words from two languages):
 - a. *sockatines* (English-Spanish bilingual child, p.c.)
 sock + *calcetines* (Spanish)
 - b. *Tú tácale* (English-Spanish bilingual child, Lindholm and Padilla, 1978)
 you (Spanish) talk (English) to him/her (Spanish)

3. Lexical transfer:
 - a. Transfer from Spanish to English:
That isn't appetizing to me (English-Spanish bilingual child, p.c.)
(calque from Spanish *No me apetece* 'I don't feel like it')
 - b. Transfer from English to Spanish:
¿Puedo salir con pies de oso? (English-Spanish bilingual child, p.c.)
'Can I go outside with bear (bare) feet?'

Bilingual children also show evidence of cross-linguistic influence (CLI) in many other areas of language: pragmatics (Sorace and Serratrice, 2009; Syrett et al., 2017), semantics (Liceras, Fuertes and de la Fuente, 2012), syntax (Nicoladis and Gavrilu, 2015), and phonology (Lleó and Kehoe, 2002). The first researchers studying childhood bilingualism assumed that utterances such as these were evidence that the children had a single linguistic fused system that eventually separates into two. This position is known as the *Unitary System Hypothesis*. More recently, researchers have argued the opposite, namely that bilingual children have separate grammatical systems by the time they begin producing multiword utterances. The following section will review the shift from one hypothesis to the other in the research on early grammatical development in bilingual children.

2. Historical overview

One of the earliest proponents of the Unitary System Hypothesis was Leopold (1970, originally published in 1949), who conducted a longitudinal study of his daughter Hildegard's acquisition of German and English. Leopold claimed that at the earliest stage, Hildegard did not distinguish between her languages and used 'one medium of communication.' He based this claim on the many mixed utterances which Hildegard produced early on, such as the examples in (4), which he argued supported the idea that initially Hildegard was unable to separate her German and English grammars.

4. a. *All babies Bett*
'All babies sleep.' [Hildegard 1;11]
- b. *Three meows lost Handschuhe*
'Three meows lost gloves.' [Hildegard 2;0]
- c. *I can't give you any Kuß because I have a Schmutznase*
'I can't give you any kisses because I have a cold(?)' [Hildegard 3;03]
- d. *Look how many Platz there is; is this viel?*
'Look how many spaces there is; is this a lot?' [Hildegard 3;09]

Volterra and Taeschner (1978) also supported the Unitary System Hypothesis, based on findings from their longitudinal study of two Italian/German bilingual children between the ages of 1;05–3;06, and 1;02–2;06. They proposed that there are three stages in bilingual development, beginning with a stage in which the bilingual child has a single grammar and lexicon with words from both languages. At the second stage, the authors proposed that the child has two lexical systems, but only one grammar. At this point, the authors reported, the children mixed their languages within utterances, as seen in (5):

5. a. *Spetta lass zu Lisa komm ida, bene? Bene.*
'Wait, let closed, Lisa comes back OK?' [Lisa 2;05]

- b. *Mami vuole Strickzeug, vuole Arbeit, si?*
'Mommy want knitting, wants work yes?' [Lisa 2;07]
- c. *Giulia gemacht a caseta per a böse Wolf*
'Giulia made a little house for the bad wolf' [Giulia 2;02]
- d. *Mami ich will prendere ja?*
'Mommy, I will take it, yes?' [Giulia 2;05]

The third stage is reached when the child 'speaks two languages differentiated both in lexicon and in syntax' (Volterra and Taeschner, 1978, p. 312).

Redlinger and Park (1980) also argued in favour of the Unitary System Hypothesis, based on data from four bilingual children (one German/Spanish bilingual, one English/German bilingual, and two French/German bilinguals). Redlinger and Park found that as the children grew older, the rate of language mixing declined from 20–30% to 2–5%. The authors found that the majority of children's single word mixes were nouns (around 40%).

Only one of the children, Danny, combined verbal inflection from one language with stems from the other language. Some examples of Danny's mixing within the verb phrase are shown in (6). His exact age at the time of each utterance was not provided in the article but was between 1;11 and 2;06 years.

- 6. a. *pfeift ing*
'Whistling.'
- b. *Die Mädchen 's going night-night*
'The girl's going night-night.'
- c. *Der pusht der kleine Josef.*
'He pushes the little Joseph.' [Danny, 1;11–2;06]

Redlinger and Park concluded that their results 'suggest that the participants were involved in a gradual process of language differentiation and were in agreement with those of previous investigators supporting the one-system approach to bilingual acquisition' (Redlinger and Park, 1980, p. 344).

Vihman (1985) also argued that data from her Estonian/English bilingual son Raivo supported the one-system hypothesis, claiming that he initially had a single lexicon made up of Estonian and English words, and that he mixed English function words in Estonian sentences. She found that mixed utterances accounted for 10–57% Raivo's early speech production between the ages of 1;08 and 2;00 years, a much higher percentage of mixed utterances than would typically be found in the speech of adult bilinguals. Furthermore, Vihman found that Raivo's mixed utterances were mostly comprised of mixes between closed-class words and content words, unlike adult code-switches, which consist primarily of nouns from one language switched into utterances from the other. Vihman proposed that children's mixing rates decline and become more adult-like as the children grow older and their grammars gradually separate.

This hypothesis was challenged by researchers who argued that the grammars of bilingual children are separate by the time that they begin producing multiword utterances (Lindholm and Padilla, 1978; Meisel, 1994; De Houwer, 1990; Paradis and Genesee, 1997; and others). They defended this position, known as the Separate Development Hypothesis, for several reasons: first, they stated that mixed utterances should not be used as evidence for or against grammatical fusion, given that 'mixing' implies there are elements from two discrete systems to be mixed (Meisel, 1989). Also, the frequency of mixed utterances must be taken into

account. As De Houwer (1990) noted, Leopold used his daughter's mixed utterances as proof of an initial hybrid grammatical system, but he could just as easily have pointed to the much higher percentage of non-mixed utterances to argue against this conclusion. In the case of Volterra and Taeschner's 1978 study, Meisel (1989) has argued that their mixed data could also be considered instances of grammatical transfer from the child's dominant language to the weaker one.

An additional argument in favour of the Separate Development Hypothesis comes from the finding that language development in bilingual children proceeds along the same paths as in monolinguals, suggesting that bilingual grammars are not qualitatively different from monolingual grammars at the earliest stages (De Houwer, 1990; Ezeizabarrena, 1996; Paradis and Genesee, 1997; among others).

The Separate Development Hypothesis is also supported by evidence of language differentiation from the beginning of bilingual development. According to De Houwer,

in all aspects of language use investigated that provided unambiguous opportunities for discovering either the presence or absence of inter-linguistic interaction, we were able to show that Kate's developing morphosyntactic knowledge of Dutch could not function as a basis for her speech production in English or vice versa.

(De Houwer, 1990, p. 338)

De Houwer points out that among Kate's utterances with syntactic mixes, many involved repairs (self-corrections), in which the child began using a word or morpheme in the language other than the one she was using, and corrected herself. Bilingual children are also sensitive to the context in which languages are spoken (Lanza, 2004; Genesee, Nicoladis and Paradis, 1995), and that they have metalinguistic awareness of speaking two codes from an early age.

Finally, bilingual children have little difficulty in recognizing and following correct syntactic rules for each language and producing multiword utterances that respect the word order of each grammar. For example, Barreña and Idiazabal (1997) found that their Basque/Spanish bilingual subject Mikel used an SVO word order in Spanish in 87% of his utterances between ages 1;09 and 3;00, which corresponds to monolingual norms and to the unmarked word order of Spanish.

In Basque, which has an unmarked SOV word order (although its word order is fairly free), Mikel produced the object preverbally in 46% of his utterances, and postverbally in 54% of possible contexts. This was very similar to the word order patterns in the speech of a monolingual Basque-speaking child in the same study, who produced objects preverbally 51% and postverbally 49% of the time.

3. Critical issues and topics: explanatory factors in cross-linguistic influence

Despite the evidence that bilingual children's languages are separate from the beginning of speech production, it is also clear that their grammars mutually influence each other. Current research in the field has focused on why and when cross-linguistic influence occurs in developing grammars, and to what extent it persists into adulthood. This change in focus from investigating whether a child's languages are initially separate to examining how the languages of a bilingual child continue to influence each other throughout his/her lifetime has been informed by research showing that adult bilinguals access their two languages non-selectively rather than separately during lexical, phonological, and syntactic tasks (Costa et al., 2003; Loebell

and Bock, 2003; Meijer and Fox Tree, 2003; Kroll, Bobb and Wodniecka, 2006; Desmet and Declercq, 2006; *inter alia*). Despite abundant evidence of cross-linguistic influence in adult bilingual grammars, in many cases it is sufficiently subtle that such influence is only detectable through experimental measures such as reaction time. In contrast, even without experimental methods it is obvious that adult bilinguals generate many fewer language mixes than child bilinguals of the kind seen in examples (1) and (2), which were the subject of study in the earliest research on childhood bilingualism. Current research in early bilingualism has identified several factors which may favour cross-linguistic influence in children's speech, as well as reasons why it can take forms which would not be grammatical for adults. These studies will be reviewed in the following sections of this paper.

Paradis and Genesee (1997, p. 3) defined cross-linguistic influence (CLI) in children as long-term influence from the grammar of one of a bilingual's languages to the other, resulting in developmental patterns that are qualitatively or quantitatively different from those of a child who is monolingual. CLI can take the form of delay, acceleration, or transfer, which is the use of a grammatical property of one language in the other language (1997, p. 3). These types of influence do not include *code-mixing* or *code-switching*, which is when words or morphemes from both of a bilingual's languages are used in the same utterance, as seen in examples (1) and (2).

An example of delay which has been argued to be caused by cross-linguistic influence is the acquisition of Spanish clitics by Basque/Spanish children, who drop more clitics in Spanish than monolingual children do (Larrañaga and Guijarro-Fuentes, 2012). Grammatical acceleration due to cross-linguistic influence has been found in the production of German articles by bilingual children acquiring German and Italian, who omit fewer determiners in German than monolingual children acquiring German (Kupisch, 2007). Similar accelerative effects have been found in the productions of English wh-questions earlier in Spanish/English bilingual children than in monolingual English-speaking children (Hsin, 2012) and the early ability of Dutch-Italian bilingual children to interpret Dutch indefinite objects under the scope of negation, which emerges sooner than in Dutch monolinguals (Meroni, Smeets and Unsworth, 2017).

It is important to note that not all instances in which bilingual acquisition differs from development in monolingual children need be caused by cross-linguistic influence. In the words of Gathercole (2007, p. 225): 'while interaction (between a child's languages) may be a sufficient condition for acceleration or delay, it is not a necessary condition for either.' In many cases, researchers have attributed slower development in bilingual children to their having received less input in each language than monolingual children. This input-related delay has been reported in research on bilingual children acquiring Spanish and English between seven and ten years old, who were significantly less accurate than age-matched Spanish monolinguals on grammaticality judgment tasks assessing knowledge of that-trace effects (the ability of a subject to be extracted from a clause headed by that), grammatical gender and mass/count nouns (Gathercole, 2002, 2007). Gathercole (2007) also reported that bilingual Welsh/English bilingual children aged five to nine years made more errors in producing grammatical gender and recognizing subjects and objects in sentences if they came from homes where only English or English/Welsh were spoken, compared to children from homes where only Welsh was spoken. In a study of bilingual children acquiring French and English between the ages of 2;11 and 5;01 years, Nicoladis and Marchak (2011) found that they produced fewer target-like feminine noun-adjective agreement and determiner-noun agreement than age-matched monolingual children, a difference that the authors attributed to the bilinguals' having received fewer input types and tokens in each language. Similar findings have been reported for the

acquisition of ergative case and agreement in Basque/Spanish bilingual children (Austin, 2007, 2009), the production of verbal and nominal morphology in Turkish/Dutch bilinguals (Blom, 2010), and the development of grammar and vocabulary in Spanish/English bilingual children (Hoff et al., 2012). All these authors concluded that the delay that bilingual children experience in reaching some grammatical milestones is caused by bilinguals receiving less exposure to input than monolinguals do rather than by influence from one language on another, on the assumption that the bilinguals hear fewer hours of input each day than monolingual children, since the input is divided between two languages. These input-related effects are particularly pronounced in the acquisition of morphological forms that have low token frequency in the input (Nicoladis, Palmer and Marentette, 2007; Paradis et al., 2011; Nicoladis, Song and Marantette, 2012).

Lanza (1992) also pointed to the role that parental discourse strategies play in providing feedback to bilingual children regarding the appropriateness of language mixes. In her 1992 longitudinal case study of Siri, a Norwegian-English bilingual child, Lanza examined how often Siri mixed her languages in conversations with her parents, as well as her parents' responses. Lanza categorized these responses as guiding the child towards a more monolingual context after the child code-switches (such as minimal grasp, where the adult requests clarification from the child) or a more bilingual context, in which the parent responds to a child's code-switch with a code-switch of his/her own. Interestingly, Lanza found that although Siri was a Norwegian-dominant bilingual, she used more English words with her Norwegian-speaking father than Norwegian-speaking mother, which Lanza concludes was the result of Siri's father encouraging a more bilingual strategy than her mother did in English.

The next section will provide an overview of recent research investigating why and when cross-linguistic influence appears in bilingual children's grammars.

4. Current approaches and research

In the past few decades, researchers have identified several factors which may favour cross-linguistic influence in young bilinguals' grammars, including grammatical maturation, language dominance, the presence of partial grammatical overlap, the involvement of interface areas of language, the use of cross-linguistic structures as a relief strategy, cross-linguistic priming and language processing difficulties. The following section will discuss grammatical maturation and bilingual bootstrapping, two hypotheses which attempt to explain why children produce non-adult language mixes early in development.

Grammatical maturation and cross-linguistic influence

Köppe and Meisel (1995) and Toribio and Brown (1995) proposed that children's early non-adult mixes such as (7) were attempts at code-switching. These mixes are non-adult like because the switch from one language to another occurs within syntactic phrases.

7. *moi je va à la küche*
me I go to the (French) kitchen (German)
(Ivar, 2:05, Köppe and Meisel, 1995, p. 292, citing Veh, 1990, p. 79)

According to these researchers, children produce this type of language mix because they lack an awareness of the pragmatic constraints on code-switching, as well as the syntactic categories needed for adult-like code-switching. Following Radford's small clause hypothesis,

these authors argued that children initially lack functional projections (or grammatical categories), and therefore that their code-switching is not constrained by functional heads the way that it is in the speech of adult bilinguals. For this reason, Köppe and Meisel claimed that once there was evidence of the emergence of the functional category INFL in their French/German participants' speech, mixing of function words declined sharply, as Vihman's proposal would predict. They claimed that this happened around age 2;06 for one of their participants Ivar, and at 2;00 for the other, Annika. This hypothesis has the advantage of offering a theoretically plausible explanation for why children produce non-adult mixes as well as why they cease to use them as they mature that is based on more general patterns seen in cognitive and linguistic development. However, contrary to the predictions of Köppe and Meisel's hypothesis, children produce these mixes even after verbal inflection and case-marking appear in their grammars, clear evidence of INFL (Austin, 2001). Nevertheless, this line of analysis opens up many interesting questions, such as the exact nature of the mechanisms that allow adult bilinguals to avoid these types of mixes, and how they develop in children.

Cross-linguistic structures as a relief strategy in child grammars

Gawlitzeck-Maiwald and Tracy (1996) and Bernardini and Schlyter (2004) proposed alternative analyses of children's non-adult mixes in the Bilingual Bootstrapping Hypothesis and the Ivy Hypothesis. In a longitudinal study of a bilingual child learning German and English, Gawlitzeck-Maiwald and Tracy (1996) proposed that their participant had an Inflection Phrase (IP) and a Verb Phrase (VP) in German, but only a VP in English, and that she was transferring a German IP to an English VP in utterances such as (8):

8. *Ich hab gemade you much better.*
 'I have made you much better.' [Gawlitzeck-Maiwald and Tracy, 1996, p. 914]

The authors found mixed-IP utterances such as the one in (8) until the child was 2;09, at which point, they argued, she acquired an IP in English as well. Gawlitzeck-Maiwald and Tracy referred to this use of the L1 functional categories in the L2 as *bilingual bootstrapping*, or a relief strategy for filling in grammatical gaps in one of the bilingual child's languages.

Bernardini and Schlyter (2004) made a similar proposal, the *Ivy Hypothesis*, in a study of code-mixing in five bilingual children acquiring Swedish and French or Italian for whom Swedish was the children's dominant language. They argued that unbalanced child bilinguals use their dominant language as a kind of syntactic anchor for the weaker one in utterances such as (9):

9. a. *den kan inte mangia la coda*
 'it cannot (Swedish) eat the tail (Italian)' [Lukas, 2;07]
 b. *där bouch-en*
 there (Swedish) mouth(French)-the (Swedish) [Léo, 1;10]

These authors also claimed that grammatical transfer can occur from the stronger to the weaker language in all areas of syntax.

In Austin (2015) I proposed that Basque/Spanish children used transfer from Spanish to Basque in the form of preverbal complementizers in Basque as a relief strategy. In Basque, a head-final language, clause-final complementizers are obligatory, as shown in

(10). Out of 20 bilingual children age 2;00–3;06, preverbal complementizers were used by four children between the ages of 2;08 and 3;02, as shown in the following examples in (10a) and (11):

10. a. *Ze dako (dakot) moko*
 COMP have.ABS3sg.ERG3sg mucus
 ‘That I have a runny nose.’ [IU, 2’10]
- b. Target: *Dakod-alako moko*
 have.ABS3sg.ERG1sg.COMP mucus
 ‘Because I have a runny nose.’
11. a. **ze etorri da G. ta Jennifer*
 that.COMP come AUX.ABS3sg G and Jennifer
 ‘that G. and Jennifer have come’ [RB 2;08]
- b. Target: *etorri direla G. ta Jennifer*
 see AUX.ABS3pl-COMP G. and Jennifer

In contrast, pre-verbal complementizers were never used by monolingual children or adults. Five other bilingual children in this age range didn’t use them at all, and their production was not correlated with the children’s MLU in Basque. I interpreted the use of these preverbal complementizers to be a type of relief strategy temporarily used by some bilingual children when they need to use a construction that they have not yet acquired. They did not seem to reflect the wholesale transfer of the head-initial properties of Spanish to Basque, given that for other grammatical categories the bilingual children’s grammars were consistently head-final.

Language dominance and cross-linguistic influence

Petersen (1988) and Nicoladis and Genesee (1997) linked early language mixing to dominance of one of the child’s languages. Petersen found that Thea, her Danish/English bilingual participant’s mixed utterances only contained functional morphemes from English, her dominant language, with Danish content morphemes, and never mixed Danish grammatical morphemes with English lexical items. An example of an utterance with mixing in the VP and in the object NP is shown in (12).

12. *I’m laving pandekages*
 I’m making pancakes [Thea, 3;2 Petersen, 1988, p. 480]

Similarly, Lanza (1992, 2004) found that her Norwegian/English speaking participant Siri only used functional morphemes from Norwegian, her dominant language, in mixed utterances, as seen in (13a) and (13b).

13. a. *looker*
 look+s
 ‘looks’ [Siri 2;04] (Lanza, 1992, p. 640)
- b. *hairer*
 hair + the
 ‘the hair’ [Siri 2;06] (Lanza, 2004, p. 144)

Siri mixed differently in conversation with her English-speaking mother and Norwegian-speaking father; she only used grammatical morphemes from the other language with her mother, in speaking English.

In contrast, Meisel (2007) found no evidence of qualitative differences in developmental patterns between the weaker language of bilingual children who had a dominant language (despite having been exposed to both from birth) and the two languages of balanced bilingual children. In a survey of the literature on language dominance and CLI, Nicoladis (2015) noted that the findings in the literature are inconsistent regarding the impact of dominance on CLI, in part for methodological reasons; dominance as a variable has not always been measured the same way. In addition, the effect of dominance seems to vary among bilingual children, and may require a large participant group in order to be detected.

To summarize, bilingual children's use of non-adult mixes (the seemingly indiscriminate mixing of languages) and some instances of grammatical transfer seem to be temporary stages that disappear by the age of four years. Presumably, these non-adult mixes disappear as a consequence of cognitive maturation and greater knowledge of each language. Children may also cease to use grammatical transfer as a relief strategy as their proficiency increases in their weaker language.

Several other types of cross-linguistic influence will be reviewed in the following sections, along with proposals that attempt to explain their occurrence.

Grammatical overlap and cross-linguistic influence

Several researchers have claimed that grammatical overlap between a child's languages favours cross-linguistic influence by reinforcing a non-adult like analysis in one of the languages. For example, Döpke (1998) found that bilingual children acquiring German and English used SVO word order more often than monolingual children acquiring German. Döpke claimed that the bilingual children's exposure to SVO word order in English led them to over-produce SVO order in German, even in contexts which required SOV order. Hulk (1997) and Müller (1998) had similar findings for bilingual children acquiring Dutch/French and German/French, respectively; bilingual children in both studies produced a word order that was marginal or unacceptable in language A but that was acceptable in language B.

To explain this pattern, Müller suggested that transfer could be used as a relief strategy when a bilingual child is faced with several competing interpretations of a syntactic structure. Under these circumstances, bilingual children may produce more instances of one of the options provided by the ambiguous input than a monolingual child would if it coincides with a grammatical option in the other language.

Hulk and Müller (2000) and Müller and Hulk (2001) found that bilingual children acquiring German and either French or Italian dropped more object pronouns than monolingual children acquiring these Romance languages, a result which they attributed to cross-linguistic influence from German. However, they argued that the influence is indirect in nature, and that input from German, in which object drop is grammatically acceptable, led the bilingual children to retain the option of pragmatically licensed null objects provided by a Minimal Default Grammar (Roeper, 1999) longer than a monolingual child would. Hulk and Müller hypothesized that partial overlap between a bilingual child's languages may favour cross-linguistic influence, along with the involvement of an interface between the semantics and grammar; this factor will be discussed in the following section. A study by Kupisch (2007) of four German/Italian bilingual children found that there was acceleration in the developmental pattern of German determiners in these children relative to monolingual children, whereas the bilinguals showed only a slight slowing in the acquisition of articles in Italian. Kupisch argued that both

language dominance and the existence of a syntactic overlap are decisive factors in determining the direction of language transfer in bilingual children.

However, other researchers have not found that partial overlap between a child's grammars necessarily predicts where CLI will occur. Unsworth (2003) did not find evidence of cross-linguistic influence in the distribution of root infinitives in German in a German/English bilingual child, despite it being a domain with partial overlap between German and English. Furthermore, several researchers have found evidence of cross-linguistic evidence in areas where there is no overlap between a child's languages. In their research on syntactic development in a Cantonese-English bilingual child, Yip and Matthews (2000, 2007) argued language dominance but not grammatical overlap favoured cross-linguistic transfer. These authors found evidence of syntactic transfer from Cantonese to English in their Cantonese-dominant bilingual participant aged 1;06–3;06 in the form of *wh*-in-situ interrogatives, null objects and pronominal relatives. Because these structures are not found in monolingual English development, Yip and Matthews claimed that their data point to qualitative as well as quantitative differences between bilingual and monolingual acquisition.

Interface areas and cross-linguistic influence

In addition to the presence of partial grammatical overlap discussed in the previous section, Hulk and Müller (2000) and Müller and Hulk (2001) proposed that areas of grammar involving the Complementizer system (or C-domain) are particularly susceptible to cross-linguistic influence in bilingual development. The C-domain involves operations at the interface between the syntax and pragmatics, and these areas of grammar have a protracted developmental course in monolingual and bilingual children as well as children with SLI, and aphasic adults (Meisel, 1992; Platzack, 2001). Examples of areas of grammar which involve the syntax-pragmatics interface include the expression of null vs. overt subjects in pro-drop languages and argument topicalization. In a similar vein, the Interface Hypothesis (Sorace and Filiaci, 2006) posited that near-native L2 learners experience fossilization and residual optionality in interface areas of grammar and therefore persist in errors such as overproducing overt pronouns in Italian, even at very high levels of proficiency, because of difficulty integrating different modules of language. A subsequent version of this proposal reformulated this hypothesis to claim that areas of interface between syntax and modules of language outside of the grammar were more difficult for advanced L2 learners than grammar-internal interfaces, such as the syntax-morphology interface (Sorace and Serratrice, 2009). The predictions of these hypotheses were also consistent with studies of syntactic convergence in child and adult L2 grammars which claimed that interpretable features (which are found in the peripheral or interface areas of grammar) are most vulnerable to cross-linguistic influence and attrition (Sorace, 2003; Bull-ock and Toribio, 2004; Montrul, 2004; Sánchez, 2004).

The studies that have tested the predictions of the Interface Hypothesis have found mixed support for this approach. While areas of grammar that have interfaces with pragmatics may be susceptible to cross-linguistic influence and persistent errors, such influence does not seem to be limited to these cases, and can also be found in syntactic domains such as verb raising and clitic placement in child and adult bilinguals which do not share interfaces with pragmatics (Pérez-Leroux, Cuza and Thomas, 2011; Cuza, 2013). Furthermore, cross-linguistic influence has not been found in the acquisition of interface-related areas of bilingual and L2 acquisition where the Interface Hypothesis would predict it to occur (Zwanziger, Allen and Genesee, 2005; Donaldson, 2011, 2012; Slabakova, Kempchinsky and Rothman, 2012; Domínguez and Arche, 2014).

In a review of the Interface Hypothesis, Montrul (2011) noted that while this approach makes correct predictions for persistent CLI and protracted acquisition in some interface-governed domains such as overt subject expression, other areas of grammar may have less clear or even conflicting predictions because they bring together multiple interfaces, both internal and external:

Although one may agree that a specific phenomenon (e.g., the expression of subjects in null subjects grammars) falls squarely under a particular interface (e.g., the syntax – discourse interface), many other linguistic properties are hard to characterize and may indeed involve multiple interfaces. And if we cannot extend the Interfaces Hypothesis beyond the study of null/overt subjects, topic and focus and genericity, and to other structures and linguistic domains, then theoretical and empirical advances, and falsification of the theory, are impossible.

(Montrul, 2011, p. 603)

In summary, the research findings are equivocal regarding the vulnerability of interface domains to CLI. In the next sections, we will review other potential factors which may predict where cross-linguistic influence occurs in child grammars.

Competition and cross-linguistic priming in child grammars

Recent research has investigated whether factors such as processing errors and cross-linguistic priming can account for the occurrence of CLI. In a series of experiments, Nicoladis and colleagues have investigated CLI in bilingual children acquiring languages with conflicting word orders in adjective phrases. In a study of 35 French/English age 3;03 to 5;01, Nicoladis (2006) found that the children produced more phrases in French with adjective-noun order than age-matched monolingual English-speaking children did, which was expected, given that both adjective-noun and noun-adjective orders are allowed in French, and influence from English would reinforce the adjective-noun order. Contrary to expectations, there was also influence from French to English, and the bilingual children produced more ungrammatical noun-adjective phrases in English than the monolinguals, particularly in cases where the corresponding adjective in French has a post-nominal position, such as ‘closed’ and ‘green.’ Nicoladis interpreted these results to be evidence of speech errors resulting from cross-linguistic competition that occurs at the lemma level when a speaker is selecting the correct syntactic frame for a phrase. In a subsequent study, Nicoladis, Rose and Foursha-Stevenson (2010) found that cross-linguistic influence at the lemma level did not occur in the speech of three- to five-year-old French/English bilinguals when they produced descriptive phrases in French and English of moving objects such as ‘dancing cow.’ French and English speakers prefer different expressions for these concepts; in English, a gerund+ noun phrase such as ‘dancing cow’ is preferred, whereas in French a relative clause headed by a noun is more common (*vache qui danse*, or ‘a cow that dances’). Although both types of phrase are possible in both languages, bilingual children produced the phrase preferred in each language, showing that without shared conceptual orientation as well as overlapping structure, cross-linguistic competition did not occur at the lemma level.

Nicoladis and Gavrilu, (2015) examined the production of adjective phrases in 30 bilingual children learning Welsh and English between three and six years old (average age 4;10 years) and age-matched monolingual English speakers. The bilingual children were dominant in Welsh, and spoke Welsh at home. Adjective phrases in these languages were

chosen as a test case for whether grammatical overlap is necessary for CLI because these constructions do not present overlapping options to bilingual children; in Welsh, adjectives are produced postnominally, whereas in English they occur before the noun. Nicoladis and Gavrilu found significantly more differences in reversals in noun-adjective order in the bilingual children than in the monolinguals, suggesting that overlap need not be present for CLI to occur. They analyzed the noun-adjective reversals as instances of speech errors caused by competing grammatical structures in the adult input. The authors also found that word order reversals were more likely to occur in bilinguals who were strongly dominant in Welsh than more balanced bilinguals, although the difference was not significant. This suggests that dominance may make CLI more likely to occur, but does not trigger cross-linguistic influence on its own.

Several studies have shown that cross-linguistic priming occurs in adult bilinguals when they are presented with phrases or sentences across languages that share the same word order, but not phrases that differ structurally, even if they are semantically identical (Loebell and Bock, 2003; Branigan et al., 2006; Hartsuiker et al., 2016). Recent research has found that cross-language priming effects can also be seen in bilingual children, and that younger children are more susceptible to priming than older ones (Wolleb, Sorace and Westergaard, 2018). A study of 65 Spanish/English children age 5;02–6;05 (mean 5;11) found that passive Spanish sentences primed passive English sentences (using the structurally equivalent *fue escrito* ‘it was written’ type passives), but not vice versa (Vasilyeva et al., 2010). The authors proposed that this asymmetry results from the fact that *fue escrito* passives are much less frequent and much more formal than constructions like *se escribió*, which is the most common way to say ‘it was written,’ suggesting that cross-linguistic priming occurs when phrases are structurally equivalent and roughly equally common. Another study found that four- and five-year-old bilingual children who speak Spanish and English are more likely to use an ungrammatical word order in Spanish (adjective-noun) after hearing an adjective phrase with that order in English, indicating that cross-linguistic influence can be primed, at least in children (Hsin, Legendre and Omaki, 2013). In a study of 38 bilingual French/English five-year-olds, Hervé, Serratrice and Corley (2016) found that pragmatically infelicitous structures (left-dislocated constructions) could be primed across languages, and that this effect was mitigated by the amount of exposure to each language the children had. Wolleb, Sorace and Westergaard (2018) reported that double object constructions (e.g., ‘I gave you the book’) could be primed across languages in 38 Norwegian-English bilingual children age 4;07–8;05 (mean 6;01). The results also showed that the effects of within-language priming were not significantly different from between-language priming in these children, which the authors interpret to mean that inhibitory mechanisms are not recruited during the between-languages tasks (although language control is present, given that the children are able to respond in the target language).

While significant advances have been made in the understanding of cross-linguistic influence in young bilinguals’ grammars, at this point there does not seem to be a single factor invariably linked to cross-linguistic influence (Serratrice, 2013; Hulk, 2017). The area of grammar involved, the degree of overlap between grammars, the amount of input in each language, and dominance in one language, all may contribute separately to the likelihood of CLI or could work in conjunction, in ways that are not fully understood. The interaction between ‘child-internal and child-external factors’ (Hulk, 2017) will also need to be explored in future research in order to understand how CLI is impacted by cognitive development and the environment in which the bilingual child is growing up as well as linguistic factors.

5. Main research methods

The earliest research on bilingual language development consisted of longitudinal case studies of children's natural speech, often conducted with the researchers' own children. This was true of the diary studies conducted by Ronjat (1913) of his son Louis, Leopold (1970)'s study of his daughter Hildegard's development in German and English, and Fantini's (1974) study of his bilingual Spanish/English speaking son, Mario. Through the early 2000s, case studies of natural speech samples continued to be the most common paradigm used in research on childhood bilingualism, predominantly focusing on middle class children with well-educated parents who were acquiring European languages (e.g., Vihman, 1985; De Houwer, 1990; Lanza, 1992; Meisel, 1994; Döpke, 1998). In part, this method was chosen because of the logistical difficulty of collecting longitudinal natural speech samples from multiple children; the transcription of such samples alone requires about five hours for every recorded hour of speech. While this research produced ground-breaking insights into bilingual development, nonetheless the predominance of case studies made it difficult to generalize findings across larger populations of bilinguals, or to understand on how bilingualism affects children from different demographic groups as well as children acquiring typologically distinct languages.

To this end, for the past 15 years, researchers have been using a variety of experimental methods to elicit data from bilingual children rather than relying exclusively on spontaneous speech samples, as well as to assess language comprehension. These methods included picture naming tasks (Nicoladis, 2006), grammaticality judgment tasks (Sorace and Serratrice, 2009; Lingwall, 2018), standardized assessments (Paradis, Nicoladis and Crago, 2007), and cognitive tasks such as measures of executive control (Bialystok and Martin, 2004). These experimental methods have revealed many subtle forms of cross-linguistic influence, such as the effects of cross-language priming (e.g., Vasilyeva et al., 2010). These methods have allowed researchers to include greater sample sizes and to use more powerful statistical analyses. Recent research on CLI has also included more studies of bilingual children who speak non-European languages, such as Cantonese (Yip and Matthews, 2000), Inuktitut (Zwanziger, Allen, and Genesee, 2005), and Quechua (Sánchez, 2004).

6. Conclusions

Research in the past 40 years has made significant strides in discovering why cross-linguistic influence is so pervasive in the grammars of young bilinguals. First understood as the product of a fused linguistic system which separated into two languages in maturity, CLI is now thought to be a pervasive characteristic of both child and adult grammars, which share syntactic representations and are co-activated in bilinguals of all ages. The non-adult mixes typical of the speech of young bilinguals which received much attention in the earliest studies of child bilingualism have been argued to reflect a kind of relief strategy that is employed when a child struggles to produce a form in one of his/her languages and can produce a form in the other instead. These mixes disappear with maturity. Research on CLI in bilingual children and adults has also provided insights into the nature of shared grammatical representations, which have interesting implications for theories of grammar and language change more generally.

7. Future directions

Our understanding of the role that cross-linguistic influence plays in bilingual grammars has shifted greatly since the first studies on childhood bilingualism. In addition to the research

currently being conducted, which seeks to understand and specify the circumstances under which CLI is likely to occur (cf. Blom, Cornips and Shaeffer (2017) and studies therein), researchers are investigating the extent to which shared representations are a pervasive feature of bilingual grammars, from infancy to adulthood. Studies such as Vasilyeva et al. (2010), Hsin, Legendre and Omaki (2013), Hervé, Serratrice, and Corley (2016), Nicoladis and Gavrilá (2015), and Wolleb, Sorace and Westergaard (2018) have found models of adult bilingual grammars with shared syntactic representations useful for explaining cross-linguistic priming in child bilinguals too. Such research opens the possibility that at least some of the cross-linguistic influence seen in children persists into adulthood, and should therefore be considered a permanent feature of bilingual grammars rather than the outcome of non-systematic speech errors. It also indicates that the grammars of bilingual speakers may differ in some respects from those of monolingual speakers of the same language. Recent studies also suggest that CLI may enhance or inhibit the acquisition of third languages, depending on the degree of overlap in question (Westergaard et al., 2017; Lorenz et al., 2018). Future research will need to determine just how much overlap exists between CLI in child and adult grammars, which in turn will provide insight into the nature of non-selectivity and its consequences for bilingual grammars. Such research will also require new models of how grammars store and activate syntactic representations (e.g., Lillo-Martin, de Quadros and Pichler, 2017; Putnam, Carlson and Reitter, 2018), both as an explanation for the bilingual patterns discussed in this chapter as well as for phenomena such as syntactic priming in monolinguals.

Enhanced models of cross-linguistic priming, co-activation and shared representations of syntax in bilinguals will also help us understand the connection between bilingualism, language innovation and language change. This is particularly true in situations of widespread individual bilingualism, such as in heritage language speakers whose grammars are changed as a result of language contact (Montrul, 2002, 2008; Putnam and Sánchez, 2013; Boon, 2014; Montrul, Bhatt and Girju, 2015), and minority language communities in which most or even all speakers of the minority language are bilingual, such as Welsh (Gathercole, 2007), Inuktitut (Allen, 2007) and Basque (Barreña, Ezeizabarrena and García, 2008). In some circumstances, intensive language contact may even spur the development of new languages such as Light Walpiri (O'Shannessy, 2005, 2016). Further research is needed to reveal how co-activation of shared syntactic representations in individuals can lead to syntactic convergence and eventually, linguistic change in bilingual communities.

8. Further reading

Blom, E., Cornips, L. and Schaeffer, J., eds. (2017). *Cross-linguistic influence in bilingualism: In Honor of Aafke Hulk*, vol. 52. Amsterdam: John Benjamins.

This edited volume brings together research on cross-linguistic influence in child and adult grammars in a variety of language and dialect combinations. These papers also examine how CLI influences development in a wide range of bilingual learners, including children with Specific Language Impairment.

Serratrice, L. (2013). Cross-linguistic influence in bilingual development: determinants and mechanisms. *Linguistic Approaches to Bilingualism*, 3(1), pp. 3–25.

This article surveys the literature on cross-linguistic influence in child grammars, with a focus on CLI in null subject and topic drop languages.

Lillo-Martin, D., de Quadros, R. and Pichler, D. (2017). The development of bimodal bilingualism. *Linguistic Approaches to Bilingualism*, 6(6), pp. 719–755.

This keynote article proposes the Language Synthesis Model as a framework for understanding bilingual language representations that can account for patterns observed in CLI in bimodal bilinguals.

Putnam, M., Carlson, M. and Reitter, D. (2018). Integrated, not isolated: Defining typological proximity in an integrated multilingual architecture. *Frontiers in Psychology*, 8, p. 2212.

This article proposes a new model for bilingual grammars which assumes that they are partially separate but joined by a shared grammar; the model takes the form of a parallel, distributed, multi-vector network based on Hsin (2014).

O'Shannessy, C. (2016). Distributions of case allomorphy by multilingual children. *Linguistic Variation*, 16(1), pp. 68–102.

This article discusses the role that CLI in the speech of child and adult bilinguals plays in furthering language change, focusing on Light Warlpiri speakers.

9. Related topics

Borrowing, convergence, mixed languages

Abbreviations

| | |
|------|------------------------------|
| 3 | third person |
| ABS | absolute |
| AUX | auxiliary |
| CLI | cross-linguistic influence |
| COMP | complementizer |
| ERG | ergative |
| INFL | inflection |
| IP | inflection phrase |
| L1 | first language |
| L2 | second language |
| MLU | mean length of utterance |
| NP | noun phrase |
| O | object |
| PL | Plural |
| S | subject |
| SG | Singular |
| SLI | Specific Language Impairment |
| V | verb |
| VP | verb phrase |

References

- Allen, S. (2007). The future of Inuktitut in the face of majority languages: Bilingualism or language shift? *Applied Psycholinguistics*, 28(3), pp. 515–536.
- Arnaus Gil, L., Eichler, N., Jansen, V., Patuto, M. and Müller, N. (2012). The syntax of mixed DPs containing an adjective: Evidence from bilingual German-Romance (French, Italian, Spanish) children. In: K. Geeslin and M. Díaz-Campos, eds., *Selected Proceedings of the 14th Hispanic Linguistics Symposium*. Somerville, MA: Cascadilla Proceedings Project, pp. 242–257.

- Austin, J. (2001). Language differentiation and the development of morphosyntax in bilingual children acquiring Basque and Spanish. PhD. Cornell University.
- Austin, J. (2007). Grammatical interference and the acquisition of ergative case in bilingual children learning Basque and Spanish. *Bilingualism: Language and Cognition*, 10(3), pp. 315–331.
- Austin, J. (2009). Delay, interference and bilingual development: The acquisition of verbal morphology in children learning Basque and Spanish. *International Journal of Bilingualism*, 13(4), pp. 447–479.
- Austin, J. (2015). Transfer and contact-induced variation in child Basque. *Frontiers in Psychology*, 5, p. 1576.
- Barreña, A., Ezeizabarrena, M.J. and García, I. (2008). Influence of the linguistic environment on the development of the lexicon and grammar of Basque bilingual children. In: C. Pérez-Vidal, M. Juan-Garau, and A. Bel, eds., *A portrait of the young in the new multilingual Spain*, 1st ed. Clevedon, UK: Multilingual Matters, pp. 86–110.
- Barreña, A. and Idiazabal, I. (1997). *Diferenciación bilingüe precoz de códigos: Los objetos en español y euskara*. M.S. University of Salamanca and University of Basque Country.
- Bernardini, P. and Schlyter, S. (2004). Growing syntactic structure and code-mixing in the weaker language: The ivy hypothesis. *Bilingualism: Language and Cognition*, 7(1), pp. 49–69.
- Bialystok, E. and Martin, M.M. (2004). Attention and inhibition in bilingual children: Evidence from the dimensional change card sort task. *Developmental Science*, 7(3), pp. 325–339.
- Blom, E. (2010). Effects of input on the early grammatical development of bilingual children. *International Journal of Bilingualism*, 14(4), pp. 422–446.
- Blom, E., Cornips, L. and Schaeffer, J., eds. (2017). *Cross-linguistic influence in bilingualism: In honor of Aafke Hulk*, vol. 52. Amsterdam: John Benjamins.
- Boon, E. (2014). *Heritage Welsh: A study of heritage language as the outcome of minority language acquisition and bilingualism*. PhD. Harvard University.
- Branigan, H., Pickering, M., McLean, J. and Stewart, A. (2006). The role of local and global syntactic structure in language production: Evidence from syntactic priming. *Language and Cognitive Processes*, 21(7–8), pp. 974–1010.
- Bullock, B.E. and Toribio, A.J. (2004). Convergence as an emergent property in bilingual speech. *Bilingualism: Language and Cognition*, 7, pp. 91–93.
- Costa, A., Colomé, A., Gómez, O. and Sebastián-Gallés, N. (2003). Another look at cross-language competition in bilingual speech production: Lexical and phonological factors. *Bilingualism: Language and Cognition*, 6(3), pp. 167–179.
- Cuza, A. (2013). Crosslinguistic influence at the syntax proper: Interrogative subject-verb inversion in heritage Spanish. *International Journal of Bilingualism*, 17(1), pp. 71–96.
- De Houwer, A. (1990). *The acquisition of two languages from birth: A case study*. Cambridge: Cambridge University Press.
- Desmet, T. and Declercq, M. (2006). Cross-linguistic priming of syntactic hierarchical configuration information. *Journal of Memory and Language*, 54(4), pp. 610–632.
- Domínguez, L. and Arche, M.J. (2014). Subject inversion in non-native Spanish. *Lingua*, 145, pp. 243–265.
- Donaldson, B. (2011). Left dislocation in near-native French. *Studies in Second Language Acquisition*, 33(3), pp. 399–432.
- Donaldson, B. (2012). Syntax and discourse in near-native French: Clefts and focus. *Language Learning*, 62, pp. 902–930.
- Döpke, S. (1998). Competing language structures: The acquisition of verb placement by bilingual German-English children. *Journal of Child Language*, 25, pp. 555–584.
- Ezeizabarrena, M. (1996). *Adquisición de la morfología verbal en euskera y castellano por niños bilingües*. Bilbao: Servicio Editorial de la Universidad del País Vasco.
- Fantini, A. (1974). *Language acquisition of a bilingual child: A sociolinguistic perspective*. PhD. University of Texas at Austin.
- Gathercole, V. (2002). Monolingual and bilingual acquisition: Learning different treatments of *that*-trace phenomena. In: D.K. Oller and R. Eilers, eds., *Language and literacy in bilingual children*, 1st ed. Clevedon: Multilingual Matters, pp. 220–254.

- Gathercole, V. (2007). Miami and North Wales, so far and yet so near: A constructivist account of morphosyntactic development in bilingual children. *The International Journal of Bilingual Education and Bilingualism*, 10(3), pp. 224–247.
- Gawlitzeck-Maiwald, I. and Tracy, R. (1996). Bilingual bootstrapping. *Linguistics*, 34(5), pp. 901–926.
- Genesee, F., Nicoladis, E. and Paradis, J. (1995). Language differentiation in early bilingual development. *Journal of Child Language*, 22(3), pp. 611–631.
- Hartsuiker, R., Beerts, S., Loncke, M., Desmet, T. and Bernolet, S. (2016). Cross-linguistic structural priming in multilinguals: Further evidence for shared syntax. *Journal of Memory and Language*, 90, pp. 14–30.
- Hervé, C., Serratrice, L. and Corley, M. (2016). Dislocations in French – English bilingual children: An elicitation study. *Bilingualism: Language and Cognition*, 19(5), pp. 987–1000.
- Hoff, E., Core, C., Place, S., Rumiche, R., Señor, M. and Parra, M. (2012). Dual language exposure and early bilingual development. *Journal of Child Language*, 39(1), pp. 1–27.
- Hsin, L. (2012). Accelerated acquisition in Spanish-English bilinguals: The structural transfer hypothesis. In: J. Choi, E. Hogue, J. Punske, D. Tat, J. Schertz, and A. Trueman, eds., *Proceedings of the 29th West Coast Conference on Formal Linguistics*, Somerville, MA: Cascadilla Proceedings Project, pp. 108–116.
- Hsin, L. (2014). *Integrated bilingual grammatical architecture: Insights from syntactic development*. PhD. Johns Hopkins University.
- Hsin, L., Legendre, G. and Omaki, A. (2013). Priming cross-linguistic interference in Spanish – English bilingual children. In: S. Baiz, N. Goldman, and R. Hawkes, eds., *Proceedings of the 37th annual Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press, pp. 165–177.
- Hulk, A. (1997). The acquisition of French object pronouns by a Dutch/French bilingual child. In: A. Sorace, C. Heycock, and R. Shillcock, eds., *Language acquisition: Knowledge, representation and processing. Proceedings of the GALA 1997 Conference on Language Acquisition*. Edinburgh: University of Edinburgh, pp. 521–526.
- Hulk, A. (2017). Note on cross-linguistic influence: Back to ‘MULK.’ In: E. Blom, L. Cornips, and J. Schaeffer, eds., *Cross-linguistic influence in bilingualism: In honor of Aafke Hulk*, vol. 52, 1st ed. Amsterdam: John Benjamins, pp. 16–24.
- Hulk, A. and Müller, N. (2000). Bilingual first language acquisition at the interface between syntax and pragmatics. *Bilingualism: Language and Cognition*, 3(3), pp. 227–244.
- Köppe, R. and Meisel, J. (1995). Codeswitching in bilingual first language acquisition. In: L. Milroy and P. Muysken, eds., *One speaker – Two languages: Cross-disciplinary perspectives on code-switching*, 1st ed. Cambridge: Cambridge University Press, pp. 276–301.
- Kroll, J., Bobb, S. and Wodniecka, Z. (2006). Language selectivity is the exception, not the rule: Arguments against a fixed locus of language selection in bilingual speech. *Bilingualism: Language and Cognition*, 9(2), pp. 119–135.
- Kupisch, T. (2007). Determiners in bilingual German-Italian children: What they tell us about the relation between language influence and language dominance. *Bilingualism: Language and Cognition*, 10(1), pp. 57–78.
- Lanza, E. (1992). Can bilingual two year olds code switch? *Journal of Child Language*, 19, pp. 633–658.
- Lanza, E. (2004). *Language mixing in infant bilingualism: A sociolinguistic perspective*. Oxford: Oxford University Press.
- Larrañaga, P. and Guijarro-Fuentes, P. (2012). Clitics in L1 bilingual acquisition. *First Language*, 32(1–2), pp. 151–175.
- Leopold, W. (1970). *Speech development of a bilingual child: A linguist's record*. New York: AMS Press.
- Liceras, J., Fuertes, R. and de la Fuente, A. (2012). Overt subjects and copula omission in the Spanish and the English grammar of English – Spanish bilinguals: On the locus and directionality of interlinguistic influence. *First Language*, 32(1–2), pp. 88–115.
- Lillo-Martin, D., de Quadros, R.M. and Pichler, D.C. (2017). The development of bimodal bilingualism. *Linguistic Approaches to Bilingualism*, 6(6), pp. 719–755.

- Lindholm, K. and Padilla, A. (1978). Language mixing in bilingual children. *Journal of Child Language*, 5(2), pp. 327–335.
- Lingwall, A. (2018). *Grammatical acceptability among Spanish monolingual and Spanish-English bilingual children: The acquisition of Ser and Estar*. PhD. Rutgers University.
- Lleó, C. and Kehoe, M. (2002). On the interaction of phonological systems in child bilingual acquisition. *International Journal of Bilingualism*, 6(3), pp. 233–237.
- Loebell, H. and Bock, K. (2003). Structural priming across languages. *Linguistics*, 41(5), pp. 791–824.
- Lorenz, E., Bonnie, R.J., Feindt, K., Rahbari, S. and Siemund, P. (2018). Cross-linguistic influence in unbalanced bilingual heritage speakers on subsequent language acquisition: Evidence from pronominal object placement in ditransitive clauses. *International Journal of Bilingualism*, 23(6), pp. 1410–1430.
- Meijer, P. and Fox Tree, J. (2003). Building syntactic structures in speaking: A bilingual exploration. *Experimental Psychology*, 50(3), pp. 184–195.
- Meisel, J. (1989). Early differentiation of languages in bilingual children. In: K. Hyltenstam and L. Obler, eds., *Bilingualism across the life span. Aspects of acquisition, maturity, and loss*, 1st ed. Cambridge: Cambridge University Press, pp. 13–40.
- Meisel, J., ed. (1992). *The acquisition of verb placement: Functional categories and V2 phenomena in language acquisition*. Dordrecht: Kluwer.
- Meisel, J., ed. (1994). *Bilingual first language acquisition: French and German grammatical development*. Philadelphia: John Benjamins.
- Meisel, J. (2007). The weaker language in early child bilingualism: Acquiring a first language as a second language? *Applied Psycholinguistics*, 28(3), pp. 495–514.
- Meroni, L., Smeets, L. and Unsworth, S. (2017). Cross-linguistic influence in scope ambiguity. In: E. Blom, L. Cornips, and J. Schaeffer, eds., *Cross-linguistic influence in bilingualism: In honor of Aafke Hulk*, vol. 52, 1st ed. Amsterdam: John Benjamins, pp. 182–206.
- Montrul, S. (2002). Incomplete acquisition and attrition of Spanish tense/aspect distinctions in adult bilinguals. *Bilingualism: Language and Cognition*, 5(1), pp. 39–68.
- Montrul, S. (2004). *The acquisition of Spanish: Morphosyntactic development in monolingual and bilingual L1 acquisition and adult L2 acquisition*. Amsterdam and Philadelphia: John Benjamins.
- Montrul, S. (2008). *Incomplete acquisition in bilingualism: Re-examining the age factor*. Amsterdam and Philadelphia: John Benjamins.
- Montrul, S. (2011). Multiple interfaces and incomplete acquisition. *Lingua*, 121(4), pp. 591–604.
- Montrul, S., Bhatt, R. and Girju, R. (2015). Differential object marking in Spanish, Hindi, and Romanian as heritage languages. *Language*, 91(3), pp. 564–610.
- Müller, N. (1998). Transfer in bilingual first language acquisition. *Bilingualism: Language and Cognition*, 1(3), pp. 151–171.
- Müller, N. and Hulk, A. (2001). Crosslinguistic influence in bilingual language acquisition: Italian and French as recipient languages. *Bilingualism: Language and Cognition*, 4, pp. 1–21.
- Nicoladis, E. (2006). Cross-linguistic transfer in adjective – noun strings by preschool bilingual children. *Bilingualism: Language and Cognition*, 9(1), pp. 15–32.
- Nicoladis, E. (2015). Measuring language dominance in bilingual children: Implications for crosslinguistic influence. In: J. Treffers-Daller and C. Silva-Corvalán, eds., *Language dominance in bilinguals: Issues of measurement and operationalization*, 1st ed. Cambridge: Cambridge University Press, pp. 219–234.
- Nicoladis, E. and Gavrilu, A. (2015). Cross-linguistic influence in Welsh – English bilingual children’s adjectival constructions. *Journal of Child Language*, 42(4), pp. 903–916.
- Nicoladis, E. and Genesee, F. (1997). The role of parental input and language dominance in bilingual children’s code-mixing. In: E. Hughes, M. Hughes, and A. Greenhill, eds., *Proceedings of the 21st annual Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press, pp. 422–432.
- Nicoladis, E. and Marchak, K. (2011). Le carte blanc or la carte blanche? Bilingual children’s acquisition of French adjective agreement. *Language Learning*, 61(3), pp. 734–758.

- Nicoladis, E., Palmer, A. and Marentette, P. (2007). The role of type and token frequency in using past tense morphemes correctly. *Developmental Science*, 10(2), pp. 237–254.
- Nicoladis, E., Rose, A. and Foursha-Stevenson, C. (2010). Thinking for speaking and cross-linguistic transfer in preschool bilingual children. *International Journal of Bilingual Education and Bilingualism*, 13(3), pp. 345–370.
- Nicoladis, E., Song, J. and Marentette, P. (2012). Do young bilinguals acquire past tense morphology like monolinguals, only later? Evidence from French – English and Chinese – English bilinguals. *Applied Psycholinguistics*, 33(3), pp. 457–479.
- O’Shannessy, C. (2005). Light Warlpiri: A new language. *Australian Journal of Linguistics*, 25(1), pp. 31–57.
- O’Shannessy, C. (2016). Entrenchment of light Warlpiri morphology. In: F. Meakins and C. O’Shannessy, eds., *Loss and renewal: Australian languages since colonisation*, 1st ed. Berlin: De Gruyter, pp. 217–252.
- Paradis, J. and Genesee, F. (1997). On continuity and the emergence of functional categories in bilingual first-language acquisition. *Language Acquisition*, 6(2), pp. 91–124.
- Paradis, J., Nicoladis, E. and Crago, M. (2007). French-English bilingual children’s acquisition of the past tense. In: *Proceedings of the 31st annual Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press, pp. 497–507.
- Paradis, J., Nicoladis, E., Crago, M. and Genesee, F. (2011). Bilingual children’s acquisition of the past tense: A usage-based approach. *Journal of Child Language*, 38(3), pp. 554–578.
- Pérez-Leroux, A.T., Cuza, A. and Thomas, D. (2011). Clitic placement in Spanish – English bilingual children. *Bilingualism: Language and Cognition*, 14(2), pp. 221–232.
- Petersen, J. (1988). Word-internal code switching constraints in a bilingual child’s grammar. *Linguistics*, 26, pp. 479–493.
- Platzack, C. (2001). The vulnerable C-domain. *Brain and Language*, 77(3), pp. 364–377.
- Putnam, M.T., Carlson, M. and Reitter, D. (2018). Integrated, not isolated: Defining typological proximity in an integrated multilingual architecture. *Frontiers in Psychology*, 8, p. 2212.
- Putnam, M.T. and Sánchez, L. (2013). What’s so incomplete about incomplete acquisition?: A prolegomenon to modelling heritage language grammars. *Linguistic Approaches to Bilingualism*, 3(4), pp. 478–508.
- Redlinger, W. and Park, T. (1980). Language mixing in young bilinguals. *Journal of Child Language*, 7, pp. 337–352.
- Roeper, T. (1999). Universal bilingualism. *Bilingualism: Language and Cognition*, 2(3), pp. 169–186.
- Ronjat, J. (1913). *Le Développement du Langage Observé Chez un Enfant Bilingue*. Paris: Librairie Ancienne H. Champion, Éditeur.
- Sánchez, L. (2004). Functional convergence in the tense, evidentiality and aspectual systems of Quecha Spanish bilinguals. *Bilingualism: Language and Cognition*, 7, pp. 147–162.
- Serratrice, L. (2013). Cross-linguistic influence in bilingual development: Determinants and mechanisms. *Linguistic Approaches to Bilingualism*, 3(1), pp. 3–25.
- Slabakova, R., Kempchinsky, P. and Rothman, J. (2012). Clitic-doubled left dislocation and focus fronting in L2 Spanish: A case of successful acquisition at the syntax – discourse interface. *Second Language Research*, 28(3), pp. 319–343.
- Sorace, A. (2003). Near-nativeness. In: C. Doughty and M. Long, eds., *Handbook of second language acquisition*, 1st ed. Oxford: Blackwell, pp. 130–151.
- Sorace, A. and Filiaci, F. (2006). Anaphora resolution in near-native speakers of Italian. *Second Language Research*, 22(3), pp. 339–368.
- Sorace, A. and Serratrice, L. (2009). Internal and external interfaces in bilingual language development: Beyond structural overlap. *International Journal of Bilingualism*, 13(2), pp. 195–210.
- Syrett, K., Lingwall, A., Perez-Cortes, S., Austin, J., Sánchez, L., Baker, H., Germak, C. and Arias-Amaya, A. (2017). Differences between Spanish monolingual and Spanish-English bilingual children in their calculation of entailment-based scalar implicatures. *Glossa Special Issue: Acquisition of Quantification*, 2(1), pp. 1–19.

- Toribio, J. and Brown, B. (1995). Language contact and differentiation in child bilingualism: A syntactic analysis. In: D. MacLaughlin and S. McEwen, eds., *Proceedings of the 19th annual Boston University Conference on Language Development*. Somerville, MA: Cascadilla Press, pp. 629–642.
- Unsworth, S. (2003). Testing Hulk and Müller 2000 on cross-linguistic influence: RIs in a bilingual German/English child. *Bilingualism: Language and Cognition*, 6, pp. 143–158.
- Vasilyeva, M., Waterfall, H., Gámez, P.B., Gómez, L.E., Bowers, E. and Shimpi, P. (2010). Cross-linguistic syntactic priming in bilingual children. *Journal of Child Language*, 37(5), pp. 1047–1064.
- Veh, B. (1990). Syntaktische Aspekte des code-switching bei bilingualen Kindern (Französisch-Deutsch) im Vorschulalter. Ms. University of Hamburg.
- Vihman, M. (1985). Language differentiation by the bilingual child. *Journal of Child Language*, 12, pp. 297–324.
- Volterra, V. and Taeschner, T. (1978). The acquisition and development of language by bilingual children. *Journal of Child Language*, 5, pp. 311–326.
- Westergaard, M., Mitrofanova, N., Mykhaylyk, R. and Rodina, Y. (2017). Crosslinguistic influence in the acquisition of a third language: The linguistic proximity model. *International Journal of Bilingualism*, 21(6), pp. 666–682.
- Wolleb, A., Sorace, A. and Westergaard, M. (2018). Exploring the role of cognitive control in syntactic processing: Evidence from cross-language priming in bilingual children. *Linguistic Approaches to Bilingualism*, 8(5), pp. 606–636.
- Yip, V. and Matthews, S. (2000). Syntactic transfer in a Cantonese-English bilingual child. *Bilingualism: Language and Cognition*, 3, pp. 193–208.
- Yip, V. and Matthews, S. (2007). *The bilingual child: Early development and language contact*. Cambridge: Cambridge University Press.
- Zwanziger, E., Allen, S. and Genesee, F. (2005). Crosslinguistic influence in bilingual acquisition: Subject omission in learners of Inuktitut and English. *Journal of Child Language*, 32(4), pp. 893–909.

First language attrition in the twenty-first century

How continued L1 contact in the digital age fuels language attrition theorizing

Merel Keijzer

1. Introduction

First language (L1) attrition is the (partial) language loss attested in healthy individuals who stop routinely using their mother tongue, most often following a move to a second language (L2) environment. Over the years, the phenomenon has attracted much interest, from the general public who have (directly or indirectly) experienced attrition, but also from researchers who see the field of attrition as ‘promising for the exploration of links between the brain, mind and external factors that are also of interest in multilingualism’ (Köpke, 2007, p. 10). Because attrition incurs a sudden shift from a single (where only the L1 is spoken on a daily basis) to a dual language setting (where the L2 is used alongside the L1), the resulting changes in language and cognitive control can be uniquely studied (cf. Green, 2011).

It then seems counterintuitive that past studies have not taken full advantage of this radical shift from single to dual language use patterns, instead targeting speakers with a minimal L2 immersion period of ten years. Moreover, by imposing this ten-year threshold as an inclusion criterion, the L1 vs. L2 dynamic that speakers who become immersed in an L2 environment experience cannot be examined either. This becomes crucial in the current age: whereas the subjects of earlier attrition studies were mostly migrants who were relatively isolated from their L1 environment, speakers nowadays have (social) media, applications like Skype and Facetime, and cheap airline tickets to ensure that they are never truly cut off from their L1 environments. These effects of globalization have led to a revision in labels too: immigrants now often refer to themselves as expats. The question then becomes if attrition as a phenomenon can still be studied in isolation the way it has been approached in the past. Pivotal in this respect is recent work that explores the influence of an L2 on the L1 in the initial phase of L2 learning and labels this attrition; such studies examine to what extent the L1 inhibition that is needed to avoid interference in the initial phases of an international move or even in the initial phases of L2 mastery in the home environment has consequences for L1 retrieval (cf. Levy et al., 2007; Linck, Kroll and Sunderman, 2009). It is essential to relate this work to L1 attrition theories.

This chapter presents a state-of-the art review of the field of first language attrition. This is done by first of all providing a brief overview of the field so far, with a special focus on the current psycholinguistic perspective of L1 changes in the early stages of L2 development. The critical questions and issues that dominate the field are presented, as well as a review of methods that have been and are currently used to tackle these questions. The chapter ends with an outlook to future directions.

2. Historical overview

For almost 40 years, researchers have tried to uncover L1 attrition. The question whether it is possible to lose your mother tongue is intriguing and very topical for the now more than 232 million people worldwide residing in another country than their birth country (United Nations, 2013), and therefore often have to use another language or language variety. Research into the phenomenon has tended to focus on people who have long been away from their native language environment, with an imposed threshold of ten years as the norm (Schmid, 2013). The initial attrition studies to observe this time trajectory were case studies as well as theoretical contributions that outlined how methods used in second language acquisition research could be adapted to fit language attrition research (Lambert and Freed, 1982). Although these case studies kept shaping the field for the years following the initial attrition explorations, for instance in the form of letters sent over the years by speakers who left their first language environments (e.g., Hutz, 2004), in the 1990s attrition studies multiplied, mostly in the form of PhD projects. Because of the short duration of PhD projects, these were almost invariably cross-sectional investigations, where anywhere between 20 and 50 attriters (defined as people who moved away from their first language environment) were compared to a control group of speakers still residing in the home country (but see Schmid, 2013 for a detailed account as to why establishing a baseline or control group is methodologically challenging). The only truly longitudinal work involving multiple speakers who were followed over several decades was conducted by de Bot and Clyne (1989, 1994). The most poignant findings to emerge from this work are the subtle nature of attrition (with overtly noticeable changes often being hard to detect) and the substantial individual variation that characterizes it: while residing in another language environment for a long time causes extensive attrition in some speakers, it appears to leave others virtually untouched. This latter outcome is hard to reconcile with cases that show dramatic attrition early on, i.e., within five years of the move (e.g., Hutz, 2004). Faced with this discrepancy, the field of attrition around the year 2010 appeared to have become stuck at a point where the subtle and variable long-term attrition reported in many studies could not be united with the substantial immediate attrition reported by others. And yet a unified account of how attrition operates across time is a prerequisite to answer two of the most pivotal and outstanding questions the field faces: (1) Is the main cause of L1 attrition reduced L1 or increased L2 use? And (2) Is L1 attrition permanent or temporary?

This does not mean that insights into the mechanisms underlying attrition have not been obtained. Past studies have successfully shed light on many questions, one of the main ones being which language domains are most susceptible to erosion. There is now a general consensus that the first noticeable changes occur in the lexicon and that the morphological and syntactic domains are generally more resilient to changes (cf. Schmid, 2012). Within the lexical domain, attrition manifestations have been found to range from prolonged pauses and more disfluent speech when searching for words to code-switches from the L2 (Köpke and Schmid, 2004). The reason for the syntactic and morphological domains being less prone to erosion – with syntax appearing most robust of the two (Håkansson, 1995; Hutz, 2004) – may

well be that these domains are more embedded into the structural system of language; whereas low-frequency words may drop below a given activation threshold, morphology and syntax will always need to be used when speaking the L1 (Keijzer, 2007). The order of attrition has sometimes been approached from an ‘implicational hierarchies’ perspective: if syntax is affected, then effects in the morphological and lexical domains will most likely also be found. Conversely, if lexical attrition is attested, this does not necessarily mean that attrition will have encroached on the syntactic and morphological domains (Keijzer, 2010). The position of phonetics and phonology in relation to attrition has received less attention, but the available evidence suggests that phonological attrition is not a default consequence of moving abroad (cf. de Leeuw, Schmid and Mennen, 2010; Hopp and Schmid, 2013). Despite these consistent findings relating to the order of attrition, past work has also stressed the individuality of attrition processes: although the order in which domains are affected remains largely the same, individuals do differ greatly as to the time trajectory during which they show attrition effects.

In an attempt to explain the individual variation, past studies have related different predictor variables to the magnitude of attrition attested. Although language internal factors have been taken into account (e.g., the typological proximity of the two languages under investigation – cf. Keijzer, 2008), most earlier work has concentrated on relating language external factors to attrition outcomes. In one of the most comprehensive accounts to date, Schmid and Dusseldorp (2010) outline the factors that have typically been related to attrition: biological age, firstly, has been used as a predictor in two ways: age at time of emigration and age at testing. The time in between, i.e., length of residence in an L2 environment, has also been seen as an important factor in predicting attrition. It has been consistently shown that young children show invariably more attrition than adults, a finding that follows from a (reverse) critical period: it is claimed to be easier to pick up an L2 in early childhood but also easier to lose an L1 (Bylund, 2009). Equally important is that children not yet literate have no means to maintain their L1 through written sources such as books or the internet (Schmitt, 2010). To test attrition rather than incomplete acquisition, most attrition studies target individuals who were at least 15 years old upon emigration. Past work has not systematically treated age at testing as important other than as an indication of a speaker’s length of residence since the move. Recently, however, Higby and colleagues (2019) pointed out that many attrition subjects have been older adults; a direct consequence of the ten-year threshold coupled with the requirement that individuals should have been at least 15 when moving to an L2 environment. Ageing is known to lead to cognitive and processing speed changes, making it hard to distinguish between ageing and attrition effects.

Length of residence has not uniformly been established to modulate attrition outcomes but has been one of the most fiercely debated predictors. Following Ebbinghaus’ forgetting theory (1885), a number of researchers in the early years of attrition research have stipulated a retention plateau, claiming that most attrition takes place in the first few years and with a ‘perma-store’ of language knowledge, which causes attrition effects to level in subsequent years (Bahrnick, 1984). In the absence of clear evidence for such a plateau (see Murtagh, 2003 for an overview), many researchers since then have adopted a ‘longer away is more attrition’ stance.

Overriding length of residence appears to be education and, by extension, L1 proficiency reached before the onset of attrition. Again, building on the idea of a critical threshold, it has frequently been found that those individuals with a higher educational background build up a working knowledge and proficiency of their first language that acts as a safeguard against attrition (Neisser, 1984). Conversely, more highly educated individuals also tend to be the ones whose L2 proficiency is highest. This then refutes the common assumption that speakers whose L2 mastery is high will show substantial L1 attrition (Keijzer, 2007).

Another factor that has been inextricably linked to L1 attrition relates to attitudes and motivation, including linguistic emotions in a broader sense. Whereas motivation to learn the L2 and/or maintain the L1 should intuitively play a role, attitudes towards the L1 and L2 have proven most important in explaining attrition outcomes (Ben-Rafael and Schmid, 2007; Schmid and Dusseldorp, 2010). Attitudes are most often tested by means of (a modified version of) Robert Gardner's Attitude and Motivational Test Battery (AMTB, Gardner, 2010). More specifically, a speaker's level of identification with either the first or second language and culture (known as Ethnolinguistic vitality and tested through the Subjective Ethnolinguistic Vitality Questionnaire (SVQ) developed by Bourhis, Giles and Rosenthal, 1981) has been shown to relate to the magnitude of attrition (cf. Yagmur, 1997).

Finally, studies examining isolated factors (e.g., gender) in relation to attrition have not yielded clear findings (see Schmid and Dusseldorp, 2010 for a discussion). It is the mixed findings in relation to such stand-alone predictors that have led Schmid and Dusseldorp to postulate that no attrition study should look at extralinguistic variables in isolation. Instead, they propose an interplay perspective where different language external factors (e.g., age, attitudes, continued L1 use, etc.) as well as internal factors (most notably the languages involved) are incorporated in what they call the 'multifactorial web of language internal and language external influences' (Schmid and Dusseldorp, 2010, p. 126). Using a multifactorial statistical model based on Principle Component Analysis, the authors found a number of reliable factor groupings in relation to attrition: (1) identification and affiliation with the L1; (2) continued amount and type of L1 exposure; and (3) attitudes towards the L1. It should be pointed out that the languages under investigation themselves never significantly contributed to the model.

In short, past attrition studies have constituted the field to arrive at the current understanding of attrition as a complex phenomenon that results from a complex interplay of factors that are inextricably linked and whose clustering likely varies depending on individual circumstances and speakers, making it hard to predict the attrition trajectory of a given speaker (see Keijzer and de Bot, 2019 for a more detailed discussion).

3. Critical issues and topics

Various attrition questions have thus been raised and addressed over the years, but despite the insights about attrition that have been obtained thus far, two important questions remain unanswered:

- 1 Is L1 attrition mainly caused by reduced L1 or increased L2 use?
- 2 Is L1 attrition permanent or temporary?

In relation to the first question, proponents of both reduced L1 as well as increased L2 use can be found. Scherag et al. (2004, p. 98) strongly put forward that attrition is 'the loss of skills in L1 due to non-use.' In child migrants, too, Bylund (2009) found L1 non-use to be more of a decisive factor driving L1 attrition than L2 acquisition. Köpke (2004) acknowledges that frequency of L1 use is an important moderator of L1 attrition and this is underscored by Schmid and Dusseldorp (2010), who found that the factor with most predictive power over attrition was continued L1 use in professional (work) settings. At the same time, it is important to realize that many of these studies approach attrition as a storage rather than processing phenomenon: from a predominantly linguistic perspective, they view attrition as leading to changes in the representation of the L1. Only in relatively recent years have psycholinguistic and neurolinguistics approaches to attrition been fully incorporated and in a relatively short

time span the idea that language attrition is governed by processes and mechanisms that are fundamental in the human mind/brain has steadily become a pillar underlying most attrition research (Köpke and Keijzer, 2019). That has also meant a focus shift, where the bidirectional influence of the L2 on the L1 and vice versa that results from competition in processing multiple languages in one mind is stressed more than changes to the L1 alone. Under such a view, L1 non-use alone cannot account for attrition, nor can L2 ‘overuse’ be the main driver of attrition (cf. Pavlenko, 2004). To avoid this issue altogether, some attrition researchers have preferred formulations such as L2 influence on L1 attrition (Cook, 2003). But this does beg the question where and how attrition as a phenomenon is distinct from general multilingual contact studies. Most obviously, the defining characteristic of attrition over L2 influence on the L1 should be sought in the (ir)reversibility of the L1 changes implied in both terms.

Nevertheless, such a clear-cut distinction cannot be maintained either. Upon the very formation of the field, Seliger (1996, p. 606) defined L1 attrition as ‘the temporary or permanent loss of language ability,’ leaving both options open. With the advent of increasingly more psycholinguistic and neurolinguistics approaches to attrition research (cf. Köpke and Keijzer, 2019, for an overview), the strong position that non-pathological L1 attrition is permanent seems untenable. Supporting evidence comes from studies that have found the L1 to resurface in older adults that may before have been categorized as having lost parts of their first language, but whose L1 becomes stronger again as they age, a phenomenon referred to as language reversion (Schmid and Keijzer, 2009). Converging evidence of an L1 that may show retrieval difficulties in processing rather than representational alteration comes from studies that have found suppressed L1s to resurface following hypnosis (Footnick, 2007), or retraining studies that have found that even international adoptees who were adopted as young children relearned their native language more easily compared to a control group who had never been exposed to that language at an earlier point in their lives (Zhou, 2015). Collectively, these findings invoke the so-called savings paradigm: the savings paradigm was first related to language attrition by de Bot and Stoessel (2000), who contrasted the outcomes of a Dutch language test of two participants, both German adults who had lived in the Netherlands for four years when they were children, to those of speakers who had never been exposed to Dutch before. The two ‘relearners’ did markedly better than the control group. The study by de Bot and Stoessel marked the start of a series of studies on vocabulary relearning of languages acquired on the basis of natural exposure, collectively reviewed by Hansen, Umeda and McKinney (2002), who also conducted their own study on language attrition and savings of Japanese and Korean as L2s. Because the observed substantial savings effects across all these investigations (where attriters almost always learned their ‘old’ language much faster compared to speakers acquiring that language for the first time), Hansen et al. suggested that subsequent research into relearning should address the question of which variables affect the size of savings, for instance the original proficiency in the attrited language and age at learning and testing.

Attempting to reconcile these findings with other (albeit scarcer) studies that failed to find a trace of the native language in international adoptees (cf. Ventureyra, Pallier and Yoo, 2004), Schmid (2011) posits that it does not matter whether in the long run language is irretrievably lost or only appears to be ‘misplaced’; if language cannot be used when needed, communication breaks down and language is effectively lost. Although it is true that such a distinction may not matter in everyday communication, it does matter in terms of theorizing about attrition as a phenomenon in its own right or as an instantiation of bilingualism and bilingual processing accounts. With the globalization effects that have led the term expat to be preferred over migrant and with that shorter-term investigations of L1 changes following L2 influence, there has never been a better time to assess this issue.

4. Current contributions and research

On the basis of psycholinguistic processing accounts of attrition, the two preceding questions (relating to the source and (ir)reversibility of L1 attrition) are more closely related than is assumed under a representational perspective: they pertain to the bidirectional stability of the L1 and L2. The shift from representation to processing accounts to attrition means that the cognitive psychological theory of Radical Embodied Cognition becomes relevant to invoke. The reasoning underlying this theory is that the way humans process information (including language) is seen as distinctly different from computer information processors, although this analogy is often made. Epstein (2016) claims that computers really do have physical memories, which they store, retrieve, and can replace upon demand. Organisms (humans), however, do not. Humans do not have a store of words or grammatical rules, not even memories. As such, linguistic memories cannot be lost the way that past attrition work has assumed (cf. Chemero, 2013). Instead, changes in language skills are best explained by the direct interaction between what is termed organisms (i.e., people) and the world they live in. In other words, language learning, language unlearning, or even language relearning occurs as a direct result of a speaker's interaction with the environmental input.

The nature of the input especially has received abundant attention in recent years and specifically which type of input or lack thereof drives attrition. As early as 2002, van Hell and Dijkstra posited that foreign language knowledge could influence native language performance even in exclusively native contexts. To shed more light on this, recent investigations have examined short-term L2 learning, both in native language contexts and during fixed-term stays abroad. Chang (2012) has shown, for instance, that substantial – and indeed most – language attrition takes place in the initial phase following a change in L1 and L2 language exposure. The results of his study in which he examined L1 American English speakers becoming immersed in an intensive Korean as a foreign language course (while residing in Korea) showed that the participants underwent phonetic attrition in their L1 as little as six weeks into the experience. This was particularly noticeable in English stop consonants and vowels. Speakers restructured their English vowel space to assimilate to the Korean vowel space. Chang explained this short-term attrition by the need to avoid interference from the L1 in the initial period of coming to terms with a new language, a need that is aided by a strong suppression of the L1. The intriguing hypothesis that can be formulated on the basis of such findings is that good L2 learners are also good L1 forgetters: they flexibly ‘allow’ changes to their L1 in accommodating their L2. Taking this assumption a step further, researchers have examined changes to the L1 in speakers who are enrolled in L2 courses while continuing to reside in an L1 environment (e.g., Bice and Kroll, 2015), very much building on van Hell and Dijkstra's earlier findings (2002).

Such contributions place attrition within a broader framework of activation and inhibition of bilingual language use and heavily build on Anderson, Bjorn and Bjorn's (1994) theory of retrieval induced forgetting (RIF): when one retrieves a piece of information, related knowledge is inhibited in order to avoid interference. In relation to L1 attrition, the hypothesis is that when a word for a certain concept is retrieved in the L2, subsequent retrieval of the corresponding (i.e., translation equivalent) L1 item is impeded. Levy et al. (2007) were the first to apply RIF theories to language attrition; their goal was to test the theory's prediction on the basis of a sample of L1 American-English speaking students taking a semester of Spanish at college. They found that these students became progressively slower in naming an L1 English item after the same item had been named ten times in Spanish. Although the RIF framework offers important insights into the nature of forgetting linguistic information, the timeframe

under which this happens needs to be examined more closely. Linck, Kroll and Sunderman (2009) showed that RIF effects continued to exist in a group of L2 Spanish learners from an L1 American-English background even up to six months following their arrival back in the United States after they had spent an exchange semester in Spain, but how long the effects persist after that remains unknown. Such insights about timeframes are crucial for attrition theorizing.

In short, these outcomes collectively point to the fact that L1 attrition cannot be seen as separate from L2 learning: L1 attrition does not occur without L2 learning permeating the picture. The collective outcomes also indirectly present evidence for the dynamic interplay between several languages in one mind: based on learner internal factors (age, motivation, among others) but crucially also on environmental settings and stimuli – one of a speaker's two languages is more dominant in selection without the need to posit that one of the languages is irretrievably lost. Instead, competition is the main driving force of the dynamic interplay between the L1 and L2 and the environmental stimulus is thus critical. Competition is at play in both L2 influences on the L1 in an exclusively native environment as well as in attrition following a move abroad. Under this view, the speaker-internal language change resulting from both settings then differs only in terms of intensity, but the external factors are of course radically different in both cases. How this exerts different influences that extend beyond behavioural findings can be answered using psycholinguistic and neurolinguistics methods.

5. Main research methods

With the field of attrition becoming more influenced by psycholinguistic investigations of the bidirectional consequences of bilingual encounters, the methods to study the phenomenon too have seen a shift. Early attrition investigations were mostly characterized by case studies, slowly progressing into cross-sectional analyses in the 1990s, where groups of attriters were matched to speakers who had stayed in the home country (cf. Schmid, 2011 for a historical overview and see also the historical overview section in this chapter). In both case studies and group comparisons, behavioural methods were used to tap attrition. Depending on the locus of investigation, different tests were used, ranging from picture naming to verbal fluency (to examine lexical attrition), morphology elicitation tasks following 'wug'-like procedures (cf. Keijzer, 2007 for further details) or grammaticality judgement tasks to tap syntactic attrition. Augmenting these tasks, free speech has been and continues to be one of the main ways of assessing attrition. On the basis of either semi-structured interviews or story/film retelling procedures, spontaneous speech samples are collected that are then analyzed on different grounds: (dis)fluency signalled through pauses, hesitations or false starts, but also lexical diversity, morphological and syntactic complexity as well as errors or code-switching instances. These behavioural methods have been and continue to be of vital importance in understanding the phenomenon of attrition, and task outcomes have been related to participants' answers on extensive language background questionnaires. It needs to be pointed out that behavioural methods including questionnaires are inherently so-called offline tasks: participants are given specific instructions on what to do, which means that by definition a level of consciousness is involved in completing these tasks (Mertins, 2009).

As attrition is now understood to be primarily governed by mental and neural processes, so-called online methods that tap more automatized and largely unconscious processing may be a better way to examine attrition effects. Whereas online methods are often defined as involving reaction time experiments or methods such as eye-tracking, these too often build on conscious decision making. It is important to distinguish such tasks, then, from 'true' online tasks that almost almost simultaneously with language processing, record speakers' neural activation. Examples

include functional magnetic resonance imaging (fMRI) and electroencephalography (EEG), specifically the measure of Event-related potentials (ERPs) recorded as part of EEG designs (cf. Mertins, 2009). An added advantage of using such methods is that the brain generally outpaces behaviour (McLaughlin, Osterhout and Kim, 2004), allowing attrition processes to be tapped in the absence of overt attrition behaviour. Whereas fMRI evidence is still largely absent in the field of attrition (but see Keijzer, 2014 and Rossi, Prystauka and Diaz, 2019, for potential future directions), ERP evidence is gaining territory in attrition research.

The event-related potential (ERP) technique is an electroencephalogram (or EEG) measure that taps ongoing voltage changes in the brain by placing electrodes on the scalp. From the EEG brainwaves, the ERP technique extracts neural responses that are associated with ‘specific sensory, cognitive, and motor events’ (Luck, 2005, p. 4). Because the responses are time-locked to a given stimulus, they can be directly related to it, hence the name event-related potentials. With a temporal resolution of one millisecond or less under optimal conditions, ERPs are recorded almost simultaneously to the actual brain activation. Until now, the main aim in using ERP methods in attrition work has been to explore if online L1 comprehension can ever become non-nativelike. Steinhauer and Kasparian (2019) outline the different ERP components that have been investigated in the context of attrition: Loerts, Stowe and Schmid (2013), looking at grammatical gender in L1 German attriters, found that this group by and large still processed gender violations in a native-like manner, eliciting so-called P600 effects upon encountering an ungrammatical German article choice, but did show a different scalp distribution than a control group in the home country. However, whereas Loerts et al.’s set-up examined grammatical processing, much ERP-based evidence for non-native processing has relied on lexical processing. Kasparian, Vespignani and Steinhauer (2014) found a two-way distinction between *high-proficient* (i.e., showing little behavioural attrition) and *low-proficient* (showing considerable overt attrition) L1 speakers of Italian who were immersed in an L2 English environment: the low-proficient attriters did not uniformly show an expected N400 effect when encountering a lexical violation in their L1. Kasparian et al.’s finding was corroborated by Datta (2010), who tested L1 Bengali speakers immersed in the US. The participants’ ERP signatures showed a more negative deflation to English, but also Bengali words, compared to native speakers of either language. In short, the scant ERP evidence to date shows that – in the lexical domain – L1 attriters can come to process their mother tongue in a non-native like manner. Importantly, these ERP findings corroborate behavioural outcomes: Kasparian, Vespignani and Steinhauer (2014) showed that the ERP signatures of only those speakers who already showed overt attrition effects deviated from those of non-attrited speakers. That also means that the potential of neuroimaging methods to augment behavioural data is not being optimally met yet. More attention ought to be paid to research designs that use the strong suits of past behavioural work on attrition in combination with online measures such as EEG and fMRI. Such a mixed method design has the potential to shed more light on the outstanding questions of attrition, most notably the questions about the (ir)reversible nature of attrition.

6. Future directions

The field of language attrition is at a crossroad. In order to advance the field both conceptually and methodologically, what is needed now are on the one hand epidemiological studies where larger cohorts of L1 attriters are followed over a longer period of time and on the other smaller cross-sectional work that directly compares immersed versus non-immersed (in an L2-environment) speakers on the basis of the same methodological set-up. Longitudinal

designs where groups of attriters are followed over time are rare in attrition work (see the historical overview section in this chapter). Crucially, studies also never looked at the longevity of attrition effects after a return to the native environment. Whereas before such a perspective was not possible, the changed globalization practices with expats often being sent out to international locations with the certainty of returning after five to seven years opens up this research avenue. The L1 vs. L2 dynamic can be uniquely studied as speakers shift from a single (where only their L1 is spoken on a daily basis) to a dual language context (with the L2 being used alongside the L1) and back to a single language setting. Ideally, behavioural evidence emerging from such a design would be augmented by neural activation evidence, most notably changes in ERP signatures to both lexical and grammatical stimuli. Beyond ERP evidence, eyes-closed resting state EEG methodology in language learning contexts is on the rise (cf. Prat et al., 2016). Such work has never been carried out in the contexts of language attrition and currently only exists in relation to L2 acquisition but could greatly elucidate the phenomenon and trajectory of attrition, especially if a design is construed where single to dual (and back to single language) contexts are directly compared in terms of changed neural activation patterns. Indeed, if resting-state neural activation fluctuates as a function of foreign language learning, it may also show oscillations as a function of language unlearning, i.e., attrition, in interaction with L2 learning.

Complementing such work, it would be fruitful to study short vs. long-term attrition effects. Specifically, by employing the same behavioural test battery and neural activation methods (once again focusing on EEG as this has most often been applied to an attrition context) for both individuals immersed in an L2 environment and those learning a foreign language in the native environment, the role of input and exposure can be allowed to freely vary, ultimately shedding light on the unique nature of attrition within the continuum of bilingual processing work. Moving away from the longitudinal perspective that is also needed, this line of work can then cross-sectionally target individuals who vary along three dimensions: (1) immersed versus non-immersed; (2) implicit versus explicit L2 learning, i.e., naturalistically versus classroom-based; and (3) continued L1 exposure as assessed as a continuous variable. Whereas the first is a clear dichotomy, the second and third dimensions by definition form a continuum of magnitude effects. By detailing the everyday linguistic realities of individuals, it can be further explored under which circumstances language learning vs. unlearning occurs. With such new research avenues, more light can be shed on the question of the temporary versus long-term attrition effects as well as the source of attrition. Moreover, by placing L1 attrition in the broader context of bilingual interference and processing, not only can insights about bilingual processing inform attrition but attrition work can also greatly contribute to bilingual theorizing (cf. Schmid and Köpke, 2017). There has never been a more poignant time to explore these new research avenues and arrive at new theories of language attrition.

7. Further reading

Köpke, B. and Keijzer, M. (2019). Neurolinguistic and psycholinguistic approaches to language attrition. In: M. S. Schmid and B. Köpke, eds., *Handbook of language attrition*, 1st ed. Oxford: Oxford University Press, pp. 63–72.

This handbook chapter forms the introduction to a section on psycholinguistic and neurolinguistic approaches to language attrition. As such, it outlines perspectives on language attrition as a processing, as a memory and as a brain mechanism. It presents a brief historical account of how psycholinguistic and neurolinguistic attrition studies have developed and the insights they have provided, but also looks ahead at how the field can progress, also in relation to the viability of different neuroimaging techniques as applied to language attrition.

Keijzer, M. and de Bot, K. (2019). Unlearning and relearning of languages from childhood to later adulthood. In: L. Ortega and A. de Houwer, eds., *Handbook of bilingualism*, 1st ed. Cambridge: Cambridge University Press, pp. 267–286.

This handbook chapter explores in depth and from a theoretical perspective what it means to unlearn a language. Not only does it detail processing versus representational accounts of language forgetting but it also looks at relearning and in that context elaborately discusses the savings paradigm in relation to language attrition.

Schmid, M. S. (2011). *Language attrition*. Cambridge: Cambridge University Press.

This book introduces students and new scholars in the field to the topic of language attrition, most notably first language attrition. It covers various topics within the realms of linguistic and extralinguistic aspects of language attrition, but also contains hands-on sections related to conducting research on language attrition (from preliminary considerations to experimental designs and data analysis).

8. Related topics

Experimental approaches to the study of language contact and its consequences for the bilingual mind; cognitive factors; bilingual language acquisition

Abbreviations

| | |
|------|---|
| AMTB | Attitude and Motivational Test Battery |
| EEG | electroencephalography |
| ERP | event-related potential |
| fMRI | functional magnetic resonance imaging |
| L1 | First language |
| L2 | Second language |
| RIF | retrieval induced forgetting |
| SVQ | Subjective Ethnolinguistic Vitality Questionnaire |

References

- Anderson, M.C., Bjorn, R.A. and Bjorn, E.L. (1994). Remembering can cause forgetting: Retrieval dynamics in long-term memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 20, pp. 1063–1087.
- Bahrck, H.P. (1984). Semantic memory content in permastore: Fifty years of memory for Spanish learned in school. *Journal of Experimental Psychology: General*, 113(1), p. 1.
- Ben-Rafael, M. and Schmid, M.S. (2007). Language attrition and ideology: Two groups of immigrants in Israel. *Language Attrition: Theoretical Perspectives*, 22, pp. 205–226.
- Bice, K. and Kroll, J.F. (2015). Native language change during early stages of second language learning. *Neuroreport*, 26(16), pp. 966–971.
- Bourhis, R.Y., Giles, H. and Rosenthal, R. (1981). Notes on the construction of a ‘subjective vitality questionnaire’ for ethnolinguistic groups. *Journal of Multilingual and Multicultural Development*, 2(2), pp. 145–153.
- Bylund, E. (2009). Maturation constraints and first language attrition. *Language Learning*, 59(3), pp. 687–715.
- Chang, C.B. (2012). Rapid and multi-faceted effects of second language learning on first language speech production. *Journal of Phonetics*, 40, pp. 249–268.
- Chemero, A. (2013). Radical embodied cognitive science. *Review of General Psychology*, 17(2), pp. 145–150.

- Cook, V. (2003). Introduction: The changing L1 in the L2 user's mind. In: V. Cook, ed., *Effects of the second language on the first (Vol. 3)*, 1st ed. Clevedon: Multilingual Matters.
- Datta, H. (2010). *Brain bases for first language lexical attrition in Bengali-English speakers*. PhD. City University of New York.
- de Bot, K. and Clyne, M. (1989). Language reversion revisited. *Studies in Second Language Acquisition*, 11(2), pp. 167–177.
- de Bot, K. and Clyne, M. (1994). A 16-year longitudinal study of language attrition in Dutch immigrants in Australia. *Journal of Multilingual and Multicultural Development*, 15(1), pp. 17–28.
- de Bot, K. and Stoessel, S. (2000). In search of yesterday's words: Reactivating a long forgotten language. *Applied Linguistics*, 21(3), pp. 364–388.
- de Leeuw, E., Schmid, M.S. and Mennen, I. (2010). The effects of contact on native language pronunciation in an L2 migrant setting. *Bilingualism: Language and Cognition*, 13(1), pp. 33–40.
- Ebbinghaus, H. (1885). *Über das Gedächtnis: Untersuchungen zur experimentellen Psychologie*. Leipzig: Duncker and Humblot.
- Epstein, R. (2016, December 5). The empty brain. *AEON*. [Online] Available at: <https://aeon.co/essays/your-brain-does-not-process-information-and-it-is-not-a-computer> [Accessed 2 Nov. 2017].
- Footnick, R. (2007). A hidden language: Recovery of a 'lost' language is triggered by hypnosis. In: B. Köpke, M.S. Schmid, M. Keijzer, and S. Dostert, eds., *Language attrition: Theoretical perspectives*, 1st ed. Amsterdam: John Benjamins, pp. 169–187.
- Gardner, R.C. (2010). *Motivation and second language acquisition: The socio-educational model*. New York: Peter Lang.
- Green, D.W. (2011). Language control in different contexts: The behavioral ecology of bilingual speakers. *Frontiers in Psychology*, 2, p. 103.
- Håkansson, G. (1995). Syntax and morphology in language attrition: A study of five bilingual expatriate Swedes. *International Journal of Applied Linguistics*, 5(2), pp. 153–169.
- Hansen, L., Umeda, Y. and McKinney, M. (2002). Savings in the relearning of a second language: The effects of time and proficiency. *Language Learning*, 52(2), pp. 653–678.
- Higby, E., Lerman, A., Korytkowska, M., Malcolm, T. and Obler, L.K. (2019). Aging as a confound in language attrition research. In: M.S. Schmid and B. Köpke, eds., *Handbook of language attrition*, 1st ed. Oxford: Oxford University Press, pp. 121–135.
- Hopp, H. and Schmid, M.S. (2013). Perceived foreign accent in first language attrition and second language acquisition: The impact of age of acquisition and bilingualism. *Applied Psycholinguistics*, 34(2), pp. 361–394.
- Hutz, M. (2004). Is there a natural process of decay? *First Language Attrition: Interdisciplinary Perspectives on Methodological Issues*, 28, p. 189.
- Kasparian, K., Vespignani, F. and Steinhauer, K. (2014). The case of a non-native-like first language: ERP evidence of first language (L1) attrition in lexical and morphosyntactic processing. *International Journal of Psychophysiology*, 94(2), pp. 159–160.
- Keijzer, M. (2007). *Last in first out? An investigation of the regression hypothesis in Dutch emigrants in Anglophone Canada*. Utrecht: LOT Publications.
- Keijzer, M. (2008). Language attrition in Dutch emigrants in Anglophone Canada: Internally or externally-induced change? *Linguistics in the Netherlands*, 25(1), pp. 97–108.
- Keijzer, M. (2010). The regression hypothesis as a theoretical framework for first language attrition. *Bilingualism: Language and Cognition*, 13(1), pp. 9–18.
- Keijzer, M. (2014). Changes in neural activation patterns and brain anatomy as a function of non-pathological first language attrition. *Journal of Neurological Disorders*, 2(4).
- Keijzer, M. and de Bot, K. (2019). Unlearning and relearning of languages from childhood to later adulthood. In: L. Ortega and A. de Houwer, eds., *Handbook of bilingualism*, 1st ed. Cambridge: Cambridge University Press, pp. 267–286.
- Köpke, B. (2004). Attrition. *Special issue of Journal of Neurolinguistics*, 17(1).
- Köpke, B. (2007). Language attrition at the crossroads of brain, mind, and society. In: B. Köpke, M.S. Schmid, M. Keijzer, and S. Dostert, eds., *Language attrition: Theoretical perspectives*, 1st ed. Amsterdam: John Benjamins, pp. 9–38.

- Köpke, B. and Keijzer, M. (2019). Neurolinguistic and psycholinguistic approaches to language attrition. In: M.S. Schmid and B. Köpke, eds., *Handbook of language attrition*, 1st ed. Oxford: Oxford University Press, pp. 63–72.
- Köpke, B. and Schmid, M.S. (2004). Language attrition: The next phase. In: M.S. Schmid, B. Köpke, M. Keijzer, and L. Weilemar, eds., *First language attrition: Interdisciplinary perspectives on methodological issues*. Amsterdam: John Benjamins, pp. 1–43.
- Lambert, R.D. and Freed, B.F. (1982). *The loss of language skills*. Rowly, MA: Newbury House Publishers.
- Levy, B.J., McVeigh, N.D., Marful, A. and Anderson, M.C. (2007). Inhibiting your native language: The role of retrieval-induced forgetting during second language acquisition. *Psychological Science*, 18(1), pp. 29–34.
- Linck, J.A., Kroll, J.F. and Sunderman, G. (2009). Losing access to the native language while immersed in a second language: Evidence for the role of inhibition in second- language learning. *Psychological Science*, 20(12), pp. 1507–1515.
- Loerts, H., Stowe, L.A. and Schmid, M.S. (2013). Predictability speeds up the re-analysis process: An ERP investigation of gender agreement and cloze probability. *Journal of Neurolinguistics*, 26(5), pp. 561–580.
- Luck, S.J. (2005). *An introduction to the event-related potential technique*. Cambridge, MA: MIT Press.
- McLaughlin, J., Osterhout, L. and Kim, A. (2004). Neural correlates of second-language word learning: Minimal instruction produces rapid change. *Nature Neuroscience*, 7(7), p. 703.
- Mertins, B. (retrieved 2019). *The use of experimental methods in linguistic research: Advantages, problems and possible pitfalls*. Unpublished manuscript. Available at: www.germanistik.tudortmund.de/cms/Medienpool/Daten_Mitarbeiter/Mertins/Publikationen/01-Mertins_-The-Use_-15-33.pdf.
- Murtagh, L. (2003). *Retention and attrition of Irish as a second language: A longitudinal study of general and communicative proficiency in Irish among second level school leavers and the influence of instructional background, language use and attitude/motivation variables*. PhD. University of Groningen.
- Neisser, U. (1984). Interpreting Harry Bahrick’s discovery: What confers immunity against forgetting? *Journal of Experimental Psychology: General*, 113(1), pp. 32–35.
- Pavlenko, A. (2004). L2 influence and L1 attrition in adult bilingualism. *First Language Attrition: Interdisciplinary Perspectives on Methodological Issues*, pp. 47–59.
- Prat, C.S., Yamasaki, B.L., Kluender, R.A. and Stocco, A. (2016). Resting-state qEEG predicts rate of second language learning in adults. *Brain and Language*, 157, pp. 44–50.
- Rossi, E., Prystauka, Y. and Diaz, M. (2019). Investigating L1 attrition and language change: Neuroimaging perspectives. In: M.S. Schmid and B. Köpke, eds., *Handbook of language attrition*, 1st ed. Oxford: Oxford University Press, p. 166–177.
- Scherag, A., Demuth, L., Rösler, F., Neville, H.J. and Röder, B. (2004). The effects of late acquisition of L2 and the consequences of immigration on L1 for semantic and morpho-syntactic language aspects. *Cognition*, 93(3), pp. B97–B108.
- Schmid, M.S. (2011). *Language attrition*. Cambridge: Cambridge University Press.
- Schmid, M.S. (2012). The impact of age and exposure on bilingual development in international adoptees and family migrants: A perspective from Holocaust survivors. *Linguistic Approaches to Bilingualism*, 2(2), pp. 177–208.
- Schmid, M.S. (2013). First language attrition. *Linguistic Approaches to Bilingualism*, 3(1), pp. 94–115.
- Schmid, M.S. and Dusseldorp, E. (2010). Quantitative analyses in a multivariate study of language attrition: The impact of extralinguistic factors. *Second Language Research*, 26(1), pp. 125–160.
- Schmid, M.S. and Keijzer, M. (2009). First language attrition and reversion among older migrants. *International Journal of the Sociology of Language*, 2009(200), pp. 83–101.
- Schmid, M.S. and Köpke, B. (2017). The relevance of first language attrition to theories of bilingual development. *Linguistic Approaches to Bilingualism*, 7(6), pp. 637–667.
- Schmitt, E. (2010). When boundaries are crossed: Evaluating language attrition data from two perspectives. *Bilingualism: Language and Cognition*, 13(1), pp. 63–72.

- Seliger, H.W. (1996). Primary language attrition in the context of bilingualism. In: W.C. Ritchie and T.K. Bhatia. *Handbook of second language acquisition*. San Diego: Academic Press, pp. 605–626.
- Steinhauer, K. and Kasparian, K. (2019). Electrophysiological approaches to L1 attrition. In: M.S. Schmid and B. Köpke, eds., *Handbook of language attrition*, 1st ed. Oxford: Oxford University Press, p. 146–165.
- United Nations. (2013). *International migration report 2013*. [pdf] New York: United Nations. Available at: http://un.org/en/development/desa/population/publications/pdf/migration/migrationreport2013/Full_Document_final.pdf [Accessed 23 Oct. 2018].
- Van Hell, J.G. and Dijkstra, T. (2002). Foreign language knowledge can influence native language performance in exclusively native contexts. *Psychonomic Bulletin and Review*, 9(4), pp. 780–789.
- Ventureyra, V., Pallier, C. and Yoo, H-J. (2004). The loss of first language phonetic perception in adopted Koreans. *Journal of Neurolinguistics*, 17, pp. 79–91.
- Yagmur, K. (1997). *First language attrition among Turkish speakers in Sydney*. Tilburg: Tilburg University Press.
- Zhou, W. (2015). *Assessing birth language memory in young adoptees*. PhD. Nijmegen University.



Taylor & Francis

Taylor & Francis Group
<http://taylorandfrancis.com>

Part 3

Outcomes



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Borrowing

Yaron Matras and Evangelia Adamou

1. Introduction and definitions

‘Borrowing’ is a metaphor that denotes the use of a structure (i.e., phone, phoneme, morpheme, semantic value, or form-function alignment) within a particular linguistic system although it is normally associated with another linguistic system. The term ‘borrowing’ is therefore based on the underlying formal or structuralist assumption that languages are self-contained systems, and that the use of linguistic structures can and should be described in terms of their system affiliation. That assumption has, however, become ever more controversial in recent years, as scholars of contact linguistics refer to users’ ‘repertoires’ (Matras, 2009; Blommaert and Backus, 2013) and stress that bilinguals may interact in a ‘bilingual mode’ where they may choose to combine elements from both languages (Grosjean, 1989). Support for this view comes from neurolinguistic and psycholinguistic studies showing that bilingual speakers do not ‘switch off’ individual linguistic systems (e.g., Thierry and Wu, 2007; Kroll et al., 2008; Loebell and Bock, 2003). Moreover, cross-discipline perspectives in linguistics, ethnography, and education have taken an interest in what is sometimes referred to as ‘metrolingual’ or ‘translanguaging practices,’ understood as users’ ways of availing themselves of the full expressive potential of their entire repertoire of linguistic features in communicative interaction (Li Wei, 2018). As a result, there is a need to distinguish between multilinguals availing themselves of structures that form part of their overall repertoire of linguistic features, and an etymological analysis of individual items (words and structures). Depending on the approach, borrowing can therefore have either a synchronic or diachronic reading and be the object of investigation in the field of bilingualism and language contact, respectively.

Other terms that are used in connection with the crossing of language boundaries are ‘loans,’ ‘transfers,’ ‘copying,’ and ‘replication.’ Matras and Sakel (2007b), in particular, refer to two kinds of replication of a linguistic structure: replication of linguistic ‘matter’ (concrete, identifiable sound-shapes of words and morphs) and replication of mode of organizing units of speech, that is, of linguistic ‘patterns.’ The languages involved in this process are commonly referred to as ‘model’ and ‘replica’ (see Heine and Kuteva, 2005).

2. Historical overview

In the nineteenth century, Whitney (1881) responded to contemporary discussions on the possible existence of ‘mixed languages’ and, in particular, to the question of whether mixture is possible in grammar. He argued that there is nothing in the nature of language that could stand in the way of appropriation of word forms from another language, and introduced the term ‘borrowing’ to refer to such cases (Whitney, 1881, p. 10). Whitney emphasized the unequal role of the two languages in contact as, for example, in the case of English and Norman French, inflection was not borrowed together with verbs and nouns; in other words, grammar was exempted from mixture, and this also applied to pronouns, articles, prepositions, conjunctions, and numerals. Whitney further addressed the challenge to bring empirical proof that shows otherwise and to explain the conditions under which such processes occur (Whitney, 1881, p. 15). It was noted, however, that grammatical material may be borrowed along with words, as in the borrowing of productive word formation affixes with French word forms, something that allows for inflection to take the same route, as in the plural ending in the word *phenomena* (Whitney, 1881, p. 17). These were dubbed ‘secondary processes.’ Moreover, Whitney pointed out the possibility for analogous borrowing as in English *one* from French *on*, and the borrowing of *second*. He further noted the need for a scale of the comparative ease and difficulty of ‘immediate borrowing’ (Whitney, 1881, p. 19). He postulated such a scale in the following way: noun > adjective > verb (where > indicates that a category is more easily borrowed). Whitney (1881) also discussed the typological conditions that could facilitate borrowing, such as conversion possibilities among English word forms that facilitate the borrowing of verbs and adjectives. By contrast, Persian does not borrow Arabic verbs, but instead uses nouns to form quasi-verbs. Finally, Whitney (1881) addressed the replicability of morpho-syntactic patterns such as the order of heads and attributes and suggested that more formal or structural material is more resistant to borrowing.

The first half of the twentieth century is another key moment in the history of contact linguistics, with several authors focusing on borrowing. Haugen (1950), for example, postulated that the analysis of borrowing must begin with an analysis of the behaviour of bilingual speakers. He pointed out the awkwardness of the metaphor ‘borrowing’ even though its use as a technical term has remained largely unambiguous in linguistics. Borrowing is essentially the reproduction by speakers of new linguistic patterns in the context of a language other than the one in which those patterns were acquired. Every borrowing must have at some point appeared as an innovation, whether or not the speaker is conscious of that. Borrowing is thus a process of ‘innovative reproduction’ (Haugen, 1950, p. 212). Haugen distinguishes ‘importation,’ understood as the complete replication of material from another language, from ‘substitution,’ understood as replication that involves alteration of aspects of the structure of the borrowed item. He also reviews various competing and closely related terms such as ‘loanwords,’ which he distinguishes from ‘hybrids’ (where only a part of the word is borrowed), ‘loan translations’ or ‘calques’ (where form-function mapping is replicated), and ‘semantic loans’ (where word semantics are reproduced). In particular, Haugen introduces three categories: ‘loanwords’ (importation but no morphemic substitution), ‘loanblends’ (importation and partial morphemic substitution of certain features), and ‘loanshifts’ (complete morphemic substitution without importation, shift in semantic meaning of an existing word based on similarity with external words), including ‘loan translations’ (words that come into being through contact but are not directly imported from or even modelled on a specific item from the contact language). As a contemporary example of the latter (unknown to Haugen), we can identify today the German expressions *Handy* ‘a mobile phone, cell-phone’ and *Beamer* ‘a projector linked to

a computer,' which replicate English-derived word-forms, but with novel meanings. Haugen also reviews Whitney's scale of borrowing, acknowledging that all features can be borrowed, but that there is a 'scale of adoptability' that is correlated with structural organization, but does not elaborate on that point. He does, however, introduce the need for cross-linguistic research, hinting that there may be differences in borrowing among languages. Moreover, he discusses the effects of borrowing on a given linguistic system, noting that it leads to instability. Finally, he discusses phonological, phonetic, and morphological adaptation. In sum, Haugen (1950) largely sets the research agenda on borrowing by addressing key issues such as the manner of structural integration, the structural impact of borrowing, the existence of a universal scale of borrowability of categories, and cross-linguistic differences in borrowing patterns.

At roughly the same time, Weinreich (1953) highlighted the fact that individuals are the locus of contact, directing attention to the fact that languages are said to be in contact if they are used alternately by the same person. Weinreich focuses on 'interference' among languages, where elements that do not belong to the main language of a speech event are considered to be borrowed or transferred. For Weinreich, interference is the transfer within an act of speech. Much of his discussion is devoted to an overview of types of transfer in phonetics, grammar, and lexicon. He notes that transfer of borrowed morphemes is rare and that the fuller the (structural) integration of a morpheme, the less likely transfer is to occur. According to Weinreich (1953, p. 35), nouns show high borrowability because of their semantic function as designations of new things. Borrowing of lexicon is also conditioned by the frequency of words, where there is a need for affective enrichment or euphemism, a need for differentiation, and a general need for renewal. Weinreich also distinguishes 'replica' from 'transfer' (1953, p. 37ff) which corresponds to interference in the domain of grammatical relations such as word order, or agreement, or functional extension.

In the second part of the twentieth century, Moravcsik (1978) proposed a link between structural autonomy, semantic transparency, and borrowability. This results in the following implicational hierarchies: nouns > verbs, derivational > inflectional, unbound > bound. Heath (1984) distinguishes convergence from transfer of forms. He also raises questions regarding adaptation, resulting from conflicting phonological patterns. He introduces the concept of 'routines' to denote the habitual adaptation pattern of borrowed forms. Together these contributions set the agenda for structural types of borrowing, structural patterns of integration, motivations and societal conditions for borrowing.

3. Critical issues and topics

3.1 *Distinguishing borrowing from code-switching*

A major topic of discussion in the literature is the distinction between borrowing and code-switching insertions. There are two main strands of analysis: some researchers consider that borrowings are distinct from code-switching, others that the distinction is not clear-cut but is best understood as a continuum.

The proposal that code-switching and borrowing are two different phenomena is expressed by Poplack, Sankoff and Miller (1988), Poplack and Dion (2012) and in a different theoretical frame by MacSwan (2016) (also see Chapters 3 and 5). Whereas Poplack, Sankoff and Miller (1988) assess various criteria to define 'borrowing,' Poplack and Dion (2012) suggest that grammatical integration is the best diagnostic criterion and that it correlates with the degree of composition. In this view, single-word tokens tend to be treated like borrowings as they are integrated into the grammatical frame of the recipient language, unlike multiword

code-switching insertions. Building on the Minimalist Program, MacSwan (2016) proposes a lexicalist framework. In this approach, bilinguals have two lexicons that include grammatical aspects, and two phonetic forms that follow two phonological systems. In this perspective, phonological integration is the best diagnostic feature of borrowings, covering lexical borrowings that are widespread in the community as well as ‘nonce borrowings,’ that is, borrowings used only once by a given speaker.

In contrast, researchers such as Myers-Scotton (1993) and Matras (2009) consider the existence of a borrowing–code-switching continuum. For Myers-Scotton (1993), single words are code-switching insertions with the potential of becoming borrowings as they become conventionalized, that is, shared by various members of a speech community. There is, according to that view, no structural distinction between borrowing and code-switching, merely a social distinction between usage conventions. Matras (2009, p. 111) considers a variety of diagnostic features to determine an item as a borrowing or as a code-switching: degree of speaker bilingualism (monolingual vs. bilingual), degree of item composition (utterance vs. single lexeme) and of functionality (stylistic vs. default use), unique character of the referent (lexical vs. paralexical), operability (core vocabulary vs. grammatical operations), regularity of the process (single vs. regular occurrence), and structural integration (non-integrated vs. integrated).

3.2 Integration

For many researchers, the defining characteristic of a borrowing is its integration into the phonology, morphology, syntax of the recipient language and modulation of their semantic scope with respect to the recipient language’s equivalent terms. It is generally considered that if an item is integrated by means of phonology, morphology, and syntax, then it can be considered a borrowing.

Standard phonological models assume that the integration of borrowings is a phonological process that takes place in production (Hyman, 1970). More recently, however, it has been proposed that integration also takes place in perception, at a pre-phonological level (Peperkamp and Dupoux, 2003; Yip, 2006). Peperkamp and Dupoux, in particular, suggest that phonetics most likely drive the integration process based on the first language’s ‘phonetic decoder.’ It is interesting to note that borrowed items may show various degrees of phonological integration. For example, Adamou and Arvaniti (2014) report that Romani-Turkish bilinguals approximate some Turkish vowels that do not belong to the native Romani inventory such as [y], [u] and [œ] together with the Turkish words. On the other hand, Arabic loans in Turkish are often exempted from the rules of (grammatical) vowel harmony, thus the plural formation of *harf* ‘letter’ is *harf-ler*, not **harf-lar* (Al-Hashmi, 2016, p. 32). Which phonetic/phonological phenomena will be adapted to the phonetics and the phonological rules of the recipient language seems to depend on a combination of perceptual and articulatory ease. Moreover, as Matras (2009) notes, other motivations may play a role, as, for example, bilinguals familiar with the pronunciation of a word in the donor language may choose to ‘authenticate’ it by approximating its original phonology.

Structural factors may sometimes drive the choice of the integration strategy, though structural equivalence between two languages is not a pre-condition for borrowing. For example, Domari speakers borrow the full Arabic word-form for all comparative/superlative forms of adjectives, resulting in complete borrowing-based suppletion of the inventory of adjectives: *tilla* ‘big’ native Domari with *akbar* ‘bigger’ from Arabic (Matras, 2009). The motivation behind this choice appears to be that the Arabic morpho-phonological template that forms the comparative/superlative, namely áCCaC, is not easily integrated into the morphological structure of Domari, yet there is at the same time a strong cognitive motivation for the comparative/superlative procedure to undergo fusion with that of the contact language Arabic.

In a cross-linguistic study of verb borrowing integration with examples from over 400 languages Wohlgenuth (2009) identified four major strategies: (1) verb integration through light verbs, (2) indirect insertion, that is, verb integration through affixes, (3) direct insertion without any morphological or syntactic accommodation, and (4) paradigm transfer, for verbs borrowed together with their original verb morphology. Light verbs are commonly used cross-linguistically. For example, Spanish verbs are integrated into the Otomanguean language Ixcatec with the Ixcatec verbs *tse* ‘do’ or *tsu* ‘want.’ The light verbs receive the person and tense-mood-aspect markers and are followed by the Spanish verb stem, e.g., *tse pregunta* ‘ask’ (from Spanish *pregunta-r*), *tse ?adbertí* ‘warn’ (from Spanish *adverti-r*) (Adamou, 2016, p. 90). In comparison, an example of direct insertion of Spanish verbs comes from Imbabura Quechua where Quechua verbal inflection is added directly to the Spanish stems, e.g., *balurani* ‘I value’ (from Spanish *valora-r*) (Gómez-Rendón, 2007). Indirect insertion of loan verbs may take place through affixes. This is common in Romani where inflectional or derivational affixes from Greek, such as the Greek aorist *is-/-as-/-os-* and the Greek present tense *-iz-/-az-/-oz-* and *-in-/-an-/-on-*, have lost their grammatical value and serve as ‘loan verb markers’ (Elšík and Matras, 2006, pp. 324–333). These affixes were originally used with Greek verbs, as in *ir-in-av* or *ir-iz-av* ‘I return’ (from Greek *jir-íz-o*), and for past tense *ir-is-ájlom* ‘I returned’ (from Greek *jír-is-a*). They were then extended to verb borrowings from other contact languages, as in *misl-in-av*, *misl-iz-av* ‘I think’ (from Slavic *misl-*). Finally, paradigm transfer is noted among various Romani dialects spoken in Russia, Finland, and the Balkans. For example, in Romani spoken in Greek Thrace, Turkish verbs consistently take Turkish verb morphology, that is, various tense-mood-aspect markers, but also all person markers, as well as the negation and causative affixes (Adamou, 2016).

To account for preferred verb-integration strategies across the languages of the world, Wohlgenuth (2009) assesses the hypothesis that these could correlate to the degree of contact and bilingualism. Although paradigm transfer is indeed a strategy that requires a high degree of bilingualism, this hypothesis is ultimately ruled out. Matras (2009) proposes instead that the choice of inflectional system in which to anchor the predication is crucial as it symbolizes the bilingual speaker’s context-bound choice of ‘language.’ In that sense, some speakers integrate verbs merely as lexical labels while others use them, to various degrees, as predicate-initiating devices. More specifically, paradigm transfer is possible when there is full acceptance of a bilingual group identity that licenses speakers to initiate the predication in either language. An alternative account is offered by Myers-Scotton and Jake (2014) who hypothesize that verb integration into the recipient language’s morphology should be preferred as being less costly, since the bilingual would only need to control the semantic-pragmatic features and not to check the congruence between the two languages. However, psycholinguistic evidence from Turkish-Romani bilinguals who make use of the paradigm transfer strategy shows that processing morphologically non-integrated verbs in comprehension does not incur cost as long as this has become a habitual verb-integration strategy (Adamou and Shen, 2019).

3.3 Borrowability

‘Borrowability’ is a term that refers to the susceptibility of categories to borrowing. It is often expressed through borrowability scales and implicational hierarchies. Borrowability scales are primarily associated with the frequency of borrowed items by category in a cross-linguistic perspective. Implicational hierarchies, by contrast, suggest that borrowing follows a specific development, where some stages precede other stages in a non-trivial manner.

Thomason and Kaufman (1988) were the first to propose a link between structural categories of borrowing and the duration and intensity of cultural contact ranging from ‘casual contact’ and lexical borrowing, to ‘very strong cultural pressure’ and heavy structural borrowing (Thomason and Kaufman, 1988, pp. 74–75). While very influential, the proposal suffers from a number of shortcomings: the categories on the scale remain somewhat vague (e.g., ‘function words’) and no criteria are provided to quantify the degrees of intensity of contact (e.g., ‘slightly more intense contact’). More importantly, no explanatory account is offered for the link between category susceptibility to borrowing and intensity of cultural contact.

Campbell (1993) offered an evaluation of various proposed generalizations in the contact linguistics literature. While these did not address general patterns of borrowing, they have contributed to the debate on the status of counter-examples. Campbell criticizes the notion of structural compatibility as a prerequisite for borrowing. He also criticizes the notion of gaps as a motivating factor for borrowing, that is, the idea that the borrowing of a morpheme necessarily leads to the replacement of the corresponding inherited morpheme rather than overall system enrichment. He further criticizes the claim that free standing morphemes are more borrowable than bound morphemes, citing individual examples of borrowed bound morphemes (including the mixed language Ma’a, based on Thomason and Kaufman, 1988). Campbell (1993) also critiques Moravcsik’s stated universals (Moravcsik, 1978, p. 101ff) noting that the primacy of lexical borrowing as a prerequisite for non-lexical borrowing may be true empirically, but is not theoretically significant since there is no link between the two. With support from some empirical evidence, he also rejects the generalization that verbs cannot be borrowed, and argues that despite being common, the borrowing of derivational morphology is not an absolute prerequisite for the borrowing of inflectional morphology, noting that borrowing of plural markers such as Spanish *-s* is found in Quechua without any borrowing of derivational markers. Instead Campbell slightly reformulates this implicational hierarchy, saying that it is unlikely for semantically weak affixes to be borrowed (Campbell, 1993, p. 103). Finally, he offers counter-examples to the notion that grammatical items can only be borrowed along with their linear rule as, for instance, Arabic prepositions are borrowed as postpositions in Turkish (e.g., *kadar*).

Evidence from two cross-linguistic samples have significantly enriched this discussion. The first sample consists of data from 27 languages collected through responses to a typological questionnaire that was filled in by language specialists (Matras and Sakel, 2007a). The second sample comes from the Romani Morphosyntax Database bringing together responses to a questionnaire from over 75 Romani dialects in contact with 25 different languages (Elšik and Matras, 2006). The analysis of these samples confirms a connection between susceptibility to borrowing, and the truth- or presupposition value assigned by a category to propositional content. A good example is the borrowability hierarchy of connectives ‘but > or > and,’ as well as the hierarchy ‘modality > aspect > tense,’ ‘indefinites > other pronouns,’ ‘prosody > segmental phonology,’ as well as the tendency for more complex or discontinuous local relations such as ‘against’ and ‘except’ to show higher borrowability than basic local relations such as ‘on’ or ‘at.’

3.4 Motivations for borrowing

Motivations for borrowing can be defined as the factors that drive the speaker to use linguistic matter from one language in interaction settings that are generally associated with another language.

Traditionally, motivations for borrowing were conceived of as falling under two types: filling a ‘gap’ (typically including cultural loans that are associated with novel activities and artefacts) and reflecting ‘prestige’ (covering the use of items from the language of a more powerful community to achieve a wide range of social outcomes). However, recent research focuses instead on a wider variety of factors that can be summarized as follows: the need to provide labels for unique referents; association with specific activity domains; the cognitive pressure resulting from the manipulation of the presupposition domain or management of the interaction roles, and, in particular, the clash between speaker intentions and hearer expectations (Matras, 2009).

Van Hout and Muysken (1994) investigated the use of Spanish borrowings in Bolivian Quechua and found that Spanish items that were discourse-related were more susceptible to borrowing than items related to the structure of the clause; that speakers tended to borrow fewer items that are highly inflected; and that frequently used words in Spanish are more likely to be used in Quechua. Stolz and Stolz (1996) examined borrowing in roughly 30 Mesoamerican languages in contact with Spanish and noted the systematic borrowing of function words. They concluded that structural criteria or gaps are not relevant explanatory factors, but that the pragmatics of communication offer the best explanatory account. Salmons (1990) further elaborated on the pragmatic motivations of borrowing, later expanded by Matras (1998, 2009) to relate borrowings to the communicative interaction setting and to the effects that the choice of linguistic items from a given language is expected to have on the interlocutor. From this perspective, gaps are not seen as deficiencies of a given language’s expressive means. Rather, borrowings offer bilinguals the possibility to make use of their full inventory of linguistic resources by lifting the restrictions in language selection and using linguistic items in a wider variety of interactional settings. Matras (1998) more specifically accounts for this preference by referring to the role of cognitive factors, that is, to the effort that is needed to keep the subcomponents of a bilingual’s language repertoire apart. Evidence from cross-linguistic sampling of grammatical borrowing indicates that grammatical markers associated with complex language processing may be more prone to borrowing (as in the case of the adversative marker ‘but’). The explanation offered in Matras (1998) is that the interaction-level tension surrounding the act of contradicting a shared presupposition may interfere with the language selection mechanism or ‘executive control.’

Another influential account that takes into consideration cognitive factors comes from the 4-M model building on Levelt’s speech processing model (see Myers-Scotton and Jake, 2017 and Chapter 4). According to the 4-M model, the conceptually activated morphemes (the ‘early system morphemes’) are more susceptible to borrowing as they are salient in the mental lexicon along with their content morpheme heads; this is the case of determiners, derivational prepositions, and particles in phrasal verbs, as well as certain affixes (derivational and plural markers, some tense and aspect markers, some subordinating and coordinating conjunctions). In contrast, the structurally assigned morphemes (the ‘late system morphemes’) are rarely borrowed because they carry little content and are accessed late in speech production; this is the case of ‘bridges’ that join together two NPs and complementizers that join together two clauses, and ‘outsiders’ that include the agreement markers and several case markers.

3.5 Borrowing and language genealogy

Glottochronology is a method that was elaborated in the 1950s to evaluate language genealogies. This method estimated that every 14 non-cognate pairs out of hundred-word lists corresponded to a thousand-year separation between languages (Swadesh, 1955). The word

lists elaborated by Swadesh for this purpose were very influential in the discussion of borrowings but were also heavily criticized. Swadesh's word lists rely on the premises of the 'gap' hypothesis and on the assumption that some concepts are more likely to remain unaffected by language contact (e.g., body parts, close kin, body-related activities, pronouns, interrogatives, and basic concepts for nature and geography). The limits of this method to evaluate language genealogical groupings and separations have long been identified and alternative methods are currently in use such as Bayesian linguistic phylogenetic dating and ASJP chronology (see Holman et al., 2011). In addition, researchers working in current approaches no longer wish to exclude borrowings, but instead embrace the idea that language contact is ubiquitous and that it needs to be quantified and modelled (Holman et al., 2007). Following advances in contact linguistics, it has now become apparent that diverse contact settings will impact differently on language genealogies. For example, settings with patrilocal exogamy (where women move into their husband's group) can be expected to introduce structural and phonological/phonetic features as women are predominantly second language speakers of the husband's language. By contrast, trade is more likely to be associated to the introduction of lexical borrowings associated to artefacts.

4. Current contributions and research

4.1 *The markedness hypothesis*

Elšík and Matras (2006) assessed the role of the Markedness Hypothesis in borrowing. According to the Markedness Hypothesis categories that show greater complexity and lower differentiation are considered as 'marked,' and those that show lower complexity but greater differentiation are considered 'unmarked.' Elšík and Matras (2006) analyzed data from the Romani Morphosyntax Database and showed that markedness and borrowing do not correlate in a uniform way. In borrowing, replication of routine forms, which are more frequent in usage, are associated with 'unmarked' categories, accounting for the more frequent borrowing of categories such as lower numerals or nominative case, whereas reduction of processing load is associated with 'marked' categories, explaining the high borrowability of categories such as peripheral local relations, or markers of contrast, restriction or discontinuity.

4.2 *The specificity hypothesis*

Based on the analysis of a Dutch-Turkish corpus, Backus (2001) formulated the Specificity Hypothesis according to which words with specific meaning are more susceptible to borrowing than words with general meaning. Serigos (2017) tested the Specificity Hypothesis by examining English borrowings in a 24-million-word newspaper corpus of Argentine Spanish. She assessed specificity by examining the target word's environment and by considering that more specific words would have less variation in their surrounding context. The analysis confirmed the Specificity Hypothesis as it revealed that English borrowings were found in contexts that were similar and showed less variation in comparison to their native counterparts. Serigos further interpreted this finding as indirect evidence for the motivation of speakers to add nuances to the original concept.

Verschik and Backus (2012) also discuss specificity with respect to borrowing of bound morphemes. They argue that structural boundedness itself plays just a minor role in explaining constraints on borrowing, and suggest instead that bound morphemes are less likely to be borrowed because their meaning is non-specific. For example, in dialectal Russian spoken in

Estonia, the bound Estonian intensifier marker *-gi* ‘indeed’ is borrowable on account to its function.

4.3 *The Utilitarian motivation*

Based on an analysis of a cross-linguistic sample, Matras (2007) proposed the Utilitarian motivation for borrowing. According to the Utilitarian motivation, more borrowable elements are associated more strongly with interaction settings in the donor language. This is also reflected in Matras’s (2007) proposal that ‘unique referents’ such as terms for institutions, procedures, and concepts related to activities carried out in particular settings are more closely associated with specific languages and therefore more susceptible to borrowing, as opposed to general/core vocabulary. Similarly, numerals in formal contexts, higher cardinal numerals, and days of week are found to be more susceptible to borrowing than numerals in informal contexts, lower cardinal numerals, and times of day. The domain-related hierarchy of numeral borrowing, for example, is confirmed in a study of the Tetun Dili language of East Timor in contact with Indonesian and Portuguese. Tetun speakers use three parallel numeral systems that reflect the interaction domains with which the respective languages are associated: alongside the Tetun numeral set, which is preferred when counting family members, there is wholesale borrowing of the sets of numerals from Indonesian, in connection with education, prices, and technical features, and from Portuguese, for dates and maths (Williams-Van and Hajek, 2018).

Support for a domain-related hierarchy also comes from the results of the typological project *Loanwords in the World’s Languages* (Haspelmath and Tadmor, 2009). This database offers comparative data from 41 languages collected with a questionnaire that was filled by language specialists. The analysis of the sample shows that semantic fields related to social organization such as the modern world, possession or commerce, religion, law, and household, are more susceptible to borrowing than semantic fields related to natural or physical surroundings such as physical world, body, emotions, perception, space, and kinship. Matras (2020) interprets these preferences as evidence for the tendency of bilinguals to generalize the use of some items in all contexts of use and interlocutors. Conversely, it is less likely for speakers to borrow items that are preferred in private settings where it is easier to maintain the demarcation between sub-components of the repertoire. Support for this view comes from the corpus study conducted by Chesley and Baayen (2010). The authors investigated borrowings in French in two corpora within a ten-year time difference and found that borrowings were best preserved when they were shorter and more polysemous, and when they were used in a wide variety of contexts, including contexts that are not culturally specific.

4.4 *Usage frequency*

The usage frequency criterion has been widely discussed both in regard to the recipient language and the donor language. According to the usage-based account, the more frequent an item is in a given language, the more entrenched it becomes and the easier its activation is for a bilingual speaker.

According to Matras (2007), higher frequency, routine, and casualness of usage of words from one language discourages borrowing. This is the case of close kin and core local relations. At the same time, some studies provide compelling evidence for the role of frequency of the borrowed term in the donor language. Interestingly, frequency of an item in the donor language seems to affect the item’s structural integration. For example, Backus (2003, p. 104) analyzed a bilingual Turkish-Dutch spoken corpus from the Netherlands and noted that at

least half of the Dutch compound nouns used in Turkish sentences were high-frequency Dutch items. Backus postulates the Unit Hypothesis to account for the fact that Dutch compound nouns are integrated as such in Turkish. Similarly, based on the analysis of a bilingual speech corpus and statistical frequencies in monolingual speech, Hakimov (2016) demonstrates that usage frequency in the donor language reliably predicts the insertion of inflected words in bilingual sentences. In particular, Hakimov found that nouns are more likely to retain plural inflection when they are more frequently used in plural (as opposed to singular form) in monolingual speech.

The impact of frequency of use on bilingual processing has been discussed in the study of Turkish verbs in a Romani variety spoken in Greece. First, in a corpus study of Turkish-Romani bilingual speech it was shown that verbs from Turkish are not integrated into Romani morphology but instead keep morphological marking intact (Adamou and Granqvist, 2015). The question in this case was whether these verbs were borrowings or code-switching insertions given that speakers were bilinguals. To address this question, Adamou and Shen (2019) relied on additional behavioural data. Analysis of reaction times in sentence processing showed that Turkish-Romani bilingual speakers living in Greek Thrace process Turkish verbs differently depending on their likelihood to be used in Turkish or in Romani within their speech community. More specifically, well-established Turkish verbs were processed without cognitive cost, as in unilingual speech processing (speech in a single language), while Turkish verbs that retain a Romani counter-part were processed with cognitive costs as words that are typically associated to alternation between two languages. Adamou and Shen concluded that the data support usage-based approaches where frequency and interactional habits play a decisive role in processing.

4.5 Borrowability of morphology

Since Moravcsik (1978), it has been widely accepted that inflectional morphology is less borrowable than derivational morphology, and that within inflectional morphology, plural markers are borrowed more frequently than case markers or person agreement markers. To account for these tendencies, it has been noted that inflectional morphemes are more abstract, syntactically dependent on other elements (e.g., in agreement), and have a deictic or anaphoric function (e.g., in agreement) (see Matras, 2007; Gardani, 2008). Matras (2009) proposes the following hierarchy based on semantic transparency (also see Field, 2002 for the notion of transparency): derivation marker > classifier > plural marker > definiteness marker > case marker. He argues that derivational procedures that mark agentivity, diminutives, and noun classifiers are highly transparent and are therefore susceptible to borrowing. Plurality is relatively transparent as it serves to categorize a noun in terms of quantity. Less transparent are the definite and indefinite articles that indicate a noun's pragmatic status in discourse. Finally, case markers are the least transparent as they rely on the semantic-syntactic role of nouns in a phrase.

The aforementioned tendencies were largely confirmed in a recent cross-linguistic study dedicated to bound morphological affixes that draws on data from roughly 100 language pairs (Seifart, 2017). For instance, this overview confirms the previously noted prominence of nominal plurality markers among borrowed affixes. The special treatment of plurals as part of the lexical stem rather than as an inflectional ending in borrowed words is also apparent in Jerusalem Domari where speakers add the Domari inherited plural formation to Arabic-derived plural nouns: singular *zálame* 'man' (Arabic singular *zálame*), plural *zlām-é* (Arabic plural *zlām*) (Matras, 2009). Seifart (2017) also notes that verbal inflection is less borrowable than

nominal inflection. A rare exception to this tendency comes from speakers of Éven (Tungusic) who borrowed subject agreement markers together with mood suffixes from Sakha (Yakut) (Pakendorf, 2019). Pakendorf considers that this rare outcome results from direct borrowing of the inflectional paradigms and notes three facilitating factors: the fact that both languages are agglutinative, that Sakha verb stems are easily segmented, and that the tense-aspect-mood markers need to combine with subject agreement suffixes. Indeed, the third factor might be seen as motivated by the drive toward what Matras (2009) calls ‘fusion’ of, in this case, operational markers of modality.

4.6 Social motivations for borrowing

Two recent studies based on the analysis of large spoken corpora investigated the role of social motivations for borrowing. In the first study, Zenner, Speelman and Geeraerts (2015) analyzed a 35-hour spoken corpus from a reality TV show from 46 speakers of Dutch to identify the use of English insertions. They found that English insertions into Dutch are preferred by younger, more educated males allowing them to highlight aspects of their identities. They also showed that, beyond the classic sociolinguistic variables such as age and gender, situational factors were crucial, as for example English swearwords were particularly frequent in contexts charged with negative emotions.

The second study demonstrates that borrowing does not always come from a socially dominant language. This is illustrated in Calude, Miller and Pagel (2017), a study based on the analysis of a million-word spoken corpus of New Zealand English. The study confirms the high frequency of Māori borrowings despite the fact that Māori is a minority language. Rather, Māori has been going through a significant revitalization process since the 1970s that seems to be affecting its status in positive terms. More specifically, the authors found that social factors such as the ethnic identity of the speakers and of the audience accounted for the extent of Māori borrowing, where higher use of Māori borrowings signalled some sort of support for the Māori identity. Linguistic factors were further identified in the uses of both Māori and White/European New Zealand speakers of English: Māori borrowings were more likely to be used when they had a cultural rather than a core meaning, when they did not have a clear English equivalent, and when they were content words rather than function words.

Finally, Adamou et al. (2016) compared spoken corpora from bilingual speakers of four Slavic minority languages from settings of century-long contact with the majority languages (specifically Greek, Italian, and German). Statistical analyses revealed patterns of borrowing within each community. An additional analysis including extralinguistic factors showed that the degree of lexical borrowing did not depend on the degree of bilingualism in the community or the duration of contact, but was rather conditioned by prescriptive attitudes supported by institutions.

5. Main research methods

Early theoretical generalizations about borrowing relied on the study of language contact phenomena in individual language pairs. Dawkins (1916) for example is an important case study that drew attention to the extensive influence of Turkish on Anatolian Greek, including grammatical inflections, word order, and relative clause constructions. In addition to such qualitative studies, quantitative empirical evidence has become decisive to the development of theoretical claims in language contact. These come from structured cross-linguistic samples and corpus-based studies.

Since the 2000s, corpus-based approaches to language contact and the study of borrowing have become more central due to the growth of computerized corpora. Contact linguists work with written and spoken corpora: the former offer the possibility to explore large corpora, the latter offer direct access to the spontaneous productions of bilingual speakers and to the phonetic details that may be crucial in determining the status of an item as a borrowing. For well-described language pairs and written corpora, automatic language tag models are available using n-gram methods (calculating for example probabilities of character sequences to belong to a given language) and look-up methods (checking the languages' dictionaries) (see e.g., Guzman et al., 2016). In contrast, in the case of lesser-studied languages that lack tools for automatic annotation, manual coding is required. Along with phonetic and morpheme annotation, researchers can tag the words depending on the languages involved: for example, L1, L2, multiple or ambiguous, mixed, other language, or unknown (Adamou, 2016). In the analysis of borrowing, in particular, some corpus linguists adopt a frequency-based approach and others a concept-based approach. In the former, statistical models are used with raw frequencies of borrowed items, whereas in the latter, borrowing use is relative to the use of a specific concept. Though the concept-based approach captures an important aspect of lexical borrowing, it restricts the study to a small set of words that have a more or less clear-cut semantic equivalent.

As discussed in this chapter, cross-linguistic samples have also been created for the study of borrowing. However, the study of borrowing in this perspective presents special challenges (see Matras and Sakel, 2007a). In addition to the need for the sample to reflect the diversity of languages and language families or areas, a comparative study of borrowing must take into account all the categories that are potentially affected by contact. Since the study of borrowing implies a diachronic perspective, information is required about the history of contact of the languages under study. However, reliable diachronic information is lacking with respect to the languages of entire regions. Moreover, as borrowing is sensitive to extralinguistic factors, these too need to be taken into account (e.g., duration and intensity of cultural contact, the roles and status of the participating languages, language attitudes, and the degree of institutional support enjoyed by the languages). Finally, information on contact phenomena is not always available in descriptive grammars and discussions of specific contact phenomena do not cover the entire set of contact phenomena present in a given language. Despite such difficulties, extant cross-linguistic samples of borrowing demonstrate that it is possible to bring together language specialists and work with structured questionnaires that make systematic data comparison possible.

6. Future directions

Some researchers consider translanguaging to be a useful lens through which to enhance our theoretical understanding of borrowing. However, as MacSwan (2017) notes, proponents of the translanguaging analyses need to offer a coherent account for the abundant empirical evidence in the literature and propose a solid frame that can explain how mental lexicons and grammars are organized in a bilingual mind. Usage-based approaches seem to offer such a coherent framework as they combine sociolinguistic, psycholinguistic, and linguistic aspects to account for both diachronic and synchronic processes in language contact (see Chapter 6).

The significant progress of electrophysiological and neuroimaging methods in the past decades can also be expected to impact on the study of borrowing as it has done in bilingual language processing. Behavioural methods have already been offering novel insights into the understanding of bilingualism and language contact phenomena (see Chapter 2). In connection

with the code-switching/borrowing discussion, for example, the behavioural data in Adamou and Shen (2019) open the possibility to consider borrowings as those items that would entail no processing costs as they are expected to be used in a specific language, whereas code-switching insertions would entail processing costs due to a higher surprise effect stemming from the choice of the language.

Lastly, Natural Language Processing methods are currently being applied to the development of new automated techniques, adapted to the specificities of bi/multilingual corpora (see Guzman et al., 2016 and Chapter 1). The analysis of more sizeable bi/multilingual corpora, representing a wider variety of contact settings and language pairs could offer a much-needed empirical basis on which to test extant theoretical models.

7. Further reading

Matras, Y. (2020). *Language contact*. 2nd ed. Cambridge: Cambridge University Press.

This book offers an up-to-date introduction to the highly dynamic field of contact linguistics. It covers a wide-range of language contact aspects, ranging from bilingual child acquisition to typology.

Poplack, S. (2018). *Borrowing: Loanwords in the speech community and in the grammar*. Oxford: Oxford University Press.

This book is dedicated to the study of borrowing. Using the variationist approach, it relies on empirical data from a variety of language pairs. The book stresses in particular the key role of integration in the process of borrowing.

8. Related topics

Code-switching, mixed languages, the 4-M model, usage-based approaches, variationist methods, processing multilingual data

9. References

- Adamou, E. (2016). *A corpus-driven approach to language contact: Endangered languages in a comparative perspective*. Boston and Berlin: De Gruyter Mouton.
- Adamou, E. and Arvaniti, A. (2014). Greek Thracian Xoraxane Romani. *Journal of the International Phonetic Association*, 44, pp. 223–231.
- Adamou, E., Breu, W., Scholze, L. and Shen, X.R. (2016). Borrowing and contact intensity: A corpus-driven approach from four Slavic minority languages. *Journal of Language Contact*, 9(3), pp. 515–544.
- Adamou, E. and Granqvist, K. (2015). Unevenly mixed Romani languages. *International Journal of Bilingualism*, 19, pp. 525–547.
- Adamou, E. and Shen, X.R. (2019). There are no language switching costs when codeswitching is frequent. *International Journal of Bilingualism*, 23, pp. 53–70.
- Al-Hashmi, S. (2016). *The phonetics and phonology of Arabic loanwords in Turkish: Residual effects of gutturals*. PhD. University of York.
- Backus, A. (2001). The role of semantic specificity in insertional codeswitching: Evidence from Dutch-Turkish. In: R. Jacobson, ed., *Codeswitching worldwide II*. Berlin and New York: De Gruyter Mouton, pp. 125–154.
- Backus, A. (2003). Units in code switching: Evidence for multimorphemic elements in the lexicon. *Linguistics*, 41, pp. 83–132.
- Blommaert, J. and Backus, A. (2013). Superdiverse repertoires and the individual. *Multilingualism and Multimodality*, pp. 11–32.
- Calude, A.S., Miller, S. and Pagel, M. (2017). Modelling loanword success – a sociolinguistic quantitative study of Māori loanwords in New Zealand English. *Corpus Linguistics and Linguistic Theory*, Online access only: <https://doi.org/10.1515/cllt-2017-0010>

- Campbell, L. (1993). On proposed universals of grammatical borrowing. In: H. Aertsen and R. Jeffers, eds., *Historical linguistics 1989: Papers from the 9th international conference on historical linguistics*. Amsterdam: John Benjamins, pp. 91–109.
- Chesley, P. and Baayen, H.R. (2010). Predicting new words from newer words: Lexical borrowings in French. *Linguistics*, 48, pp. 1343–1374.
- Dawkins, R.M. (1916). *Modern Greek in Asia Minor: A study of dialect of Silly, Cappadocia and Pharsa*. Cambridge: Cambridge University Press.
- Elšík, V. and Matras, Y. (2006). *Markedness and language change: The Romani sample*. Berlin: De Gruyter Mouton.
- Field, F.W. (2002). *Linguistic borrowing in bilingual contexts*. Amsterdam: John Benjamins.
- Gardani, F. (2008). *Borrowing of inflectional morphemes*. Frankfurt: Peter Lang.
- Gómez-Rendón, J. (2007). Grammatical borrowing in Imbabura Quechua (Ecuador). In: Y. Matras and J. Sakel, eds., *Grammatical borrowing in cross-linguistic perspective*. Berlin: Walter de Gruyter, pp. 523–550.
- Grosjean, F. (1989). Neurolinguists, beware! The bilingual is not two monolinguals in one person. *Brain and Language*, 36, pp. 3–15.
- Guzman, G.A., Serigos, J., Bullock, B.E. and Toribio, A.J. (2016). Simple tools for exploring variation in code-switching for linguists. In: M. Diab, P. Fung, M. Ghoneim, J. Hirschberg, and T. Solorío, eds., *Proceedings of the second workshop on computational approaches to code switching*. Austin, TX: ACL, pp. 12–20.
- Hakimov, N. (2016). Explaining variation in plural marking of single German noun insertions in Russian sentences. In: H. Beherens and S. Pfänder, eds., *Experience counts: Frequency effects in language*. Berlin: de Gruyter, pp. 21–260.
- Haspelmath, M. and Tadmor, U., eds. (2009). *Loanwords in the world's languages. A comparative handbook*. Berlin: De Gruyter Mouton.
- Haugen, E. (1950). The analysis of linguistic borrowing. *Language*, 26, pp. 210–231.
- Heath, J. (1984). Language contact and language change. *Annual Review of Anthropology*, 13, pp. 367–384.
- Heine, B. and Kuteva, T. (2005). *Language contact and grammatical change*. Cambridge: Cambridge University Press.
- Holman, E.W., Brown, C.H., Wichmann, S., Müller, A., Velupillai, V., Hammarström, H., Sauppe, S., Jung, H., Bakker, D., Brown, P., Belyaev, O., Urban, M., Mailhammer, R., List, J.-M. and Egorov, D. (2011). Automated dating of the world's language families based on lexical similarity. *Current Anthropology*, 52, pp. 841–875.
- Holman, E.W., Schulze, C., Stauffer, D. and Wichmann, S. (2007). On the relation between structural diversity and geographical distance among languages: Observations and computer simulations. *Linguistic Typology*, 11, pp. 393–421.
- Hyman, L.M. (1970). The role of borrowing in the justification of phonological grammars. *Studies in African Linguistics*, 1, pp. 1–48.
- Kroll, J.F., Bobb, S.C., Misra, M. and Guo, T. (2008). Language selection in bilingual speech: Evidence for inhibitory processes. *Acta Psychologica*, 128, pp. 416–430.
- Li Wei. (2018). Translanguaging as a practical theory of language. *Applied Linguistics*, 39, pp. 261–261.
- Loebell, H. and Bock, K. (2003). Structural priming across languages. *Linguistics*, 41, pp. 791–824.
- MacSwan, J. (2016). Code-switching in adulthood. In: E. Nicoladis and S. Montanari, eds., *Bilingualism across the lifespan: Factors moderating language proficiency*, 1st ed. Washington, DC: American Psychological Association, pp. 183–200.
- MacSwan, J. (2017). A multilingual perspective on translanguaging. *American Educational Research Journal*, 54, pp. 167–201.
- Matras, Y. (1998). Utterance modifiers and universals of grammatical borrowing. *Linguistics*, 36, pp. 281–331.
- Matras, Y. (2007). The borrowability of grammatical categories. In: Y. Matras and J. Sakel, eds., *Grammatical borrowing in cross-linguistic perspective*. Berlin: Walter de Gruyter, pp. 31–74.
- Matras, Y. (2009). *Language contact*. 1st ed. Cambridge: Cambridge University Press.
- Matras, Y. (2020). *Language contact*. 2nd ed. Cambridge: Cambridge University Press.

- Matras, Y. and Sakel, J. eds. (2007a). *Grammatical borrowing in cross-linguistic perspective*. Berlin: Walter de Gruyter.
- Matras, Y. and Sakel, J. (2007b). Investigating the mechanisms of pattern replication in language convergence. *Studies in Language*, 31, pp. 829–865.
- Moravcsik, E. (1978). Universals of language contact. In: J.H. Greenberg, ed., *Universals of human language*. Stanford, CA: Stanford University Press, pp. 94–122.
- Myers-Scotton, C. (1993). *Social motivations for code-switching. Evidence from Africa*. Oxford: Oxford University Press.
- Myers-Scotton, C. and Jake, J.L. (2014). Nonfinite verbs and negotiating bilingualism in codeswitching: Implications for a language production model. *Bilingualism: Language and Cognition*, 17, pp. 511–525.
- Myers-Scotton, C. and Jake, J.L. (2017). Revisiting the 4-M model: Codeswitching and morpheme election at the abstract level. *International Journal of Bilingualism*, 21, pp. 340–366.
- Pakendorf, B. (2019). Direct copying of inflectional paradigms: Evidence from Lamunghin Even. *Language*, 95, pp. e364–e380. <https://doi.org/10.1353/lan.2019.0063>
- Peperkamp, S. and Dupoux, E. (2003). Reinterpreting loanword adaptations: The role of perception. In: M.J. Solé, D. Recasens, and J. Romero, eds., *Proceedings of the 15th international congress of phonetic sciences*. Barcelona: Causal Productions, pp. 367–370.
- Poplack, S. and Dion, N. (2012). Myths and facts about loanword development. *Language Variation and Change*, 24, pp. 279–315.
- Poplack, S., Sankoff, D. and Miller, C. (1988). The social correlates and linguistic processes of lexical borrowing and assimilation. *Linguistics*, 26, pp. 47–104.
- Salmons, J. (1990). Bilingual discourse marking: Code switching, borrowing, and convergence in some German-American dialects. *Linguistics*, 28, pp. 453–480.
- Seifart, F. (2017). Patterns of affix borrowing in a sample of 100 languages. *Journal of Historical Linguistics*, 7, pp. 389–431.
- Serigos, J. (2017). Using distributional semantics in loanword research: A concept-based approach to quantifying semantic specificity of Anglicisms in Spanish. *International Journal of Bilingualism*, 21, pp. 521–540.
- Stolz, C. and Stolz, T. (1996). Funktionswortentlehnung in Mesoamerika Spanisch-amerindischer Sprachkontakt (Hispanoindiana II). *STUF – Language Typology and Universals*, 49, pp. 86–123.
- Swadesh, M. (1955). Towards greater accuracy in lexicostatistic dating. *International Journal of American Linguistics*, 21, pp. 121–137.
- Thierry, G. and Wu, Y.J. (2007). Brain potentials reveal unconscious translation during foreign-language comprehension. *Proceedings of the National Academy of Sciences*, 104, pp. 12530–12535.
- Thomason, S.G. and Kaufman, T. (1988). *Language contact, creolization and genetic linguistics*. Berkeley: University of California Press.
- van Hout, R. and Muysken, P. (1994). Modeling lexical borrowability. *Language Variation and Change*, 6, pp. 39–62.
- Verschik, A. and Backus, A. (2012). Copiability of (bound) morphology. In: L. Johanson and M. Roberts, eds., *Copies versus cognates in bound morphology*. Leiden: Brill, pp. 123–149.
- Weinreich, U. (1953). *Languages in contact*. The Hague: Mouton.
- Whitney, W.D. (1881). On mixture in language. *Transactions of the American Philological Association (1869–1896)*, 12, pp. 1–26.
- Williams-Van Klinken, C. and Hajek, J. (2018). Mixing numeral systems in Timor-Leste. *NUSA: Linguistic Studies of Languages in and Around Indonesia*, 64, pp. 65–94.
- Wohlgemuth, J. (2009). *A typology of verbal borrowings*. Berlin: De Gruyter Mouton.
- Yip, M. (2006). The symbiosis between perception and grammar in loanword phonology. *Lingua*, 116, pp. 950–975.
- Zenner, E., Speelman, D. and Geeraerts, D. (2015). A sociolinguistic analysis of borrowing in weak contact situations: English loanwords and phrases in expressive utterances in a Dutch reality TV show. *International Journal of Bilingualism*, 19, pp. 333–346.

Code-switching and bilinguals' grammars

Rena Torres Cacoullos and Catherine E. Travis

1. Introduction

Code-switching (CS), using two languages in a conversation, continues to amass scholarly attention. Though CS was once viewed, if noticed at all, as a haphazard mess not amenable to analysis, it has been a growing subject over the past 50 years. Attention to CS comes from different linguistic branches and approaches, ranging from psycholinguistic experiments seeking to illuminate language processing, to grammaticality judgments serving to adjudicate on syntactic theories, to recordings of conversations revealing usage in the speech community. CS has been implicated in change in bilinguals' grammars, a conjecture that can only be tested via proper delimitation of CS and diagnostics of change.

To begin to understand the impact of CS, we need to know what it is. We find a plethora of terminologies and taxonomies and an astounding lack of agreement even on how to recognise a code-switch. A major dispute bearing on the identification of CS is whether to characterise it as alternation between two languages, each of which retains grammatical independence, or rather as insertion of one language into the other, which would be the matrix language grammatically. In the latter situation, the relation between the two languages is asymmetrical, one being morpho-syntactically dominant, whereas in the former, both languages are at work and the question becomes that of discovering the structural sites of CS. The Equivalence Constraint states that bilinguals tend to avoid CS at points of word order conflict between the two languages; the notion of equivalence, congruence, or matching between languages features in many approaches to CS (Poplack, 1980; Poplack, 2015; see also Lipski, 1978; Muysken, 2000; Deuchar, 2005; Muysken, 2015).

The use of two languages within a speaker turn is exemplified in (1), from a Spanish-English bilingual corpus. Each line of transcription represents an Intonation Unit (IU) (Du Bois et al., 1993, 47, 52–55).¹ CS here occurs at IU boundaries, following final intonation contour (marked by a period), between lines (a) and (b), and after continuing intonation (marked by a comma), between (f) and (g) (the first exemplifies inter- and the second intra-sentential CS (cf. Poplack, 1980, p. 589)). Single-word insertion, or nonce borrowing, is illustrated in

lines (c) and (d) (*carrots*). (In the examples, the translation appears on the right, with stretches of speech originally produced in English in italics.)

- 1.
- | | | |
|----|---|--|
| a. | <i>... and I used to get chile quite a bit.</i> | '... and I used to get chile quite a bit. |
| b. | <i>.. y salía todo.</i> | .. and everything grew. |
| c. | <i>... carrots,</i> | ... carrots, |
| d. | <i>unas carrots grandotas.</i> | some big carrots. |
| e. | <i>asina las sacaba.</i> | I would pull them out like this. |
| f. | <i>... y ahora la tierra como que no me sirve=,</i> | ... and now it's like the earth is no good, |
| g. | <i>something is wrong.</i> | <i>something is wrong.'</i> |
- [04, 47:10–47:22]

Achieving agreement on CS has been impeded by the fact that few bilingual corpora are usable and suitable for accountable and interpretable analysis. *Usable* corpora record the speech of members of a well-defined speech community, transcribed following defined protocols. Such corpora allow us to contextualise contact phenomena both linguistically and socially, in order to detect which phenomena represent robust patterns and explain those patterns in light of bilingual practices in the community. *Suitable* corpora capture quantities of incontrovertible multiword instances of CS as well as of single-word insertions, including nonce and established borrowings (such as *carrots* and *chile* in (1)). Beyond appropriate corpora, the second core methodological issue is the lack of agreed-upon evaluation metrics to assess competing accounts. In this chapter, we look to how bilingual speech itself can elucidate the structure of CS and its relation to other outcomes of language contact, remembering that it is the speakers using two or more languages who are the locus of contact (Weinreich, 1953/1968, p. 1).

2. Historical overview: mechanisms of change

A widespread hypothesis is that CS promotes contact-induced change, especially grammatical convergence, by encouraging shared structures or simultaneous activation of bilinguals' two languages. How would CS bring about change, though? Spelling out the proposed mechanisms of change bears directly on the predictions that can be made in order to test the hypothesis.

Let us take a single candidate for contact-induced change to illustrate how multiple accounts and even contrasting predictions have been advanced. Subject pronouns are often expected to *increase* in overall rate in null subject languages (such as Spanish or Turkish) in contact with non-null subject languages (such as English or Dutch) (e.g., Heine and Kuteva, 2005, p. 70; Doğruöz, 2014). The reasoning is that the distribution of a variant form converges with that of an analogous form in the contact language (Backus, 2005, p. 333).

Varied explanations implicate erosion of discourse-pragmatic constraints on choice of a pronominal vs. an unexpressed subject, in particular, weakening of the effect of the accessibility of the referent. One usage-based account for weakening of discourse-pragmatic constraints is that when speaking the non-null subject language bilinguals are not practicing

such constraints, which are ostensibly inapplicable in a language like English (Otheguy and Zentella, 2012, pp. 167–168). Alternatively, appealing to the intrinsic demands of bilingualism, some processing accounts consider the overt pronoun a default (Sorace et al., 2009, pp. 473–474), though presumed cognitive ‘costs’ of CS depend on bilinguals’ linguistic experience, as well as on the psycholinguistic measure employed (Adamou and Shen, 2019; Johns, Valdés Kroff and Dussias, 2019). In either case, whether erosion is attributed to insufficient practice or bilingual processing demands, the constraint and its workings must be specified, since the opposite prediction may follow. A *decrease* in pronoun rate would follow from, for example, reduced use of the first person singular pronoun *yo* ‘I’ in certain interactional contexts, as observed among Spanish speakers in the USA shifting to English (Silva-Corvalán, 1994, pp. 147–163).

It is thus worth delving into mechanisms of change, capitalising on speech data from bilinguals who make regular use of both their languages and switch smoothly between them.

A well-understood mechanism of change is grammaticalisation (e.g., Bybee, 2015, pp. 117–160). *Contact-induced grammaticalisation* is said to be a kind of grammatical replication, by which an existing structure is used with grammatical meaning modelled on another language (Heine and Kuteva, 2005, pp. 79–122). The aspectual periphrases formed with a gerund in English and Spanish provide an example. Both developed from locative expressions, following a cross-linguistic grammaticalisation path. Nevertheless, English *be* + *V-ing* is an obligatory expression of progressivity the absence of which signals habitual meaning (e.g., *I’m drinking decaf* vs. *I drink decaf*) (Bybee, 2015, p. 193). In this it is farther along the grammaticalisation path than Spanish *estar* ‘be (located)’ + *V-ndo*, which still alternates with simple verb forms to express progressive aspect. This can be seen in (2), where a simple imperfective form in (b) refers to an activity in progress at the moment of reference (note that this simple verb form is most naturally rendered as *be* + *V-ing* in English) (Torres Cacoullos, 2012, pp. 103–107).

2.

- | | | |
|----|--|---|
| a. | <i>cuando se le voló,</i> | ‘when it ((his hat)) blew off, |
| b. | <i>...(1.0) venía_[IMPF] un carro y lo tropelló.</i> | <i>...(1.0) a car was coming_[IMPF] and ran over it.’</i> |

[05, 18:59–19:03]

Degree of advancement along a shared evolutionary path constitutes a *diagnostic difference*, that is, a locus of structural divergence between the languages in contact that permits detection of contact-induced change (see Poplack, 2018, p. 26, on the ‘conflict site’). Identifying a diagnostic difference in turn allows us to pose a quantitative question to probe contact-induced grammaticalisation: does Spanish in contact with English show increased use of *estar* + *V-ndo* relative to simple verb forms in progressive aspect contexts, as compared with non-contact Spanish varieties? Not necessarily so. In Puerto Rico, *estar* + *V-ndo* rates in present progressive contexts are similar for bilinguals and monolinguals (22%, N = 384 and 27%, N = 258, respectively, $p = 0.11$ by Fisher’s exact test) (Cortes-Torres, 2005, p. 53). In New Mexico, in the southwestern USA, *estar* + *V-ndo* rates in past progressive contexts, exemplified in (2), are similarly high as in a monolingual Latin American benchmark (19%, N = 313 vs. 22%, N = 284); and they are similarly low in past habitual contexts (3%, N = 509 vs. 4%, N = 91, respectively) (Dumont and Wilson, 2016, pp. 410–412). To illuminate the existence and nature of a candidate change, note that the quantitative argument rests on co-occurrence patterns – rates of alternative forms in particular contexts.

Simplification involving decreased use of one variant form without a counterpart in the other language (alongside increased use of another with a counterpart) may also be a mechanism of contact-induced change (e.g., Silva-Corvalán, 1994, p. 3). An example is Spanish mood choice in contact with English. Consider the complement clauses in (3) and (4) with Subjunctive and Indicative, respectively, under the same main clause verb 'seem.' Studies of varieties of Spanish spoken in the USA have asserted simplification based on lower rates of the subjunctive vs. the indicative.

3. *pero parece que pudieran poner*_[SUBJ] *a sign,*
'but it seems like they could have put'_[SUBJ] *a sign,'*
[29, 39:36–39:39]
4. *parecía que estaba*_[IND] *toda detenida.*
'it seemed like I was'_[IND] *all caught up.'*
[06, 36:50–36:52]

Yet a more rigorous gauge of simplification is once again supplied by co-occurrence patterns. For example, frequency measures according to the main clause verb have exposed limited subjunctive productivity, and that this holds in Romance languages regardless of language contact. Thus, much subjunctive selection is determined by the lexical identity of the main clause verb, and subjunctive tokens largely occur with a limited number of verbs that take the subjunctive categorically, leaving little room for semantic-pragmatic considerations in mood choice (Poplack et al., 2018). LaCasse (2018) observes such lexical routinisation in Spanish, for both a US Spanish variety and a monolingual benchmark. Where there is room for variability between subjunctive and indicative, across the comparison corpora the same linguistic conditioning applies (for example, according to main clause polarity). Why the finding against contact-induced simplification reported in numerous studies of US Spanish? This difference stems from the data source in sampling bilinguals who habitually speak both languages (vs. speakers shifting to the majority language), but also from the evaluation metrics in going beyond overall rate to explore linguistic variation patterns.

Finally, *contextual distribution via code-switching*, or shifts in the contexts of use of a form in bilinguals' speech, is a mechanism of variation and change that can explain contact effects (cf., Torres Cacoullos and Travis, 2018, p. 95). An example is the phonology of cognate words. For word-initial Spanish /d/, lenited and stop allophones exist in variation depending on the preceding phonetic environment, whereas English word-initial /d/ has a stop allophone only. In Spanish in contact with English, the rate of lenition is lower in cognate words such as *doctor* than in non-cognates such as *dolor* 'pain' (Brown, 2015, p. 396). The effect of cognate status follows from cumulative usage. How so?

Variation and change is affected by the 'usage history' of words, as lexical representations are based on speakers' linguistic experience (Bybee, 2010, p. 43). The effect of an alternating phonetic context – as is the case preceding word-initial /d/ (e.g., *el dolor*, *mucho dolor* 'the pain, much pain') – is cumulative, such that words used more frequently in a phonetic environment that favours a newer variant undergo change earlier than other words (Bybee, 2015, p. 87). Factoring in bilinguals' use of English as well as Spanish, and considering the interactivity between cognates in bilingual lexical representation, Brown (2015, p. 399) shows that frequency of occurrence in contexts propitious to lenition has the same effect on /d-/ lenition in cognates as in non-cognates (left panel, Figure 14.1), but cognates are exposed less frequently to such contexts (right panel), resulting in an overall lower rate of lenition in cognates. The cognate effect

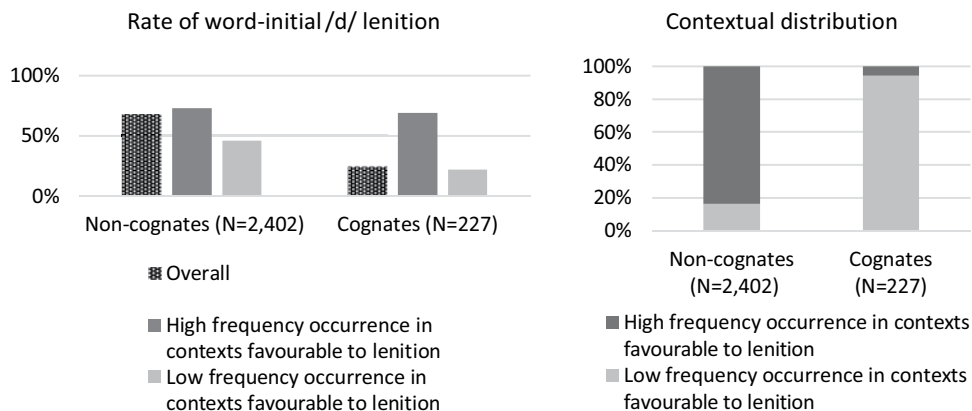


Figure 14.1 Rates of word-initial /d/ lenition (left panel) and contextual distribution (right panel) according to frequency of occurrence in contexts favourable to lenition, for cognate and non-cognate words in a US Spanish variety

here is thus a corollary of shifts in the contextual distributions of structures once bilinguals' experience in both their languages is included.

3. Critical issues and topics

Conflicting definitions and therefore delimitations of what counts as CS have hindered advances in our knowledge of the conditions on CS and the consequences it may have for grammars in contact. A far-reaching decision concerns the treatment of single-word insertions from one language into the other as CS or as spontaneous borrowing. In this section, we also discuss how patterns of variation between competing forms can be used to assess the role of surmised triggering elements in CS and the actual scope of language mixing phenomena.

Code-switching and nonce borrowing: alike and disparate

Borrowing involves lexical retrieval from the recipient language only, while CS involves drawing from two languages in real time. There is no dispute there. However, *nonce borrowing* and CS have been treated as one and the same in insertional models of CS, which are premised on the blanket assumption that single-word insertions are switches (e.g., Myers-Scotton, 1993). At the crux of the matter are content words typically unattested in dictionaries of the recipient language, as with English *carrots* embedded in a Spanish stretch of speech in (1), lines (c) and (d) (but not *chile* in line (a), as an established borrowing). Muysken's (2000) influential classification of CS distinguishes the insertional and alternational kind, responding to the awareness of at least some kind of difference between using one and more than one word from another language. What similarities or differences are there between what we can operationally identify as unattested other-language (inserted) single words versus (alternating) multiword strings?

Single-word insertions, or lone items, and multiword CS strings certainly share social and cognitive properties, such as those noted in Figure 14.2. Both entail bilingualism, neither is predictable, and the usage of both is guided by community norms.

Single-word insertions and multiword CS strings both:

1. Require bilingualism
(though they may not involve co-activation of the two languages in the same way).
 2. Are unpredictable at any given point
(though they may be preferred at particular loci).
 3. Occur (in)frequently, depending on local community norms
(though single-word insertions in aggregate tend to be more frequent than multiword strings, and hence many remarks about CS actually pertain to single-word insertions).
-

Figure 14.2 Shared social and cognitive properties of single-word insertions and multiword code-switching strings

Single-word insertions and multiword CS strings nevertheless sharply diverge in their structural linguistic properties, including for word class and word order, as well as categories and constructions. Indeed, perhaps the biggest breakthrough for the study of CS since the 1980s is the knowledge of how bilinguals morpho-syntactically treat lone items in their spontaneous language use, formalised as the *nonce borrowing hypothesis* by Poplack and colleagues (e.g., Poplack, 2018).

For example, the syntactically unattached *carrots* in line (c) in (1) is a nonce borrowing rather than a code-switch based on the tendency of English single nouns to assume the linguistic features of loanwords in this particular speech community (Torres Cacoullos and Aaron, 2003). This is demonstrably so in line (d), where the word order and nominal categories are Spanish, as *carrots* appears with a post-posed adjective and feminine gender – both diagnostic differences between Spanish and English. The aggregate patterns establishing loanword status remain true even if integration is not displayed on a case-by-case basis, as with the determinerless and unmodified *carrots* in line (c).

Consider, in contrast, the multiword CS in example (5), where again, diagnostic differences allow us to observe the integrity of the use of English and Spanish as separate languages – line (c) features the English-particular pattern of preposition stranding (*talking about*), and line (d) a proclitic indirect object (*le decían* ‘they used to call (to) **him**’), consistent with Spanish grammar.

5. (talking about nicknames for different people named Miguel)
 - a. *más antes había muchos ~Migueles,* ‘before there were a lot of Miguels,’
(6 IUs intervening)
 - b. *pa’ saber la **difference** which --* ‘to know the **difference** which --
 - c. *.. Miguel they were talking about,* .. *Miguel they were talking about,*
 - d. *le decían ~Miguel Bajo,* they used to call him Miguel Short,’
- [17, 01:12:20–01:12:24]

The treatment of other-language nouns has consequences for the delimitation of switches and syntactic analyses of CS. In line (b), CS undeniably occurs within the prosodic unit (IU) but whether it counts as occurring at the determiner-noun juncture (beginning with *difference*, as represented in (a) in (6)) or at the clause boundary (following *difference*, as in (b)) follows from the analyst’s decision on whether to categorise *difference* as a nonce borrowing. Note that neither switch site would violate the Equivalence Constraint (in both languages the

determiner precedes the noun), but to establish patterns of CS, the issue is one of discovering which ‘permissible’ switch sites are actually favoured.

6.
 - a. [*pa’ saber la*] [*difference which Miguel they were talking about*]
 - b. [*pa’ saber la difference*] [*which Miguel they were talking about*]
 ‘to know the difference which Miguel they were talking about’

Finally, let us consider whether, as has been proposed, borrowing and CS form a linguistic continuum (e.g., Matras, 2009, p. 111). Items on a continuum display a shared property to a greater or lesser degree, for example, some derived words are more compositional than others, as with English *tasteless* vs. *listless* (Hay, 2002, p. 1048). Continua also often involve a diachronic relationship, for example, lexical morphemes become grammatical morphemes, as with the grammaticalisation of ‘be located’ in progressive aspect constructions.

From a diachronic perspective, the assumption that insertional or single-word CS is the forebear of lexical borrowing may seem a natural one (Backus, 2005, p. 315). Yet it is at odds with the recognition that lexical borrowing is not contingent upon CS (Haspelmath, 2009). More decisive is the empirical evidence. From a corpus-based study spanning some 60 years (Poplack and Dion, 2012), we learn on the one hand, that single-word insertions tend to be instantaneously integrated into the recipient language on first use; and on the other, that despite on-the-spot morphosyntactic integration, most nonce borrowings do not diffuse to become dictionary-attested established loanwords.

One synchronic criterion applied to ascertain the status of single-word insertions relies on frequency, whereby concepts more often expressed by other-language-origin words are deemed loanwords whereas ones that are less often so are instances of CS (Myers-Scotton, 1993, pp. 191–204). Intuitive as it may seem, this social criterion is contradicted by the instantaneous structural integration of single-word insertions into the recipient language on first use. In addition, such a frequency criterion produces implausible classifications, for example, of some English school-related terms in New Mexican Spanish as loanwords and others as switches. To illustrate, when speaking Spanish, English *grades* is the preferred way of referring to school grades (*notas* in Spanish), as in (7), while *teacher* is not a preferred term (instead *maestra/o* is). Yet the putative single-word CS with the dictionary-listed compound *substitute teacher* in (8) is patently integrated, appearing as a bare predicate nominal, consistent with Spanish and in contradiction with English grammar.

7.

| | |
|--|--|
| <i>vieras los grades que tiene.</i> | ‘you should see the grades he has.’ |
| | [24, 10:47–4:48] |

8.

| | |
|---|--|
| <i>era Ø substitute teacher en la escuela,</i> | ‘she was (a) substitute teacher in the school.’ |
| | [21, 4:49–4:51] |

In assessing the relationship between CS and nonce borrowing we must distinguish between social-cognitive properties, which are shared, and structural properties, which diverge sharply, not as a matter of degree (as summarised in Figure 14.3). The synchronic and diachronic relationships between CS and borrowing thus dictate that they be recognised as distinct manifestations of language contact, an essential distinction for meaningful analysis of bilingual speech.

Single-word insertions and multiword CS strings differ in:

1. Word class:

Single-word insertions are disproportionately nouns;

no such disproportion has been reported for the words constituting multiword CS strings (which are often clauses).

2. Word order at the juncture between the two languages:

Single-word insertions follow the word order of the recipient language; multiword CS strings are juxtaposed at points where the two languages have the same word orders (the Equivalence Constraint).

3. Internal morpho-syntax:

The categories and constructions of single-word insertions are of the recipient language;

the internal constitution of multiword CS strings is consistent with the grammar of their respective language.

Figure 14.3 Divergent structural linguistic properties of single-word insertions and multiword code-switching strings

Code-switching motivations vs. mode: letting the data speak

Multiple explanations have been offered for why a speaker may combine two languages, be that via insertion or alternation (cf., Matras, 2009, p. 114ff for a summary). In some contexts, switching languages may be motivated by extralinguistic considerations of setting, topic, or the bilingual abilities of the interlocutors (Blom and Gumperz, 1972). It may also be used for rhetorical effect, such as to make metalinguistic commentary, with quoted speech, or for discourse organisation (Auer, 1995; Wei and Milroy, 1995; Aikhenvald, 2002, pp. 190–191). And the choice of language has often been perceived to have a social meaning related to the establishment of identity, as, for example in 'metaphorical CS' (Gumperz, 1982, Ch. 4). One of the challenges to be met by pragmatic explanations is to demonstrate that CS does carry the proposed function. To do so, it must be shown that the same external trigger or meaning tends to be absent when speakers are not engaging in CS (that is, for non-occurrences, as outlined in the following section).

It is not always the case, however, that there is a specific trigger for combining two languages. In 'intra-situational code-switching' (Poplack, 2015, p. 218), the two languages occur in a single speech event as a general discourse mode (similar to what Grosjean (2001) has referred to as 'bilingual mode'). In this case, the use of one or the other language, or a switch between them, is not any more locally predictable than telling a joke (Poplack, 1993, p. 276; Sankoff, 1998, p. 39).

How can we know whether CS is locally motivated? Bilinguals' responses to direct questions about meanings associated with different languages or about reasons for switching between them cannot be taken as evidence, as speaker intuitions are notoriously unreliable (cf., Blom and Gumperz, 1972, p. 430). This is even more so when combining two languages is stigmatised (as is suggested by labels such as *Spanglish* or *Franglais*), and bilinguals have been known to judge utterances that they themselves had produced as incorrect (Torres Cacoullous and Travis, 2018, pp. 39–40, and references therein). Thus, we need to look at how speakers actually use language.

Naturally arising commentary extracted through content analysis of a corpus may provide clues about speaker attitudes to the two languages (Torres Cacoullous and Travis, 2018, pp. 62–71). For example, in a 300,000-word New Mexico Spanish-English corpus (NMSEB),

there is not a single mention of any value attached to one or the other language (outside of a few fleeting mentions of church hymns). There is, however, commentary on unreflecting switching between the two languages, as in the following example.

9.

| | |
|--|---|
| <p>...(1.0) <i>we can</i>=, ... <i>be talking in English</i>, <i>y</i>=, ...(1.0) <i>en media conversación</i>, <i>estamos hablando español=l.</i></p> | <p>‘...(1.0) <i>we can</i>, ... <i>be talking in English</i>, and, ...(1.0) <i>in the middle of the conversation</i>, we are talking in Spanish.’</p> |
|--|---|

[27, 21:54–22:01]

Other clues come from structural analysis, for example, the distribution of the grammatical person of clause subjects. We might expect grammatical persons to differ in their distributions across the languages if they carry different social meanings, for example related to speaker involvement (Gumperz, 1982, p. 81). Specifically, the more personal language might be preferred for first and second person subjects, and the more objective language for third person subjects. In the New Mexico corpus, however, as seen in Figure 14.4, the grammatical person distribution is very similar for Spanish and English, the presumed more ‘personal’ and ‘objective’ languages respectively. This kind of operationalisation and quantitative test can reveal community-specific motivations for CS, or, in this case, the absence thereof.

Degrees of bilingual practice: community norms

The familiar notion of individual degree of bilingualism is fraught with often incommensurate ways of assessing proficiency. More apt is what we may call degrees of bilingual practice, or actual use of the two languages and strategies for combining them.

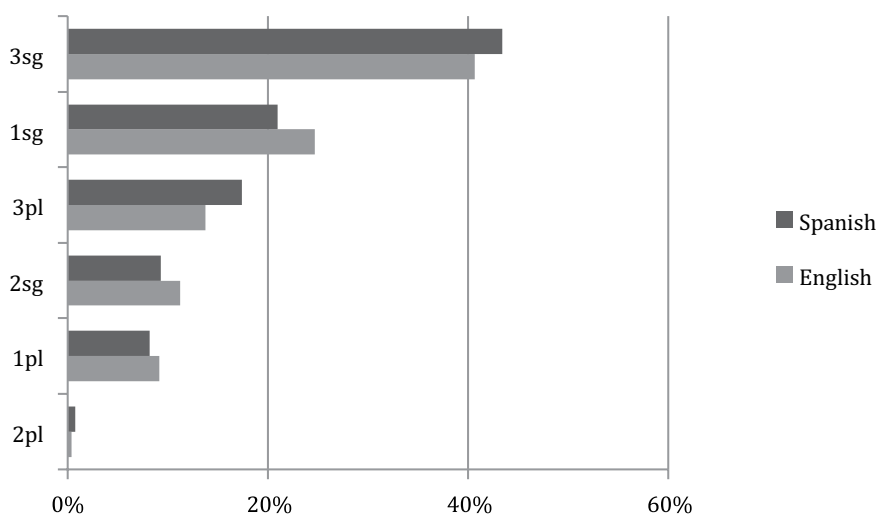


Figure 14.4 Distribution of clauses by subject person in bilinguals’ Spanish and English
Language of clause based on the language of the verb; Spanish N = 17,689; English N = 18,322
 Source: Adapted from Torres Cacoullos and Travis, 2018: Figure 9.1

Language combining strategies are governed by the norms of the bilingual speech community. For example, some communities are reported to avoid CS (e.g., indigenous Amazonian languages in the Amazonian Vaupes (Epps and Michael, 2017)).² Others use CS primarily for rhetorical purposes, such as meta-linguistic commentary, repetition or translation (e.g., French Canadians in Ottawa-Hull (Poplack, 2018, pp. 67–71)). And in yet other communities CS is a discourse mode, as is the case in New Mexico (cf., Gonzales, 1999). In the New Mexico Spanish-English bilingual (NMSEB) corpus, furthermore, the clauses are evenly divided between the two languages (see note on Figure 14.4). Thus, neither is more frequently the 'matrix language' (Myers-Scotton, 1993) nor 'more salient' (Myslín and Levy, 2015, p. 871).

Degrees of bilingual practice can be considered by situating the contact setting in its social context. In a conversational corpus of over half a million words of Chintang (a Tibeto-Burman language), for instance, Stoll et al. (2015 pp. 6, 11) find that children produce proportionally more Nepali insertions than older speakers, and syntactically integrate them less. Younger speakers are also exposed to, and use, more Nepali (Stoll et al., 2015, p. 11). With this socio-historical contextualisation of the contact setting, they interpret the difference between generations not as grammatical change, but rather as change 'in the nature of bilingualism' (2015, pp. 9, 10).

Knowledge of bilingual community practices is needed to unravel language change or continuity under contact. For example, Stanford (2008) considers clan-based dialect differences in lexical tone among the Sui, an indigenous community in southwest China. According to exogamy in this region, marriage takes place between people of different clans and women move into the village of their husband. Yet these women retain the clan-specific tonal patterns, demonstrating continuity despite intense contact.

Such quantitative descriptions of speech corpora diverge from the commonly embraced view that CS inevitably leads to grammatical convergence (e.g., Thomason and Kaufman, 1988, p. 154; Backus, 2005, p. 334). Influential was Gumperz and Wilson's (1971) study asserting that prolonged contact and constant CS between languages in Kupwar, India, had resulted in a shared syntax, with only lexical differences remaining (1971, p. 256). Some 40 years later, a study in this same region suggests a reconsideration of this claim. Instead of isomorphism, Kulkarni-Joshi (2016, pp. 168–169) documents (for both young and old speakers) considerable variability in structures in a corpus of spoken narratives. Though variability is a condition for change, 'not all variability and heterogeneity in language structure involves change' (Weinreich, Labov and Herzog, 1968, p. 188). She further points out that the inter-translatability reported by Gumperz and Wilson (1971) may be an artefact of the data collection method, which itself was based on translations between the three languages (Kulkarni-Joshi, 2016, pp. 168–169). Stories were told in one language and then retold in another, and 'translations were further edited to substitute translation equivalents so as to minimise the language distance in those instances where different expressions had been used' (Gumperz 1967, p. 54, cited in Kulkarni-Joshi, 2016, p. 169).

Systematically compiled speech corpora afford a view of community usage patterns that hold across individuals and sub-groups. To discern patterns, the key methodological concept for quantifying is the scientific principle of *accountability*: count all occurrences of the linguistic form of interest in the data set, but also all of the contexts in which the form *could have* occurred, even if it did not (Labov, 1972, p. 72).

Sankoff and Brown's (1976) account of the development of the relativiser use of *ia* in Tok Pisin, an English-based creole spoken in Papua New Guinea, is a model study for tracing 'the origin of syntax in discourse.' Counting occurrences and non-occurrences of the form, Sankoff and Brown describe how Tok Pisin *ia* developed from a place adverb (based on English

‘here’), to a deictic or demonstrative, then undergoing further extension for ‘bracketing’ use, as in relativisation and clefts (1976, p. 663). The structure is shown to arise through a grammaticalisation process in everyday speaker interactions, rather than being primarily due to other-language influence. Such insights are thanks to analysis of co-occurrence patterns in a corpus from speakers whose socio-personal history and the speech community they represent are known to the analysts.

4. Current contributions and research: code-switching in action

To substantiate the hypothesis that CS promotes convergence, we need a synchronic test of contact-induced change. The prediction to be tested is whether the grammar of one language is more similar to the other when bilinguals are code-switching than when they are not. In particular, the linguistic conditioning of variable structures should be impacted in the context of CS, and differ from that in its absence. To execute the test, we thus need a demarcation of CS, as well as measures of grammatical similarity or difference.

How to count code-switching

Quantifying CS presents an enduring challenge. First, failure to distinguish CS from borrowing risks obscuring any potential impact of CS. The question for CS-induced convergence is thus whether alternation of multiword strings (unambiguous CS) steers the grammar of one language in the direction of the other. Second, there is agreement that for CS to result in language change it must be ‘extensive’ (e.g., Myers-Scotton, 1993, p. 211). Researchers indeed often use quantitative qualifiers such as ‘abundant,’ ‘dense,’ or ‘frequent’ in reference to CS. Unless numbers are provided, though, assertions of ‘extensive’ CS remain uninterpretable. Replicable measures of CS frequency are sorely needed to enable comparisons across data sets and communities.

For single-word insertions, an overall frequency measure has been obtained by considering the proportions of tokens of other-language vs. native words in a given corpus (e.g., Tadmor, 2009, pp. 56–57; Adamou, 2016, pp. 212–214). A caveat is that overall rates will vary according to genre and topic. For example, in one study the same speaker showed a rate of single other-language words as much as seven times greater in one recording than in another (Aaron, 2015, p. 480). Topic driven differences will be exacerbated the smaller the corpus, but a way to mitigate this is to apply the principle of accountability (see Section on ‘degrees of bilingual practice’). A measure of borrowability, for example, can be obtained by making the comparison within semantic domains, that is, the proportion of tokens that are borrowed vs. native in a given domain, for example, technology-, or food-, or school-related terms (Aaron, 2015, pp. 465–468).

Frequency measures for single-word insertions are inapplicable to CS, however. Circumscribing the universe of contexts where CS *could* occur (and not only those where it actually did occur) is far from straightforward since it is impossible to predict the occurrence of CS at any particular point in the discourse (Poplack, 1993, p. 276; Sankoff, 1998, p. 39).

One way to define a universe of CS contexts is by syntactic boundary (cf., Sankoff and Poplack, 1981). An example is the boundary between a main and subordinate clause, such as a main and conditional clause (Backus, 2008, p. 254), or a main and complement clause, as in (10) and (11). In a bilingual corpus of English and Spanish, there are 66 tokens of CS between a main and complement clause, compared with 1,070 non-occurrences of CS (that is, tokens of [Main Clause + Complement Clause] that are unilingual Spanish or English, as well

| | | | |
|----|------------|-------------------------------|-------------------------------|
| f. | CS present | <i>Ø estaba como el de --</i> | (I) was like that one from -- |
| g. | | ...(1.0) @@@ | ...(1.0) @@@ |
| h. | | ...(1.4) como el este, | ...(1.4) like that one, |
| i. | | Forrest @Gump. | Forrest @Gump.' |

[04, 35:25–35:40]

Applying this measure of proximate CS to all instances of a linguistic variable in a corpus provides an indication of the distribution of language combining strategies, distinguishing contexts that are in proximity to CS (as in line (f)) from those that are locally unilingual (as in lines (b) and (c)) as well as from those that involve single-word insertions or other items for which the role of the other language is to be determined (for example, with proper nouns or with interlocutor-produced other-language material).

Figure 14.5 illustrates how such a classification may be manifested in bilingual speech data, the darkest shade representing one language, the lighter shade another, and the lightest shade all other contexts. Here we see strings of IUs produced in the one language (e.g., the IUs in lines (f)–(h) and (i)–(l)); IUs in one language containing a single item from the other (e.g., (a) and (e)); and other contexts (such as IUs containing material that can't be assigned to a language or IUs produced by an interlocutor (e.g., (b), (m)–(n)). The pertinent data sets for comparison to test the impact of CS are constituted by instances of the linguistic structure under consideration occurring in the proximity of CS vs. in a unilingual environment. In this schematised example, there are three instances of the variable of interest in the contexts to be compared; one occurs in the presence of CS (i), and two in its absence ((k) and (q)).

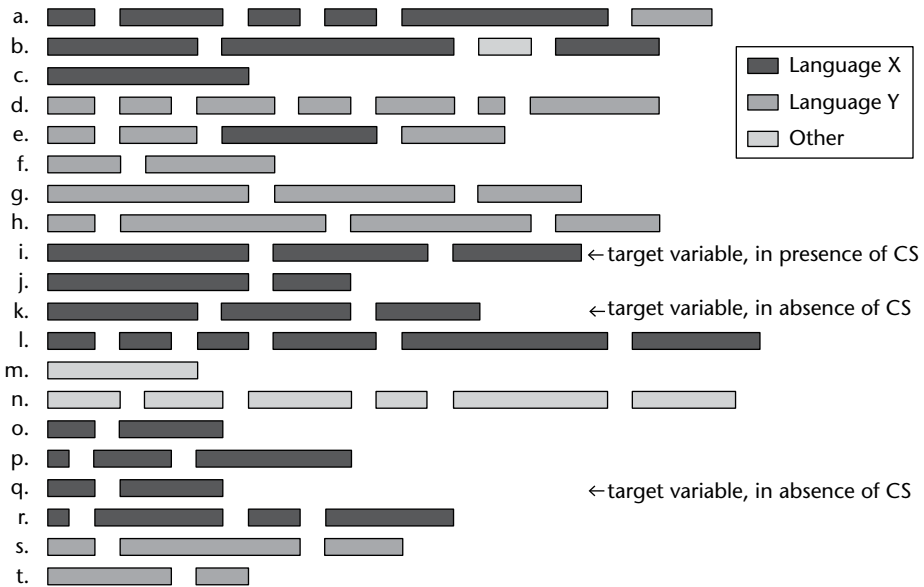


Figure 14.5 Distribution of a linguistic variable according to the proximate presence of code-switching in a bilingual speech corpus

How to measure grammatical (dis)similarity

Having demarcated CS contexts, convergence must still be determined – how can grammatical similarity or difference be measured in order to corroborate *increased* grammatical similarity? The answer is to compare the linguistic conditioning of analogous linguistic variables across data sets.

Subject expression again serves to illustrate, drawing on the findings from Torres Cacoullos and Travis (2018, Chs. 6 & 7). The variation in Spanish was seen in (12). Unexpressed subjects also occur in English, albeit within a much more limited envelope of variation than in Spanish. English unexpressed subjects are restricted to coordinate clauses, as in line (c) in (13), and, in non-coordinate clauses, to prosodic-initial position, as in (d) and (e) in (14).

13.

- a. *I go home,*
- b. *I take a shower,*
- c. *and I sit and Ø watch the novela, (('soap opera'))*

[12, 04:35–04:38]

14.

- a. *.. I was able to scramble and,*
- b. *.. find the --*
- c. *my other flashlight,*
- d. *... Ø turned it on,*
- e. *Ø worked on the one I had just broke.*

[16.1, 26:06–26:12]

Given its restricted envelope of variation, English has a rate of pronominal subjects approaching ceiling, with some estimates lying at around 97%, compared with rates ranging from 10 to 50% in Spanish (Torres Cacoullos and Travis, 2018, pp. 112, 151). Thus, the suggestion is that convergence with English would be seen in an increase in pronominal subjects (see Section 2). However, overall frequency rate increases alone are an unreliable measure of grammatical change, as they may fluctuate due to extra-grammatical factors such as genre; for this reason, the threshold for a linguistically significant increase is unknown.

To discern increased similarity, or change in the direction of English, we instead look to diagnostic differences in the linguistic conditioning of the variation, that is, where the co-occurrence patterns in Spanish differ from those in English. Comparison across languages shows that the probabilistic constraints replicated in numerous studies of Spanish correspond with cross-linguistic tendencies in subject pronoun expression: the same direction of effect for the same predictors (e.g., accessibility, priming, tense-aspect) obtains in a range of other languages quantitatively studied, regardless of language type – including English. Thus, within the restricted envelope of variation, English is actually qualitatively similar to null-subject and other traditional language types (Torres Cacoullos and Travis, 2019). What, then, are the interlingual differences that will allow us to measure potential convergence in contact varieties?

Even when probabilistic constraints are shared, diagnostic differences may be sought in the magnitude of those constraints. Here, we dissect the most replicated constraint for subject expression, accessibility, which concerns recency of mention, most often configured

in terms of coreferentiality with the preceding clause subject. For example, (15) illustrates a non-coreferential context, with a switch in subject between the target clause in (c) and the preceding clause in (a), and (16) illustrates a coreferential context. Pronouns are favoured (as a tendency, not a categorical rule) in non-coreferential, over coreferential, contexts.

Within coreferential contexts, accessibility is reinforced by linking between clauses, syntactic (via a conjunction) and/or prosodic (via intonation). In (16), the two coreferential-subject clauses are not linked – there is no conjunction; the first clause ends with final intonation (indicated with the period); and there is a pause separating the two clauses (indicated by the three dots, with the length given in seconds following). In (17), like (16), the two clauses share the same subject, but here the target clause in (b) is linked to the preceding, via both the conjunction *y* ‘and,’ and continuing intonation on the first clause (marked with a comma). In this structurally linked context, where the subject can be considered to be more accessible, pronouns are less likely to occur than in a coreferential non-linked context.

- 15.
- | | | |
|----|------------------------------|-------------------------------|
| a. | <i>y las veces que iba,</i> | ‘and whenever I visited, |
| b. | <i>pues=,</i> | well, |
| c. | <i>ahí Ø estaba siempre,</i> | (he) was always there, |
| d. | <i>you know?</i> | you know?’ |
- [04, 59:02–59:04]
- 16.
- | | | |
|----|---|--|
| a. | <i>.. y Ø se salió.</i> | ‘and (he) left. |
| b. | <i>...(1.0) al ratito Ø volvió pa’trás.</i> | ...(1.0) a bit later (he) came back.’ |
- [03, 18:44–18:48]
- 17.
- | | | |
|----|---|---|
| a. | <i>y luego de ahí Ø se iba para el otro lado,</i> | ‘and then from there (he)’d go somewhere else, |
| b. | <i>y Ø compra una cerveza.</i> | and (he)’d buy a beer.’ |
- [06, 18:42–18:44]

Figure 14.6 gives the probability with which a first- or third-person singular subject is realised with a pronoun, drawn from separate logistic regression analyses for six comparison data sets – benchmarks from conversational monolingual Spanish (CCCS) and English (SBCSAE) corpora in the sets of columns on the far left and right respectively; bilinguals’ Spanish and English on the second-most left and right columns (NMSEB); and, within these bilinguals’ Spanish, the absence and presence of CS in the middle, in accordance with the operationalisation depicted in Figure 14.5. Due to the preponderance of pronouns in English, we extracted all unexpressed 1sg and 3sg subjects, and a sample of pronominal subjects (two for each unexpressed). Thus, the relevant comparison is not the absolute values, but the relative heights of the bars within each corpus.

Note that the predicted direction of effect obtains throughout, including for English: from lowest to highest, the probability for pronominal expression goes from coreferential linked contexts (the darkest bar), to coreferential non-linked contexts to non-coreferential contexts (the lightest bar).

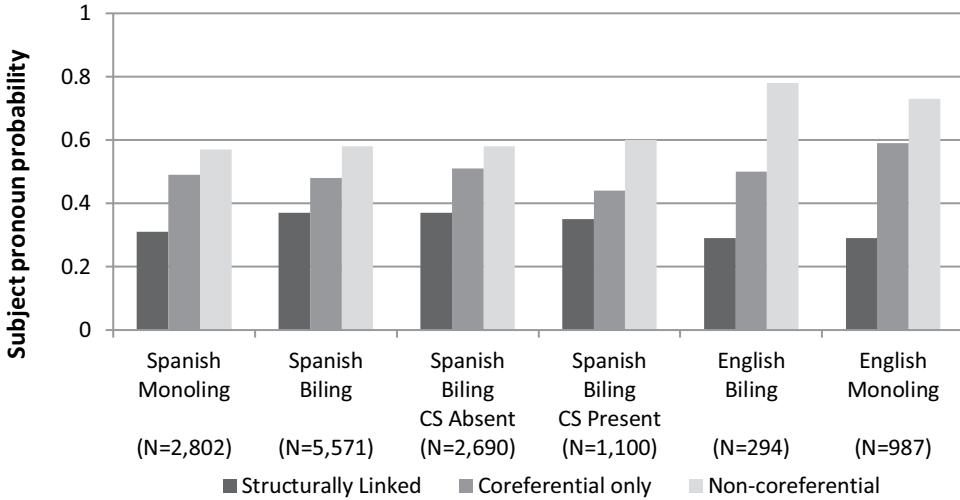


Figure 14.6 Interlingual similarities and differences in variation patterns: subject pronoun probabilities according to accessibility, in monolingual and bilingual varieties, and in presence vs. absence of code-switching

* Results from six independent logistic regression analyses including Accessibility configured in three levels as shown here, as well as Priming, Verb class, Tense, Person (for Spanish only), and Presence of English (Spanish Bilingual overall only). English based on a sample of two pronominal subjects extracted for each null subject.

(For full results and corpus information, see Torres Cacoullós and Travis, 2018, pp. 114, 184, 213–214.)

Yet this three-level measure of accessibility reveals a quantitative difference between English and Spanish in relative strength of the effect. We will first examine the two monolingual benchmarks. Accessibility is far more important for English than Spanish subject expression, being ranked higher than any other constraint within the English regression analysis but having the same or smaller effect size than other constraints in Spanish (Torres Cacoullós and Travis, 2019, pp. 672–674). The stronger effect for English can be seen here in the greater range between the probability values across the three degrees of accessibility (reflected in the heights of the bars), compared with Spanish. This interlingual difference in magnitude of effect provides a diagnostic on the basis of which to test convergence: convergence would be evidenced by a more pronounced disparity between non-coreferential and coreferential linked contexts in the likelihood of a subject pronoun in bilinguals' as compared with monolingual Spanish.

Thus, we come to the truest test of convergence, by comparing the same speakers' use of their two languages. We now juxtapose bilinguals' Spanish and English, as shown in the second sets of bars on the left and right. As we can see, the speakers maintain the same patterns as their monolingual counterparts. Not only is bilinguals' Spanish similar to monolingual Spanish, but bilinguals' English displays the stronger effect for accessibility of monolingual English. The notable corollary is that bilinguals' language varieties are no more similar to each other than the monolingual varieties are.

Given that these bilingual speakers separately retain Spanish and English patterning, the key question remains: what happens when they code-switch? If CS promotes structural similarity, then we should observe greater susceptibility to this English pattern – that is, a greater

magnitude of effect for accessibility – when speakers are engaging in CS than when they are not. Looking now to the middle two sets of bars in Figure 14.6, we see that Spanish patterns are retained, and the stronger effect of accessibility in English has not been transferred, even in the context of proximate use of English. This stringent test of the impact of CS shows that bilinguals are not impacted by the English pattern when speaking Spanish, even in the environment of using English nearby.

Bilinguals’ grammars in interaction

What does the lack of evidence for convergence mean for the bilinguals’ cognitive associations between languages? Structural priming across the two languages suggests answers.

Structural priming is where the use of one variant favours subsequent use of that same variant over alternatives. It is a robust factor in community-based studies of language-internal variation (e.g., Weiner and Labov, 1983) and also applies across languages. In the latter case, the use of one variant in one language favours subsequent use of an analogous variant in the other language, as in the dative alternation across English and German (e.g., *the lawyer sent his client the contract* vs. *the lawyer sent the contract to his client*) (cf., Gries and Kootstra, 2017). Subject pronoun expression is also impacted by cross-language priming across Spanish and English, such that the use of a coreferential English subject pronoun primes a Spanish pronoun, as in lines (a) and (b) in (18). (Travis, Torres Cacoullos and Kidd, 2017; Torres Cacoullos and Travis, 2018, Ch. 10).

18.

- | | |
|--|--|
| a. ... <i>I</i> was a statistician. | ‘... <i>I</i> was a statistician. |
| b. <i>yo</i> fui a !todos los basketball games. | ... <i>I</i> went to all the basketball games. |
| c. but <i>I</i> did all the stats. | but <i>I</i> did all the stats.’ |

[22, 11:22–11:29]

Priming provides a measure of association between linguistic classes or constructions. For example, in English ING variation between velar [ŋ] and alveolar [n] nasals, priming only occurs when the prime and target are in the same morphological class: *working* (vs. *workin’*) is primed by *kicking* but not by *morning* (Tamminga, 2016). In a seeming parallel, structural priming across languages has been taken to support the conjecture that bilinguals have a ‘shared syntax’ whereby coinciding grammatical structures ‘are represented once’ (Hartsuiker, Pickering and Veltkamp, 2004, p. 409).

If there is interaction between bilinguals’ grammars, then the result that CS does not impact Spanish subject pronoun expression in New Mexico bilinguals would appear all the more remarkable. To begin to tackle the question of bilingual associations, we zoom in on how coreferential subject priming operates for those instances of variable subject expression that occur in the proximate presence of CS. This is where there are most opportunities for cross-language priming (a full one-half of the instances with a prime (372/710) occur with a previous coreferential mention as an English pronoun, as in (18)).

The probability of a Spanish pronominal subject in the presence of CS according to the realisation of the previous coreferential mention is given in Figure 14.7.³ Spanish-to-Spanish priming is seen in that, compared with a previous unexpressed mention, a pronoun is more likely when the previous coreferential mention was a Spanish pronoun (the higher bar reflecting a higher probability). English-to-Spanish priming is seen in that a pronoun is also more likely when the previous coreferential mention was an English pronoun. It is this cross-language

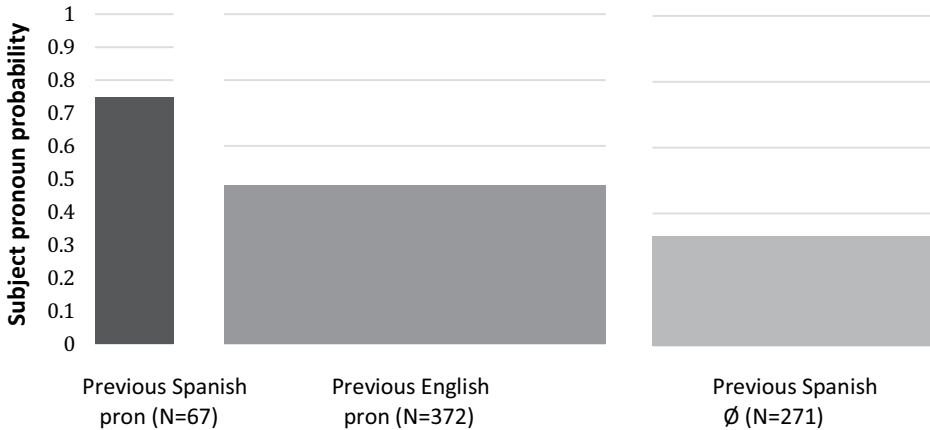


Figure 14.7 Priming across vs. within languages: subject pronoun probabilities in presence of CS, according to priming (realisation of the previous coreferential mention)

Results are from the same logistic regression analysis for the presence of CS reported in Figure 14.6; column thickness represents data distribution.

Source: Based on Torres Cacoullos and Travis, 2018, p. 184

priming that provides evidence that Spanish and English pronouns are associated for these bilingual speakers.

But also observe that the priming effect from English to Spanish is weaker than that from Spanish to Spanish, seen in the lower probability of a pronoun in the former context than the latter. This weaker effect is consistent with the independent grammatical patterns of subject expression discussed in the preceding section, and it indicates that association is a matter of degree. Using priming as a measure of degrees of association between constructions, we can say that [PRONOUN + verb] constructions are associated across Spanish and English but also that associations are stronger within the same language than across languages.

In sum, while cross-language priming attests to a connection between the two languages for these bilingual speakers, the fact that it is weaker than within-language priming corresponds to the separation of grammars shown in the previous section, even in the presence of CS. For these bilinguals, the grammars are connected, but they are not mixed. The evidence, then, is that in CS, speakers strictly alternate between two languages, each language retaining the same grammatical patterns as in the absence of CS.

5. Future directions

CS remains a contentious topic. The overriding methodological issues to be resolved involve appropriate data and systematic quantitative analysis, prioritising community norms over idiosyncratic instances and robust patterns over isolated cases.

We have seen that for there to be cross-fertilisation and cumulative advances in our understanding of the conditions on, triggers for, and grammatical consequences of CS, it will first be necessary to identify tokens of CS, delimit their boundaries, and define their units of occurrence. In demarcating CS, it will be profitable to separately evaluate single other-language words, which should not be automatically classed with multiword strings.

Defining the units of discourse to which CS may apply, or the universe of CS contexts, will allow analyses to account for not only all occurrences but also all non-occurrences of CS. Accountable reporting of CS distributions in turn will allow discovery of CS patterns: out of all places where bilinguals can switch, where they prefer to do so. For this, prosody offers a promising avenue, with the observed tendency for CS to occur at the boundary of prosodic units rather than within them, or the tendency to prosodically separate the two languages.

The conviction that grammatical convergence or mixing is an inevitable consequence of CS is ripe for review. Comparisons of language-internal variation patterns yield measures of interlingual grammatical similarity and dissimilarity, and a demonstration of change or continuity. As we've seen, just such a comparison reveals that CS – by bilinguals who make regular use of both languages and switch smoothly between them – is the alternation between structurally independent languages.

6. Further reading

Poplack, S. (1980). Sometimes I'll start a sentence in Spanish *y termino en español*. *Linguistics*, 18, pp. 581–618. (See also 2013, *Linguistics*, 51(Jubilee), pp. 11–14.)

A still unmatched quantitative community-based study, this article puts forward the basic insight that bilinguals tend to avoid CS at points of word order conflicts between the two languages, known as the Equivalence Constraint.

Poplack, S. (2018). *Borrowing: Loanwords in the speech community and in the grammar*. Oxford: Oxford University Press.

A synthesis of cumulative research across language pairs and contact situations, this book lays bare lexical borrowing as a process the principal mechanism of which is morphosyntactic integration.

Torres Cacoullos, R. and Travis, C.E. (2018/2020). *Bilingualism in the community: Code-switching and grammars in contact*. Cambridge: Cambridge University Press.

Relying on variation patterns in bilingual speech to advance quantitative diagnostics of grammatical similarity, this book demonstrates that speakers maintain distinct structures in their two languages even when engaging in code-switching.

Travis, C.E. and Torres Cacoullos, R., eds. (2015). Gauging convergence on the ground: Code-switching in the community. *International Journal of Bilingualism (Special Issue)*, 19(4), pp. 365–480.

Analyses of variation in a bilingual speech community demonstrate other-language influence in phonology but, in morphosyntax, continuity rather than change, and grammatical separation rather than mixing.

1. Related topics

Language contact in the lab, a variationist perspective, usage-based approaches, convergence

Abbreviations

| | |
|-------|---|
| 1sg | first person singular |
| 3sg | third person singular |
| CCCS | Corpus of Conversational Colombian Spanish (cf. Travis, 2005, pp. 9–25) |
| CS | code-switching |
| IMPF | imperfect |
| IND | indicative |
| IU | Intonation Unit |
| NMSEB | New Mexico Spanish-English Bilingual corpus |

SBCSAE Santa Barbara Corpus of Spoken American English (Du Bois et al., 2000–2005)
 SUBJ subjunctive

Notes

- 1 Examples are from the New Mexico Spanish-English Bilingual (NMSEB) corpus (cf., Torres Cacoulos and Travis, 2018, Chs 2 and 3). Within brackets following examples is the recording number and the beginning-ending time stamps of the lines reproduced. Transcription protocols are presented in the Appendix.
- 2 It has been suggested that 'limited code-switching and low levels of lexical borrowing' may accompany grammatical convergence (Epps and Michael, 2017), though the thresholds for high vs. low levels of CS and borrowing remain to be determined (see section on 'how to count code-switching').
- 3 Tokens for which a prime could not be identified in the previous ten clauses have been excluded (N = 390).

References

- Aaron, J.E. (2015). Lone English-origin nouns in Spanish: The precedence of community norms. *International Journal of Bilingualism*, 19(4), pp. 459–480.
- Adamou, E. (2016). *A corpus-driven approach to language contact: Endangered languages in a comparative perspective*. Boston and Berlin: De Gruyter Mouton.
- Adamou, E. and Shen, X.R. (2019). There are no language switching costs when codeswitching is frequent. *International Journal of Bilingualism*, 23(1), pp. 53–70.
- Aikhenvald, A.Y. (2002). *Language contact in Amazonia*. Oxford: Oxford University Press.
- Auer, P. (1995). The pragmatics of code-switching: A sequential approach. In: L. Milroy and P. Muysken, eds., *One speaker, two languages: Cross disciplinary perspectives on code switching*, 1st ed. Cambridge: Cambridge University Press, pp. 115–135.
- Backus, A. (2005). Codeswitching and language change: One thing leads to another? *International Journal of Bilingualism*, 9(3–4), pp. 307–340.
- Backus, A. (2008). Can a mixed language be conventionalized? In: Y. Matras and P. Bakker, eds., *The mixed language debate: Theoretical and empirical advances*, 1st ed. Berlin: De Gruyter, pp. 237–270.
- Balukas, C. and Koops, C. (2015). Spanish-English bilingual VOT in spontaneous code-switching. *International Journal of Bilingualism*, 19(4), pp. 423–443.
- Blom, J.P. and Gumperz, J.J. (1972). Social meaning in linguistic structures: Code-switching in Northern Norway. In: J.J. Gumperz and D. Hymes, eds., *Directions in sociolinguistics*, 1st ed. New York: Holt, Rinehart, and Winston, pp. 407–443.
- Brown, E.L. (2015). The role of discourse context frequency in phonological variation: A usage-based approach to bilingual speech production. *International Journal of Bilingualism*, 19(4), pp. 387–406.
- Bybee, J. (2010). *Language, usage and cognition*. Cambridge: Cambridge University Press.
- Bybee, J. (2015). *Language change*. Cambridge: Cambridge University Press.
- Cortes-Torres, M. (2005). *¿Qué estás haciendo?* La variación de la perífrasis *estar + -ndo* en el español puertorriqueño. In: D. Eddington, ed., *Selected proceedings of the 7th hispanic linguistics symposium*. Somerville, MA: Cascadilla Proceedings Project, pp. 42–55.
- Deuchar, M. (2005). Congruence and Welsh – English code-switching. *Bilingualism: Language and Cognition*, 8, pp. 255–269.
- Doğruöz, A.S. (2014). On the borrowability of subject pronoun constructions in Turkish – Dutch contact. In: M. Hilpert and J.O. Östman, eds., *Reflections on constructions across grammars*, 1st ed. Amsterdam and Philadelphia: John Benjamins, pp. 143–169.
- Du Bois, J.W., Chafe, W.L., Myer, C., Thompson, S.A., Englebretson, R. and Martey, N. (2000–2005). *Santa Barbara corpus of spoken American English*, Parts 1–4. Philadelphia: Linguistic Data Consortium.

- Du Bois, J.W., Schuetze-Coburn, S., Cumming, S. and Paolino, D. (1993). Outline of discourse transcription. In: J. Edwards and M. Lampert, eds., *Talking data: Transcription and coding in discourse*, 1st ed. Hillsdale: Lawrence Erlbaum Associates, pp. 45–89.
- Dumont, J. and Wilson, D.V. (2016). Using the variationist comparative method to examine the role of language contact in synthetic and periphrastic verbs in Spanish. *Spanish in Context*, 13(3), pp. 394–419.
- Epps, P. and Michael, L. (2017). The areal linguistics of Amazonia. In: R. Hickey, ed., *The Cambridge handbook of areal linguistics*, 1st ed. Cambridge: Cambridge University Press, pp. 934–963.
- Gonzales, M.D. (1999). Crossing social and cultural borders: The road to language hybridity. In: L. Galindo and M.D. Gonzales, eds., *Speaking Chicana: Voice, power and identity*, 1st ed. Tucson: University of Arizona Press, pp. 13–38.
- Gries, S.T. and Kootstra, G.J. (2017). Structural priming within and across languages: A corpus-based perspective. *Bilingualism: Language and Cognition*, 20(2), pp. 235–250.
- Grosjean, F. (2001). The bilingual's language modes. In: J.L. Nicol, ed., *One mind, two languages: Bilingual language processing*, 1st ed. Malden, MA: Blackwell Publishers, pp. 1–22.
- Gumperz, J.J. (1967). On the linguistic markers of bilingual communication. *Journal of Social Issues*, 23(2), pp. 48–57.
- Gumperz, J.J. (1982). *Discourse strategies*. Cambridge: Cambridge University Press.
- Gumperz, J.J. and Wilson, R. (1971). Convergence and creolization: A case from the Indo-Aryan/Draavidian border in India. In: D. Hymes, ed., *Pidginization and creolization of languages*, 1st ed. Cambridge: Cambridge University Press, pp. 151–167.
- Hartsuiker, R.J., Pickering, M.J. and Veltkamp, E. (2004). Is syntax separate or shared between languages. *Psychological Science*, 15(6), pp. 409–414.
- Haspelmath, M. (2009). Lexical borrowing: Concepts and issues. In: M. Haspelmath and U. Tadmor, eds., *Loanwords in the world's languages: A comparative handbook*, 1st ed. Berlin: De Gruyter Mouton, pp. 35–54.
- Hay, J. (2002). From speech perception to morphology: Affix ordering revisited. *Language*, 78(3), pp. 527–555.
- Heine, B. and Kuteva, T. (2005). *Language contact and grammatical change*. Cambridge: Cambridge University Press.
- Johns, M.A., Valdés Kroff, J. and Dussias, P.E. (2019). Mixing things up: How blocking and mixing affect the processing of codemixed sentences. *International Journal of Bilingualism*, 23(2), pp. 584–611.
- Kulkarni-Joshi, S. (2016). Forty years of language contact and change in Kupwar: A critical assessment of the intertranslatability model. *Journal of South Asian Languages and Linguistics*, 3(2), pp. 147–174.
- Labov, W. (1972). *Sociolinguistic patterns*. Oxford: Basil Blackwell.
- LaCasse, D. (2018). *The subjunctive in new Mexican Spanish: Maintenance in the face of language contact*. PhD. Pennsylvania State University.
- Lipski, J. (1978). Code-switching and the problem of bilingual competence. In: M. Paradis, ed., *Aspects of bilingualism*, 1st ed. Columbia, SC: Hornbeam Press, pp. 250–264.
- Matras, Y. (2009). *Language contact*. Cambridge: Cambridge University Press.
- Mettouchi, A. (2008). Kabyle/French codeswitching: A case study. In: M. Lafkioui and V. Brugnatelli, eds., *Berber in contact: Linguistic and sociolinguistic perspectives*. Cologne: Rüdiger Köppe Verlag, pp. 187–198.
- Muysken, P. (2000). *Bilingual speech: A typology of code-mixing*. Cambridge: Cambridge University Press.
- Muysken, P. (2015). Déjà voodoo or new trails ahead? Re-evaluating the mixing typology model. In: R. Torres Cacoullous, N. Dion and A. Lapierre, eds., *Linguistic variation: Confronting fact and theory*, 1st ed. New York: Routledge, pp. 242–261.
- Myers-Scotton, C. (1993). *Duelling languages*. Oxford: Oxford University Press.
- Myslín, M. and Levy, R. (2015). Code-switching and predictability of meaning in discourse. *Language*, 91(4), pp. 871–905.
- Otheguy, R. and Zentella, A.C. (2012). *Spanish in New York: Language contact, dialect levelling, and structural continuity*. Oxford: Oxford University Press.

- Poplack, S. (1980). "Sometimes I'll start a sentence in Spanish *y termino en español*": Toward a typology of code-switching. *Linguistics*, 18(7–8), pp. 581–618.
- Poplack, S. (1993). Variation theory and language contact: Concepts, methods and data. In: D.R. Preston, ed., *American dialect research*, 1st ed. Amsterdam: John Benjamins, pp. 251–286.
- Poplack, S. (2015). Code switching (linguistic). In: J.D. Wright, ed., *International encyclopedia of the social & behavioral sciences*, 2nd ed. Oxford: Elsevier, pp. 918–925.
- Poplack, S. (2018). *Borrowing: Loanwords in the speech community and in the grammar*. Oxford: Oxford University Press.
- Poplack, S. and Dion, N. (2012). Myths and facts about loanword development. *Language Variation and Change*, 24(3), pp. 279–315.
- Poplack, S., Torres Cacoullós, R., Dion, N., de Andrade Berlink, R., Digesto, S., LaCasse, D. and Steuck, J. (2018). Variation and grammaticalization in Romance: A cross-linguistic study of the subjunctive. In: W. Ayres-Bennett and J. Carruthers, eds., *Manuals in linguistics: Romance sociolinguistics*, 1st ed. Berlin: De Gruyter, pp. 213–246.
- Sankoff, D. (1998). A formal production-based explanation of the facts of code-switching. *Bilingualism: Language and Cognition*, 1(1), pp. 39–50.
- Sankoff, D. and Poplack, S. (1981). A formal grammar for code-switching. *Papers in Linguistics: International Journal of Human Communication*, 14(1), pp. 3–45.
- Sankoff, G. and Brown, P. (1976). The origins of syntax in discourse: A case study of Tok Pisin relatives. *Language*, 52(3), pp. 631–666.
- Shenk, P.S. (2006). The interactional and syntactic importance of prosody in Spanish-English bilingual discourse. *International Journal of Bilingualism*, 10(2), pp. 179–205.
- Silva-Corvalán, C. (1994). *Language contact and change: Spanish in Los Angeles*. Oxford: Clarendon Press.
- Sorace, A., Serratrice, L., Fillaci, F. and Baldo, M. (2009). Discourse conditions on subject pronoun realization: Testing the linguistic intuitions of older bilingual children. *Lingua*, 119(3), pp. 460–477.
- Stanford, J. (2008). A sociotonic analysis of Sui dialect contact. *Language Variation and Change*, 20(3), pp. 409–450.
- Steuck, J. and Torres Cacoullós, R. (2019). Complementing in another language: Prosody and code-switching. In: J.A. Villena-Ponsoda, F. Díaz-Montesinos, A.-M. Ávila-Muñoz and M. Vida-Castro, eds., *Language variation – European perspectives VII*. Amsterdam: John Benjamins, pp. 219–231.
- Stoll, S., Zakharko, T., Moran, S., Schikowski, R. and Bickel, B. (2015). Syntactic mixing across generations in an environment of community-wide bilingualism. *Frontiers in Psychology*, 6, p. 82.
- Tadmor, U. (2009). Lexical borrowing: Concepts and issues. In: M. Haspelmath and U. Tadmor, eds., *Loanwords in the world's languages: A comparative handbook*, 1st ed. Berlin: De Gruyter Mouton, pp. 55–75.
- Tamminga, M. (2016). Persistence in phonological and morphological variation. *Language Variation and Change*, 28, pp. 335–356.
- Thomason, S.G. and Kaufman, T. (1988). *Language contact, creolization and genetic linguistics*. Berkeley: University of California Press.
- Torres Cacoullós, R. (2012). Grammaticalization through inherent variability: The development of a progressive in Spanish. *Studies in Language*, 36(1), pp. 73–122.
- Torres Cacoullós, R. and Aaron, J.E. (2003). Bare English-origin nouns in Spanish: Rates, constraints and discourse functions. *Language Variation and Change*, 15(3), pp. 289–328.
- Torres Cacoullós, R. and Travis, C.E. (2011). Testing convergence via code-switching: Priming and the structure of variable subject expression. *International Journal of Bilingualism*, 15(3), pp. 241–267.
- Torres Cacoullós, R. and Travis, C.E. (2018). *Bilingualism in the community: Code-switching and grammars in contact*. Cambridge: Cambridge University Press.
- Torres Cacoullós, R. and Travis, C.E. (2019). Variationist typology: Shared probabilistic constraints across (non-)null subject languages. *Linguistics*, 57(3), pp. 653–692.
- Travis, C.E. (2005). *Discourse markers in Colombian Spanish: A study in polysemy*. Berlin and New York: De Gruyter Mouton.

- Travis, C.E., Torres Cacoullos, R. and Kidd, E. (2017). Cross-language priming: A view from bilingual speech. *Bilingualism: Language and Cognition*, 20(2), pp. 283–298.
- Wei, L. and Milroy, L. (1995). Conversational code-switching in a Chinese community in Britain: A sequential analysis. *Journal of Pragmatics*, 23(3), pp. 281–299.
- Weiner, E.J. and Labov, W. (1983). Constraints on the agentless passive. *Journal of Linguistics*, 19(1), pp. 29–58.
- Weinreich, U. (1953/1968). *Languages in contact: Findings and problems*. The Hague: Mouton.
- Weinreich, U., Labov, W. and Herzog, M.I. (1968). Empirical foundations for a theory of language change. In: W.P. Lehmann and Y. Malkiel, eds., *Directions for historical linguistics: A symposium*, 1st ed. Austin, TX: University of Texas Press, pp. 95–188.

Appendix: transcription conventions (Du Bois et al., 1993)

| | | | |
|-----------------|---------------------------------|-------|------------------------------------|
| Carriage return | new Intonation Unit | .. | short pause (0.5 secs) |
| . | final intonation contour | ... | medium pause (0.5–0.7 secs) |
| , | continuing intonation contour | ...() | timed pause (longer than 0.7 secs) |
| ? | appeal intonation contour | = | lengthened syllable |
| -- | truncated intonation contour | ~ | pseudonymised proper noun |
| @ | one syllable of laughter | (()) | researcher's comment |
| @word | speech produced while laughing! | | pronounced with notably high pitch |

Acknowledgements

We gratefully acknowledge support from the National Science Foundation (BCS-1019112/1019122, 1624966; <http://nmcodeswitching.la.psu.edu>) and the ARC Centre of Excellence for the Dynamics of Language. We thank the editors of the volume for their comments.

Convergence

Björn Wiemer

1. Introduction

Convergence is a counterpart to variation as encountered in any kind of speech community. In sociolinguistics and communicative theories, convergence is related to accommodation between speakers in face-to-face encounters (Giles, Coupland and Coupland, 1991). If speaker communities get ‘mixed up’ (for whatever reason), this, as a rule, increases variation and diversity, from which even new varieties may emerge. Subsequently, variation (often seemingly random) can either diminish or acquire more clear-cut distribution (see Section 3.3). This implies convergence, and standardization is only one of its many facets. Variationist linguistics (Poplack, this volume) provides tools for studying details of convergence (although the results and positions taken by particular researchers differ). Moreover, convergence bears a clear relation to contact linguistics, where it means an increase in similarity between varieties in contact.

The term ‘variety’ (or ‘lect’) comprises entire languages (or dialects) and their particular varieties down to idiolects. ‘Contact,’ in turn, can apply at varying degrees and in different dimensions, but usually it concerns speakers of rather comprehensive groups. Furthermore, convergence may refer either to varieties or to structural patterns. Henceforth, I will use ‘variety’ as a maximally broad umbrella term for any sort of ‘lect,’ without any principled difference between ‘dialect,’¹ ‘language,’ etc. Moreover, I will not particularly distinguish between bi- and multilingualism, unless explicitly stated.

Therefore, convergence is a very general notion² which implies two premises to be clarified. The first premise concerns the typology and range of phenomena: in which dimension, or environment, does convergence occur? Is it a societal frame (e.g., a large multilingual speech community or a small isolated village), is it a macro-areal pattern (as known from typological studies; see Section 2.5), is it a diastratic contrast (standard language vs. dialects vs. superregional koiné), or a combination of both in a middle-sized area (e.g., the Balkans, the Caucasus, or EMSEA languages; for the latter cf. Bisang, 2006, 2015)? More generally, are we dealing with ephemeral phenomena, or do repetitive transfers (replications) of patterns, or choice of variants, turn into socially firmly entrenched usage (see Sections 2.3, 2.4, 4.1)?

The second premise has to do with the reasons why convergence becomes an issue at all: the very notion of convergence presupposes that there is some background on which a particular feature (or feature cluster) or some kind of linguistic behaviour becomes salient. With respect to some areal (diatopic), diastratic,³ or other sociolinguistic dimension, features of some areally close varieties strike the eye, because they deviate from some kind of norm, or standard of comparison, that surrounds them. Convergence thus presupposes a figure-ground contrast, both in terms of expectations (predictability) of how language works in actual communication and in geographic terms (for the latter cf. Wiemer, 2004, 2019).

Thus, the main disciplines dealing, in one way or another, with convergence are dialect geography, areal linguistics (typology), but also sociolinguistics inasmuch as the rise of super-regional or urban varieties, dialect levelling, and the like are at stake. The difference between dialect geography and areal typology is arguably one of geographic size and of relative genealogical distance (see Sections 2.1, 2.5, 5.1), while the difference between dialectology and sociolinguistics (called ‘urban dialectology’ by many Anglo-Saxon scholars, e.g., Chambers and Trudgill, 1980/1998; cf. Barni and Bagna, this volume) is rather one of tradition (Wiemer and Seržant, 2014, pp. 14–15).

As a concomitant to the figure-ground contrast, convergence implies diachronic change: some varieties have become more similar to each other (while dissimilating from other varieties) than they used to be before. Convergence has been pointed out mainly for situations in which speaker communities of some varieties V_{1-i} of a group of related languages L_A have come into contact with speaker communities of varieties V_{1-i} of another language group L_B (+ $L_C \dots L_I$). If, then, $L_A, L_B (\dots L_I)$ have originally been sufficiently dissimilar from each other (in comparison to yet other language groups) and currently their varieties V_{1-i} are more similar to each other than to varieties V_{j-n} of groups $L_A, L_B (\dots L_C)$, respectively, the reasons for the observed convergence appear to be evident.

‘Divergence’ usually is seen as the converse of convergence. There are, however, instances when this converse relation does not apply, or when it depends on the perspective taken (see Section 3.1). There seems to be consensus that both convergence and divergence are epiphenomena and that, trivially (resulting from the first premise), one always has to specify at which level of a variety’s structure or regarding which habits of a speech community convergence/divergence occurs. (See the classification in Höder, 2014b, pp. 41–43.)

2. Historical overview

Historical-comparative linguistics and traditional dialectology, with their roots in the eighteenth to nineteenth centuries, paid attention primarily to divergence, since their programme aimed at the areal (= diatopic) manifestation of an increase of differentiation within earlier dialect continua and the disclosure of an assumed original state of an idealized continuum. This seems to have changed slightly with Schmidt’s wave theory (Schmidt, 1872) and with Schuchardt’s assumption that centripetal forces lead to levelling (Germ. *Ausgleich*; Schuchardt (or no split at all), 1885). The modern understanding of ‘diffusion’ (in dialectology and areal linguistics) rather corresponds to Schuchardt’s centrifugal forces, but these equally imply convergence if neighbouring areas are taken into account (see Sections 2.1, 2.5). Similar considerations hold true for the rise of koinés or regiolects (see Section 2.2), which result from the mixing, or nivellation, of dialect differences and lead to compromise varieties of superregional value (Siegel, 1985). These however rather belong to the diastratic dimension, as do *linguae francae* and creoles; cf. Hinskens, Auer and Kerswill (2005, pp. 3–15) and Mufwene (this volume).

Of course, diastatic and diatopic variation can, and usually do, interfere and lead to levelling (i.e., convergence).

2.1 *Transitional dialects and secondary convergence*

In Slavic dialectology, convergence was implicit as an overarching feature in the rise of transitional dialects (following Durnovo, Sokolov and Ušakov, 1915; Małecki, 2004 [1934]). Presumably, these have arisen from a slow process whereby subdialects of larger adjacent dialect zones A and B became similar to each other to an extent that they could no longer be clearly assigned to either zone A or B. As a main structural property of transitional dialects, it has been pointed out that features do not mix haphazardly in speech, but are shaped uniformly either according to dialects from group A or from group B; this ‘mixture’ brings about stable hybrids (compare ‘fudging’ used in Anglo-Saxon sociolinguistics; Hinskens, Auer and Kerswill, 2005, p. 10).⁴ Dialectologists have restricted these processes to closely related dialects (e.g., East vs. West Slavic), but they sometimes do occur even across such boundaries, e.g., in Rusyn. This East Slavic dialect group (its closest ‘relative’ is Ukrainian) is spoken in the Carpathian region and scattered under different roofs, among them Slovak and Polish (both belong to West Slavic). The frequency of various realizations of the BE-auxiliary (clitic, affix, lack) employed for marking the past tense resembles the different realizations in the contact roof varieties regardless of closeness; cf. Rabus (2019), similarly Danylenko (2012) on West Ukrainian dialects.

In a more global perspective, transitional zones have been characterized as being based not on ‘the bundling of unique features emanating from a discernible geographic core, but the presence of multiple overlaps from the neighbouring regions’ (Haig, 2017, p. 397). However, they also bear features which cannot be explained just by diffusion from the surrounding areas, but ‘are specific compromise responses to the conflicting typological profiles of the neighbouring languages’ (2017, p. 398).

Regardless of their size, transitional zones have been conceived of as salient areas sandwiched between two sufficiently similar (and larger) dialects (or languages). This reasoning implies that two larger areas (A, B) must have formed as earlier results of dissimilation (thus, divergence) within an even more ancient and homogeneous dialect/language continuum. The chronological relation can be schematized as follows:

1. (i) dialect continuum
 - > (ii) increasing divergence: split-up into salient (sufficiently dissimilar) dialects A, B. . . (= loss of homogeneity), often accompanied (or conditioned) by topographic or political-administratory borders
 - > (iii) (secondary) convergence: mutual assimilation of subdialects from adjacent parts of larger dialect areas A, B. . .
 - > (iv) rise of transitional dialects sufficiently distinct from dialects A, B. . .

Examples of such transitional zones are the border regions between Belarus and Ukraine (Polissja), between south-western Silesia (Polish) and northern Moravia (Czech), or the broad transitional zone between the two main (north-eastern vs. western) dialects of Belarusian. An example from areal linguistics is East Anatolia (Haig, 2017).

Secondary convergence, as modelled in (1), has also been implicit to scenarios that mitigate between assumed genealogical affiliation and relatedness via intense contact. Known

cases are some of the hypotheses advanced to explain the rise of the ‘Balto-Slavic unity’ (cf. Birnbaum, 1970; Holzer, 1998), as well as the formation of the Sino-Tibetan family (LaPolla, 2003). However, the time frame to be needed for convergence within an area has been debated (Nichols, 1992; Watkins, 2003).

2.2. *Koineisation and diastratic levelling*

While the concept of transitional dialects implies secondary convergence in a (predominantly) diatopical dimension, koineisation is a prime example of how diatopic and diastratic convergence co-occur. Koinés represent superregional standards and usually arise in reaction to needs and habits in administration and trade, but sometimes also for reasons of religious unity or of national revival. Concomitantly, koineisation enhances distance in a diastratic dimension and may lead to diglossic situations (i.e., an extreme type of diastratic divergence). Examples are Czech, Norwegian (*bokmål, nynorsk*), classical and other varieties of Arabic (Versteegh, 2014², pp. 60, 182–184), Classical Greek (Horrocks, 1997). Bubenik (2001) and Tuten (2001) supply critical surveys of terms and processes related to koiné(isation), Kouwenberg (2001) and Mufwene (this volume) discuss the relation to creoles.

A good example of koineisation is provided by the history of Czech. In the mid-nineteenth century, a superregional standard variety of Czech was created (rather anachronistically) on the basis of heavy grammatical borrowing from Renaissance Czech (fourteenth to fifteenth centuries). The consequence was a split between newly created standard Czech (*spisovná čeština*) and the spoken language of the cities, which, in turn, led to a superregional koiné based mainly on Central Bohemian dialects (*obecná čeština*). These two varieties entered into a diglossic relationship, and both of them were opposed to dialects; this is a clear case of divergence. However, subsequently, apart from dialectal differentiation successively levelling out, the border between *spisovná čeština* (= high variety) and *obecná čeština* (= low variety) started getting shaky, and it presently has more or less turned into register (i.e., diaphasic) variation which incorporates elements of both varieties. As a consequence of this convergence, diglossia has vanished (Bermel, 2000).

Therefore, mergers of patterns from high and low variety can lead to a new standard variety. This happened to the predecessors of modern Russian from the sixteenth up to the end of the eighteenth century (Keipert, 1999; Uspenskij,³ 2002; Rabus, 2014). Modern Standard Russian has incorporated many Church Slavonic (= South Slavic) elements, not only in the lexicon (e.g., many parallels of cognate stems with a clear-cut meaning distribution between concrete and abstract, e.g., *rovny* (genuinely East Slavic) ‘even (surface)’ vs. *равный* (South Slavic) ‘equal (before the law)’), but even in patterns of derivation (e.g., present active participles, morphological alternations in verb stems of certain aspect pairs with reflexes of **dj*, e.g., *ubed-i-t*^{PFV} – *ubežd-a-t*^{IPFV} ‘convince,’ as in Bulgarian, but *opered-i-t*^{PFV} – *operež-a-t*^{IPFV} ‘precede, outstrip,’ genuinely East Slavic).

Convergence which levels out both diastratic and diatopic variation appears to be rather typical for vernaculars with a sufficient degree of structural affinity, usually conditioned by genealogical closeness. Paradigm examples are the rise of superregional, mostly urban vernaculars (i.e., varieties acquired as L1) in the East Slavic domain, like Russian *prostorečie*, or Russian-Belarusian (*trasjanka*) and Russian-Ukrainian (*suržyk*) ‘mixtures’ (see Section 3.3). Remarkable is the high degree of convergence of derivational patterns in (meanwhile extinct) Lithuanian insular dialects in Belarus: intense contact and asymmetric bilingualism with East

Slavic speakers led not only to a frequency increase in the application of patterns of stem derivation inherited by both Slavic and Baltic, but even to morphological reanalysis of borrowed (Slavic) derivational suffixes and their productive application to inherited (Baltic) verb stems (Wiemer, 2009).

2.3 Sociolinguistics and discourse pragmatics

Concepts related to convergence have been employed in sociolinguistics and discourse pragmatics as well, partly adopted from social psychology and with a differentiation into short-term and long-term phenomena. Short-term phenomena are often discussed as ‘accommodation,’ which occurs in face-to-face encounters, for instance in child-parent interaction, or in mutual adaptation between speakers of different dialects (compare, e.g., the *Synchronisierungstheorie* in Schmidt and Herrgen, 2011), which often is connected to conscious choices at different levels (Hinskens, Auer and Kerswill, 2005, pp. 5–7). Tiny adaptational steps may add up to more long-lasting changes, as was noted already by Paul (1968 [1880], p. 23). To some extent, this line of thinking has been revived in connection with Matras’ notion of pivot-matching (Matras, 2011, p. 149; see Section 4.1).

2.4 Identification models and metatypy

Since Weinreich (1953) several approaches have appeared which share the following assumption (henceforth: M = model language, R = replica language):

2. A polysemous form or construction *a* in M is replicated in R by first identifying it with a form *b* in R in one function α and then extending the use of *b* to one or several other functions β of *a* where *b* has not been used before (at least not as the major use pattern). (Wiemer and Wälchli, 2012, p. 27)

We may call them identification models (see Section 4.1). Such identification may lead to isomorphism, i.e., a blueprint for specific structures whose slots can be filled with material from several varieties which are related at different degrees. If such blueprints accumulate in some region, because changes are mutually reinforced, isogrammatism results (e.g., on the Balkans; Gołąb, 1990; Lindstedt, 2000; Friedman, this volume).

A related phenomenon is metatypy. The term was coined by Ross (1996, 2001, 2007, 2008),⁵ who observed complex contact situations in New Guinea. Metatypy amounts to ‘the wholesale restructuring of a language’s semantic and syntactic structures as a result of language contact’ (Heine and Kuteva, 2005, p. 180). This process can affect all levels of linguistic structure, from phonology to gender systems, change from left-headed to right-headed syntax and clause-chaining devices (Foley, 2013); the result is isomorphism, which favours easy intertranslatability. Ross (2008, pp. 153, 158) observed that metatypy proceeds from larger to smaller structures (clause combining > clausal syntax), however NP-internal changes precede word order changes. At any rate, metatypy relies on the calquing of grammatical contrasts; it often involves massive loan translations, but no re-lexification or MAT-borrowing (see Section 4.1). This makes metatypy diametrically opposed to processes known for creoles and mixed lects (Mufwene, this volume; O’Shannessy, this volume).

Metatypy can reshape the typological profile of R, whereas M remains basically unaffected. That is, at least in the cases described by Ross, speakers of R adapted to M, but not vice versa. Here one should be careful not to apply this notion just for situations of clearly asymmetric

bilingualism, e.g., if we are dealing with the progressive decay of a minority non-standard variety (= R) under the pressure of a standard, or otherwise dominant variety, of another language (= M) (Tosco, 2008, p. 116). This was the case with Arvanitika (= R), i.e., Albanian in contact with Greek (= M); cf. Sasse (1985). However, metatypy was also applied to much more complex contact scenarios, with mutual direction of influence, more languages, and much less obvious identification of Ms and Rs. In fact, many researchers conceived of metatypy as extreme cases of linguistic areas, e.g., the Vaupés region (NW Amazonia; Aikhenvald, 2002) or the Balkans (Lindstedt, 2000, pp. 241–243).

Moreover, Foley points out that metatypy can blur original genealogical distinctions: ‘The overall superficial typological similarity of many of the Papuan languages throughout the region and the even greater similarities between unrelated languages in many sub-areas within are plausibly the result of long-term grammatical metatypy.’ (2013, p. 803) However, Matras (2011, p. 144) cautions that metatypy should not be taken as a feature to define linguistic areas (see Section 2.5), since metatypy involves only PAT-borrowing. Thus, there are also cases of extreme asymmetric convergence which combine massive PAT- and MAT-borrowing and lead to a change of typological profile, such as Nahuatl (Olko, Borges and Sullivan, 2018).

2.5 *Convergence in typology*

Studies such as Dryer (1989) and Nichols (1992) have brought to light that typological key features – such as causative-anticausative prominence, word-order patterns, head – dependent marking, accusative/ergative alignment and splits – are distributed unevenly on a (sub)continental scale (cf. also Meakins, this volume). This kind of macro-areal convergence (or clustering) has been confirmed for middle-sized areas, e.g., Europe (Haspelmath, 2001). However, one practically always finds pockets, or intersections with other areas, which show that areas looking homogeneous from a bird’s eye perspective become increasingly heterogeneous if one zooms in. Moreover, typological studies, as a rule, neglect diastatic differentiation, which sometimes is orthogonal to diatopic convergence.

Apart from the ‘zooming effect,’ one has to be careful in relating a synchronic state to its diachronic background, in particular whether macro-areal effects of convergence arose from one or more hotbeds (before they might have ‘coalesced’). For instance, infinitive loss (with its consequences for clause linkage) is a typical feature of both the Balkans and the Near East; if we did not know details of diachronic development, a synchronic macro-areal perspective would suggest this to be one area (Stolz, 2002). For analogous studies aided by distance measures see Section 4.3. The problem is always whether a cluster grew from one centre to the outside or developed as the result of multiple overlap of different continua (or smaller clusters), i.e., as a contact superposition zone (Koptjevskaja-Tamm and Wälchli, 2001). This makes obvious a parallel to dialect geography: do dialectal continua arise just from ‘centripetal waves,’ or do these interfere with waves from the periphery – where contact with more distant varieties is more probable – so that secondary convergence, or transitional, zones arise (see Section 2.1)?

Nichols (1992) introduced a distinction between spread and residual zones. Spread zones are comparatively large – usually with few topographic obstacles – and characterized by low genealogical density; they arise (and can vanish) quickly and are created by one dominant group of speakers who impose their variety on others. Residual zones, on the contrary, demonstrate high genealogical density and are often situated in less accessible regions (high mountains, swamps, etc.); there is no dominant group, and the age of the varieties spoken at such a place is older than in spread zones. A typical example of a spread zone is the Ukrainian steppe,

while the Caucasus is a paradigm example of a residual zone. Both spread and residual zones show convergence (in comparison to a larger surrounding), but for different reasons. Transitional zones are neither clear spread nor clear residual zones, in fact they sometimes have been treated as being located between areas considered to be more salient (e.g., East Anatolia; Haig, 2017). The wit is that this could make a transitional zone an area of itself.

After all, the question remains what defines an area. The label ‘convergence area’ was used by Emeneau (1956) and Weinreich (1958), the term *Sprachbund* (Language League, Language Alliance) has been used for cases like the Balkans (cf. Kouwenberg, 2001; Heine and Kuteva, 2005, pp. 173–178; and Friedman, this volume, for surveys). However, these terms hardly have any intrinsic value (Stolz, 2002; Campbell, 2006; Matras, 2011, p. 145f.). The ideal would amount to establishing areas exclusively on the basis of linguistic features – with ‘subtracted’ genealogical inheritance and typologically common features – and without knowledge of historical circumstances; however, this can hardly ever be achieved (Tosco, 2008, p. 115). There is no method to establish areas as distinct entities from a random sample of languages, or on the basis of random features (Bickel and Nichols, 2006; Bisang, 2010). Thus, ultimately, areas are established by linguists (Matras, 2009, p. 274; Nau, 2012, p. 465) on the basis of any sort of features (and their combinations) that are considered salient on some areal or genealogical backdrop. This includes prosody, but also lexicalization patterns, i.e., patterns of how concepts are lexicalized, e.g., by employing metonymic or metaphoric shifts. Compare, for instance, the Lower Volta Basin (Ameka, 2006, pp. 137–139): ‘receive+eat/hear’ → ‘believe,’ ‘see/look+way’ → ‘expect’; or Australian languages, which consistently derive cognitive verbs (‘think,’ ‘know,’ ‘believe,’ etc.) from ‘hear’ (not from ‘see,’ as in Europe). Cf. Evans and Wilkins (2000) and also Koptjevskaja-Tamm and Schapper (this volume).

As a compromise between rigid conditions (Tosco, 2008) and any loose sense of the notion, linguistic areas represent cases of convergence for which ‘the density of shared isoglosses in a multiplicity of languages’ catches the eye (Matras, 2011, p. 157); this corresponds to the figure-ground consideration (see Section 1). After all, rather broad definitions like (3) seem to be accepted:

3. A linguistic area can be recognized when a number of geographically contiguous languages share structural features which cannot be due to retention from a common proto-language and which give these languages a profile that makes them stand out among the surrounding languages.
(Haspelmath, 2001, p. 1492)

This quote implies that there are features which make the target area differ from an *immediate* surrounding. Since this need not necessarily be features compared on a continent- or worldwide scale (and they may be much more fine-grained and language-specific than a worldwide comparison would allow), this principle can apply several times for increasing (or decreasing) geographical size (or number of languages). What we then get is the *matrěški*-principle (Wiemer, 2004, 2019).

Haspelmath’s definition in (3) furthermore requires that we can tell apart genealogical heritage from contact-induced changes and from development conditioned by universal tendencies. However, very often this cannot be done, e.g., since time-depth of data is insufficient. The question ‘What happened?’ cannot be answered also for lack of data (or reliable reconstructions) on the sociolinguistic conditions of convergence, in particular no single source (‘initiator’) can be identified (Lindstedt, 2000; Nau, 2012, among others). Consequently, in many (most?) cases of areally outstanding clusters, research is restricted to the circumstantialist

position (Campbell, 2006). This position, however, is required by a rigid procedure of establishing areas (Tosco, 2008).

3. Critical issues and topics

In principle, among issues and methods of research we can distinguish those which are concerned primarily with the mechanisms responsible for convergence, and those which focus on the social (including political, historical) circumstances which are favourable or disfavourable for convergence. For reasons of space, I will treat these aspects jointly.

Convergence ultimately originates in individual, often *ad hoc*, or nonce, innovations of multilinguals, who rely on their creativity (cf. Matras, 2009; see Section 4.1.3). But, as for any change to become palpable, such innovations have to spread among a speech community – also among those who do not share the same languages with those multilinguals who triggered the erstwhile change – before they can eventually be entrenched, as a rule regardless of multilingual speech habits (Bower, 2013, p. 421; Höder, 2014b, p. 44). Accordingly, in research related to convergence we observe different foci on either the creativity (or cognitive abilities) of speakers bringing in innovations, or on the role of social factors favouring (or disfavouring) the spread (i.e., diffusion) of innovations, or of minor usage patterns.

Silva-Corvalán (2008) presents an account of both cognitive and social dimensions.

3.1 Factors: social psychology vs. cognition

Attitudinal factors (emphasized, e.g., by Hinskens, Auer and Kerswill, 2005, pp. 39–41) and a basically cognitive motivation of convergence (emphasized, e.g., by Matras, 2009) do not represent a contradiction, but complementary aspects of a complex phenomenon. Emphasis on cognitive motivation means that multilingual speakers tend toward building a unified diasystem for the varieties they use. This thought was brought up by Weinreich (1953); cf. also Breu (2012, pp. 301–310, 2018). The formation of a diasystem, on the one hand, can be regarded as simplification, or elimination of contrasts between two (or more) systems, in the sense that organizational effort is minimized, but the systems are held structurally autonomous. On the other hand, the overall system is managed even in cases of communicative and cognitive pressure; this seems to explain implicational hierarchies of borrowings (see Section 4.2 and Matras, this volume).

Emphasis on sociopsychological factors, on the contrary, often implies an increase in complexity, inasmuch as speakers let themselves be guided by (usually subconscious) considerations of solidarity, or dissociation, with relevant speaker communities. This bars accommodation (which may nonetheless occur on a subconscious level averaged over many speakers); cf. Yakpo (this volume). Hinskens, Auer and Kerswill (2005, p. 40) argue that sociopsychological factors might play a more important role if the involved varieties are related closely (e.g., neighbouring dialects of the same roof language), while cognitive motives favouring convergence become more prominent if the varieties are more distant from each other in terms of structural and/or genealogical affinity. Moreover, one may hypothesize that the relative prominence of one or the other type of motives depends on the way speakers have grown multilectal: compound (i.e., simultaneously acquired) multilectal competence may give more weight to cognitive factors, whereas co-ordinate (i.e., successively acquired) multilectal competence favours attitudinal factors. In addition, the proportions between mono- and multilectal speakers per community may influence the equilibrium between cognitive and sociopsychological factors (in-group loyalty etc.) as well.

Such an equilibrium is assumed also in recent work carried out in Diasystematic Construction Grammar (Höder, 2014a). Here, pro-diasystematic change favours convergence. In contrast to more ‘classical,’ rather structuralistic models inspired by Weinreich, a diasystem is no longer conceived of as the interaction of ‘a monolingual system in contact with a different – and separate – system,’ but as ‘a change *within one multilingual system*’ (Höder, 2014b, p. 58, emphasis original). Thus, pro-diasystematic change leads to ‘a reduction in the overall number of idiosyncrasies and a corresponding increase in the number of diasystematic elements across the relevant speaker group,’ i.e., of language-unspecific elements, while the complexity of the individual language systems involved may in principle remain unchanged (2014b, p. 46). Diasystematicity – as located in the minds of multilinguals – distinguishes between common elements in a kind of intersection between the involved varieties and leaves space for specific elements of each involved variety. Pro-diasystematic change may even superficially lead to divergence, e.g., in the phonetic realization of diphthongs in the speech of bidialectal Low German–High German speakers: as a consequence of frequent correspondences between Low and (local) High German phonetic realizations of cognate word forms, these correspondences are regularized (and, thus, converge) even for original ‘interlectal homonyms’; e.g., [‘a:baɪt] ‘work’ originally sounded the same in both varieties, but now bidialectal speakers pronounce [‘a:beɪt] in their Low German speech. The varieties superficially diverge, but from the point of view of underlying rules they converge (Höder, 2014b, pp. 56–57).

3.2 *Correlations between types of convergence and contact conditions*

The substrate–adstrate (or superstrate) distinction is, by itself, unrelated to relative sociolinguistic dominance, but it influences the outcome of interference. A substrate presupposes a complete language shift and the disappearance of an erstwhile Lx (usually by abandoning Lx as L1 in the raising of children). In (4a) the difference between Ly and Ly’ makes the substrate, which ‘derives’ from the extinct Lx. By contrast, an adstrate is supplied by a language (i.e., speaker group) that persists, but influences another language; see (4b):

- 4a. substrate: $Lx \rightarrow Ly \quad > \quad \emptyset_{Lx} Ly'$
 4b. adstrate: $Lx + Ly \quad > \quad Lx + Ly'$

In the adstrate case, depending on the proportions between speakers of Lx and Ly, the percentage of bilinguals on each side, and on many other possible social factors, Lx may undergo changes as well (Lx’). But the crucial point is that both Lx and Ly persist. For instance, the loss of morphological cases in Balkan Slavic can be explained as the result of an incomplete acquisition of Slavic by originally Balkan Romance speakers (= substrate), whereas the retention of practically the entire paradigm of morphological cases in Molise Slavic (despite massive Italian influence for about 400 years under conditions of asymmetric bilingualism) becomes understandable from the fact that its speakers, though adapting to Italian on many levels, have continued using the two languages in parallel (= adstrate); cf. Breu (1994, pp. 45–48, 2003, 2012).

On an average, the distinction between substrate and adstrate is easier to make if the involved varieties are genealogically distant; it becomes blurred not only in the levelling of dialects, but also of diglossic situations: the low variety is subject to an influx of loanwords and structures from the high variety (see Section 2.2). Bilingualism is unidirectional inasmuch as hardly anyone learns the low variety (Matras, 2009, p. 237), but the low variety is the L1-vernacular which imbibes more and more features of the high variety. Thus, what results

is not one altered variety (Ly') at the expense of some vanished Lx, but rather some kind of merger of Lx and Ly, and bilingualism disappears. See Section 3.3 for parallels in new dialect formation.

In otherwise similar settings, multilingualism does not disappear but becomes extremely complex. This happens if there is no dominant variety or *lingua franca*, but rather some very local koinés (endowed with a certain prestige), and if we observe compromises between intense inter-group contacts and strong in-group identities (Lindstedt, 2000, p. 239 on the Balkans). Lindstedt (2000, p. 231) comments that 'the origins of most grammatical Balkanisms are not to be sought in the internal development of any one of these languages, but rather in the multilingual contact situation itself.' Concomitantly, MAT borrowings (see Section 4.1) are exceptional. Moreover, '[a] strong second language of a bilingual individual may and does influence the first language even when it is used in a monolingual setting' (2000, p. 240). Noticeably, no language shift occurs; consequently, there are no substrates. Concomitantly, there is no possibility (and no sense) of identifying model and replica language (see Section 4.1).

In their often-cited study on the multilingual village Kupwar (India), Gumperz and Wilson (1971) concluded that conditions comparable to those in the Balkan core region had led to a high degree of intertranslatability, or even metatypy (see Section 2.4). However, Kulkarni-Joshi (2016) shows this result to be largely an artefact caused by the methodology which targeted speech varieties of forced bilingual interaction. More diversified speech samples and a more comprehensive comparison with historic sources and neighbouring villages 'revealed shared patterns of variation' (2016, p. 170). These patterns rather indicate complexification, although their parallels might be indicative of convergence. Thus, apart from methodological caveats, the lesson is that variation (for some features) may become more complex, the patterns underlying this variation can nonetheless converge and even remain stable over some area and time.

3.3 *New dialect formation*

Various non-standard varieties develop 'when there is a mixture of dialects leading to a single new dialect which is different from all inputs' (Hickey, 2003, p. 214). The conditions and mechanisms of the rise of New Englishes are probably best studied. Research on colonial Englishes has focused mainly on the sociolinguistic conditions which favour convergence. Levelling of dialects under conditions of long-distance migration and colonization can be considered as a specific, namely 'immigrant' type of koineisation (Siegel, 2001). Thus, Trudgill (2006, p. 162) generalizes that 'relatively weak social network ties' supply 'the breeding ground for rapid supralocal linguistic change.' His prime example is New Zealand and Australian English, but the same conclusion seems to apply as (part of) an explanation for the rise of convergent features in areas of intense and long-lasting contact between comparatively unrelated varieties. A good example is, again, the core region of the Balkans, where Arumanian and Meglenorumanian shepherds constantly migrating from place to place have been considered the probably most important part of a multilingual (or multidialectal) population responsible for diffusion.⁶ Similarly, Aikhenvald (2002, p. 13) emphasizes that the diffusion of patterns (and of their replication) occurs predominantly in multilingual communities, while one-to-one language contact rather leads to a simple levelling of structure. That is, a constellation where different communities intersect creates effects of family resemblance with overall weak ties, and this, in turn, favours the appearance of *new* areal patterns (not identical to any source). In general, network structures on a mesolectal level⁷ have been identified as crucial for diffusion of innovations and decrease of variation (Hinskens, Auer and Kerswill,

2005, pp. 37–38). These considerations possibly prove adequate regardless of the genealogical closeness of the involved varieties and of the size of the area (see Section 5.1), and they have been made responsible for (de)creolisation (Hinskens, Auer and Kerswill, 2005, pp. 13–14).

As for New Zealand English, Trudgill (2006) concluded that the following stages (and processes) characterize the formation (via convergence) of a new sufficiently distinct and uniform variety of English: (1) rudimentary levelling (via face-to-face communication) and imperfect (or partial) accommodation among adult speakers (immigrants from the British Isles) > (2) seemingly random variation among the first New Zealand-born generation > (3) koineisation among the following generations (consisting of levelling, unmarking, and reallocation) > (4) focusing process.

It is intriguing that during stage (2) the choice of variants appeared to be random only if idiolects were considered; however, if investigated collectively, i.e., for a larger portion of the speaker community, variation turned out to be rather well-structured, as it reflected ‘the proportions of such variants in the dialect mixture’ around the idiolects (2006, p. 159). Remarkably, similar observations have been pointed out for the rise of a superregional mixed Belarusian-Russian variety, sometimes called *trasjanka* (see Section 2.2), notwithstanding different socio-political conditions. Despite an allegedly random distribution of Russian vs. Belarusian realizations of features in Belarusian cities and much individual variation (even in idiolects), highly significant and regular hierarchies apply for the proportions of choices between tokens with a ‘Russian’ or ‘Belarusian’ provenance. This applies to speaker populations with different preferences for Russian or Belarusian (or their ‘mixing’). These hierarchies are stronger for phonemic features than for autonomous lexical words (with function words occupying an intermediate position); cf. Hentschel (2013). Therefore, convergence can be disclosed (with statistical methods) also for larger (and heterogeneous) feature bundles even if the respective varieties are very closely related (see Sections 5.1, 5.2).

4. Current contributions and main research methods

Apart from obvious cases, language contact can work as a catalyst of convergence on levels that seem less spectacular and evident. This applies not only to convergence between varieties that have anyway been very close in areal and/or genealogical terms from the start, such as the levelling of dialect continua or the rise of urban koinés. We should also be aware that macro-areal patterns (e.g., along the Pacific Rim, or in Eurasia) can only be explained if we account for them as results of long-term rapprochements between varieties that were dissimilar in areal and/or genealogical and, thus, also in structural terms; rapprochements occurred on many different local levels (regions) and summed up to a chain of family resemblances over a much larger area. In order to disclose such patterns, one must not only look at spectacular phenomena, and one has to account for distributional properties and their change (e.g., if minor patterns increase and gain in productivity).

4.1 Mechanisms of convergence: identification models

As concerns specific phenomena of convergence (and associated processes), several theoretical frameworks are on the market. All of them contrast model and replica languages (M and R, respectively),⁸ and all of them more or less concentrate on the moment of innovation, not on propagation (see Section 3):

4.1.1 Code Copying

This approach (Johanson, 2002) distinguishes the phonological ('material') shape and semantic aspects as well as combinatorial and frequential properties (the latter two probably tie up closely). Global copies are those which include all four aspects, copies are selective if only some of them apply.

4.1.2 PAT vs. MAT borrowing

PAT(tern) borrowing applies if R replicates a structure and/or function(s) from M without the phonological shape of items in M, while MAT(ter) borrowing includes the phonological shape (Matras and Sakel, 2007; Matras, 2009). This opposition is roughly identical to the more traditional difference between loan/borrowing and calque (cf. Matras, 2013, p. 68 for some terminological overview); it also considerably overlaps with the distinction between global and selective copies in Johanson's framework (see Section 4.1.1).

4.1.3 Pivot-matching

The mechanism of pivot-matching (Matras and Sakel, 2007; Matras, 2009, pp. 240–243, 2011, 2013, pp. 70–72) is subordinate to PAT-borrowing inasmuch as 'pivotal features of the model construction are identified and replicated in the replica language' (Matras, 2009, p. 241); 'there is no one-to-one correspondence between the morphemes of the equivalent lexemes'; each pattern is created in its own self-contained system and selected according to communicative appropriateness, only the pivot is shared (Matras, 2009, p. 247, 2011, p. 151). Pivot-matching is itself insensitive to the substrate-adstrate difference (Matras, 2009, p. 258). This model, which is designed to comprise also contact-induced grammaticalization (see Section 4.1.4), allows for abrupt (not only gradual) changes, but simply because it focuses on the moment of creative invention and is agnostic as to whether the innovation further spreads, in particular grammaticalizes, or not. However, Matras (2011, p. 157f.) suggests some socio-linguistic conditions on which the spread of pivot-matched constructions depends, together with a chain-effect 'across an entire network of contiguous or partly contiguous languages.' Most crucial are 'lax normative attitudes in a multilingual community with flexible identity boundaries' (Matras, 2013, p. 72).

4.1.4 Contact-induced grammaticalization (CiGxn)

Essential for this line of research is the formulation of constraints on the directionality of changes from lexical to grammatical expression. Contact is recognized as a trigger of grammaticalization (Matras, 2009, p. 238), but the processes are basically the same as in 'internal grammaticalization.' Initially, Heine and Kuteva (2005) made a difference between Ordinary and Replica CiGxn: in the latter case, it was not only a pattern, or pivot, that was borrowed, but even the entire grammaticalization process was assumed to be replicated by speakers of R. This idea was abandoned in Kuteva and Heine (2012), where the authors presented an integrative model of CiGxn by splitting the processes in two phases (propelling vs. accelerating forces).

Importantly, a more general model of pattern replication should account for cases which (at least superficially) violate unidirectionality, as assumed for grammaticalization, e.g., if categories are lost prior to a theoretical end of a grammaticalization process (Matras, 2013, p. 71).

4.1.5 Polysemy copying

A shift in meaning for some unit in R is inspired by M, but without a change in the grammatical status of the item in question. Usually, this leads to an extension of the lexical meaning (as in loan translations) or of the functional inventory of the respective unit. Note that the logical counterpart, namely monosemization in R on the basis of equivalents in M, occurs as well, e.g., in Molise Slavic (Breu, 2003, pp. 354–363).

4.1.6 Interlingual identification of linguistic subsystems

Interlingual identification of linguistic subsystems (both as process and result) includes signs and/or categories (Gast and van der Auwera, 2012). It is essentially a refined version of polysemy copying but was designed to cover different types of CiGxn as well (2012, pp. 389–390). See Breu (2018: §2) for a (somewhat implicit) application.

4.1.7 Comparison of approaches

The last three of the aforementioned mechanisms share (at least implicitly) assumptions known from semantic maps. Although the entities that are identified, copied, or assimilated differ in terms of complexity, these approaches face the same problem, namely: often it cannot be tested whether we are dealing with the replication of a grammaticalization process or with some kind of ‘wholesale copying’ of functions which are anyway close in semantic space. Contiguity in cognitive space is what semantic maps are designed to test; implicational hierarchies, as assumed in grammaticalization clines, are based on the same contiguity assumptions. Now, since the functions of replica categories are very unlikely to leapfrog over contiguous domains of semantic space (or hierarchies), linguists would be unable to discern how a polysemous replica came about (Wiemer and Wälchli, 2012, pp. 27–44). Admittedly, examples in which such ‘jumps’ obviously occurred do exist, for instance, the Estonian *saama*-future discussed by Metslang (2017). Less sure is the use of the HAVE+anteriority participle as a compound past in extinct Polabian Slavic varieties under the influence of German (Löttsch, 1967). Yet another case is the MAT or PAT-replication of verb particles from German, e.g., in Sorbian varieties (Giger, 1998), e.g., Lower Sorbian *prěd-chytaś* (alongside *wu-chytaś*) ‘reproach’ (Germ. *vor-werfen*), Upper Sorbian *nutř widžeć* ‘understand, realize’ (Germ. *ein-sehen*).

Furthermore, Ziegeler (2017) proposes to dissolve, or circumvent, naïve implications carried by Replica CiGxn in Heine and Kuteva (2005), by assuming implicational hierarchies between more and less concrete meanings of a source expression:

The presence of an implicational hierarchy does not imply that the contact speaker has access to historical sources of a grammaticalizing item, only that the historical sources are often comparable with lexical roots of forms that may, at a particular time, still be visibly co-existing with the more grammaticalized form. This produces a situation of polysemy enabling grammaticalization pathways to be reconstructed by the speaker in contact.

(Ziegeler, 2017, p. 343)

This kind of polysemy is reminiscent of Hopper’s (1991) notion of layering: the lexical source is still ‘alive.’ It may, thus, supply a cue for contact speakers by which a function which is rather marginal in M can be retrieved and be grammaticalized, even ‘to a frequency exceeding that of the source language’ (Ziegeler, 2017, p. 344).

Moreover, all processes mentioned previously often lead to isomorphism (Wiemer and Wälchli, 2012, pp. 37–43), but they do not need to. The formation of isomorphism can be explained from the rise of diasystems (shared by a sufficient number of bilingual speakers for reasons of communicative economy), which then have to be accepted by monolingual speakers (or speakers with other ‘sets’ of multilingual competence); see Section 3.1.

4.2 Implicational hierarchies of borrowing and convergence

Implicational relations based on semantic maps should not be confused with implicational hierarchies which have been more widely discussed in connection with MAT-borrowings.

In general, analytical structures are easier to borrow than synthetic (inflectional, derivational) ones, and, by the same token, they are more apt to converge via contact. There are two entirely different, and probably largely independent, reasons for this. The first arises from the tight correlation between morphological bondedness and the formation of paradigmatic sets (i.e., for slots to follow in a sequence). This circumstance is directly reflected in borrowing scales (Thomason and Kaufman, 1988; Thomason, 2001). These sometimes show an ordering as in (5a) (cf. Romaine, 1995, pp. 64–67 for discussion):

5a. lexical item > morphology (derivational > inflectional) > syntax.

The second reason is a purely functional one, and it leads to a diametrically opposite ordering (see (5b)). As suggested in Stolz and Stolz (1996), discourse-oriented structures occupy the highest position on implicational hierarchies of convergence which slope down to units of decreasing structural scope:

5b. discourse > clause > phrase > word.

The discrepancy between (5a) and (5b) is rather superficial, it results partially from too vague notions of analyticity–syntheticity and partially from a neglect of discourse-oriented notions in (5a). The scale in (5b) appears to better reflect reality. For confirmation and discussion cf. Matras (1996, 2007, 2009, pp. 153–165, 243–245). Matras (2011, pp. 151–154) suggests that units farther to the left of the scale (5b) are both more easily inferable from discourse and more readily transferred from one language to another in an individual’s multilingual repertoire. In general, a systematic comparison of implicational hierarchies reveals three guiding motives: first, most likely (resp. first attested) are units with the highest cognitive and/or communicative pressure, in particular if the speaker’s authority demands explicit reference and responsibility for propositional content; second, units serving monitoring and control over interaction (like discourse markers, fillers, focus particles) will be borrowed and routinised in the first place; third, units which are more tightly associated with (or occur more frequently) in M will become entrenched in R before units which are more familiar in R (e.g., remote kin > close kin, unique referents > core vocabulary, higher > lower cardinal numerals); cf. Matras (2013, pp. 79–82).

4.3 Quantification and explaining (decrease of) distance

Parameters of social and physical geography (i.e., demographic data) were employed in Trudgill’s (1983, pp. 73–78) ‘gravity model,’ which accounts for population-based criteria such as the comparative size of inhabitants of different cities, or areas, and their relative social impact

(from ‘above’ and from ‘below’). To this one might add considerations like the proportion of bi- or multilectal speakers, who are the primary agents of adoptions from contact varieties in everyday speech and from whom these adoptions may spread among monolectal speakers. More sophisticated, computer-aided approaches toward modelling diffusion of innovations are now increasingly being developed in research that combines geography with linguistic variation (e.g., Wieling and Nerbonne, 2015; URPP – Language and Space).

By and large, we hitherto lack the application of good methods measuring convergence phenomena, comparable to those used in dialectometry (e.g., Heeringa and Nerbonne, 2001). Since these methods are mainly based on token frequency, they require sufficiently annotated, primarily spoken, corpora (Backus and Hakimov, this volume). Only recently have areal typology (Dahl and Wälchli, 2016) and contact linguistics (Adamou, 2016) started to employ variationist tools, the former with parallel, the latter with comparable corpora. Distance measures and aggregate analyses can serve to show how much varieties cluster and thus, for instance, how convergence zones on a larger geographic background arise, but also how other factors may rule out areal convergence. Thus, Szmrecsanyi (2012, p. 215), on the basis of aggregate analyses of British English dialects, showed that ‘mere geographic distance is a poor predictor of aggregate morphosyntactic variability,’ the factor ‘as-the-crow-flies distance’ is often superseded by ‘least-cost travel time’ (cf. also Nerbonne, 2013 on Dutch dialects). On the other hand, token-based aggregate and collostructional analyses have shown that accepted dialect divisions should be reordered, particularly if morphosyntactic features are accounted for (cf. Szmrecsanyi, 2013 for British English; Uibo et al., 2013 for Estonian).

Thus, quantifying methods allow not only to relativize the proportions and significance of convergence phenomena, but also to engage in data-mining, by which hitherto undiscovered, or underestimated, patterns of convergence or, conversely, of skewed distributions can be disclosed (Butt et al., 2012; Bickel, 2015; Backus and Hakimov, this volume). Factors can be weighed against each other, which is particularly called for if one wants to go beyond the most salient features and to account for the relation between token and type frequency, which, in turn, is a prerequisite for productivity measures.

Distances between varieties of a larger area applied to type-based comparisons of feature bundles can be established with the help of Neighbor Net (Huson and Bryant, 2006). Thus, for instance, European languages with prominent preverb strategies group into two clusters of prefixal perfectivization centring in Slavic and Kartvelian, but for different reasons; between both clusters the combinations of relevant parameters differ (Arkadiev, 2014, 2015).

In turn, token-based measures of features have furthermore been used to establish rates of lexical (i.e., MAT) borrowings. These rates were compared with the degree of PAT-replication in identical corpora of high-contact varieties by Adamou (2016). She concluded that the amount of contact words (loans) in a corpus is not, at least not directly, related to the extent of pattern replication. This allows for an interesting hypothesis: ‘although lexical borrowing is said to precede pattern replication in the various borrowing scales, it appears that the extent of lexical borrowing and pattern replication evolve independently’ (Adamou, 2016, p. 162).

5. Future directions

We can identify three issues in need of more and better research: (1) Can methods applied to macro-areas (see Sections 2.5, 4.3) reasonably be applied to considerably smaller areas (Section 5.1)? (2) How do convergence and divergence relate in a continuum, or agglomerate,

of varieties, with or without a roof language? Together with this, (3) how can we determine the relative weight of diatopic and diastratic (possibly also diaphasic) variation in convergence (Section 5.2)? All three issues require usage-based methods known from the variationist paradigm, which aim at disclosing distributional patterns within (variously defined) speech communities (Backus and Hakimov, this volume; Poplack, this volume). In particular, a desideratum is the study of productivity of patterns and the reasons why some of them become entrenched, while others fade out.

5.1 Does area size imply a difference in quality?

Here opinions diverge. Some scholars are optimistic that methods of macro-areal typology are equally apt for describing linguistic distances, or for discovering micro-areal clines and clusters (e.g., De Vogelaer and Seiler, 2012, p. 10; Borin, 2013, p. 5). Others are sceptical about this (e.g., Dahl, 2001, p. 1462; Wälchli, 2012), the main reason being that macro-areal patterns usually result from cumulative effects of polygenetic origin (i.e., at different places and times), whereas in micro-areas and dialect continua ‘constant spread of a feature across an originally homogeneous area from a single starting point,’ thus the wave model, seems to provide a more adequate explanation (Wälchli, 2012, pp. 264–265, also 235). This, however, is an empirical question which has rarely, if ever, been seriously examined. To test it, it is necessary (among other things) to compare the effects of ‘waves’ within one original continuum against effects of convergence on its edges, where one would expect contact with originally different zones. Moreover, we have to reckon with secondary convergence within larger zones (see Section 2.1 on transitional zones).

Bowern (2013) discusses evidence for a similar point: extrapolating dialect-contact models to languages and subgroups often proves problematic, i.e., genealogical relatedness ‘matters.’ A reason might be that the diffusion of innovations in speech communities of closely related varieties is supported by a higher percentage of bilingual speakers than in contact situations between genealogically more remote languages (2013, p. 416). After all,

contact-induced change between related languages might be higher than between unrelated languages because the ease of contact [via familiarity with grammatical patterns and vocabulary; BW] is greater. Alternatively, contact between unrelated languages [e.g., as the consequence of long-distance migration and/or colonisation; BW] might be higher than between related languages because of the social processes that led to the languages coming into contact with one another.

(2013, p. 421)

Moreover, genealogical relatedness is frequently accompanied by structural affinity and areal vicinity; therefore, again, testing the impact of relatedness (against intensity of contact) on borrowability and diffusion often turns out to be empirically unfeasible.

5.2 The relation between diatopic and diastratic convergence

In traditional wave models, peripheral (and insular) dialects have been regarded as more conservative, since innovations reach them the latest, or not at all. This amounts to saying that, on average, such dialects are less affected by processes favouring convergence. In addition, we usually find the viewpoint that borders cause divergence between varieties on both sides of a border, and that this accompanies convergence on the dialect-standard axis (rather in favour of

the standard) and across dialects on either side (e.g., Auer, 2004; Hinskens, Auer and Kerswill, 2005, pp. 29–32). What about, however, enhanced contact with varieties of neighbouring languages to which peripheral (and insular) dialects may be exposed much more than varieties in the ‘centre’? Obviously, there are attitudinal factors at play that vary for different sociolinguistic constellations. It appears worthwhile to establish, for different contact situations, the proportion between conservatism and innovativeness (which each can be subject to stimuli from contact varieties outside of the continuum of the roof language), as well as of speed of diffusion, in order to find more reliable predictors for convergence, especially in contact between non-standard varieties.

Concomitantly, one wonders whether zones of convergence are, in general, more stable within a historically grown dialect continuum (or a set of varieties under one ‘roof,’ or standard) or rather in border regions between languages where local varieties of different languages overlap. This would give life, in a new key, to Schuchardt’s metaphor of centrifugal and centripetal forces (see Section 2). Consider, for example, the so-called Alpine *come*-passive (cf. Gaetà, 2018), the spread of a *go*-future in French and Continental Germanic (cf. Hilpert, 2008: §4.2), or the mutual adoption of features in the border region, and beyond, between Latvian (Baltic, Indoeuropean) and Estonian (Finnic, Uralic) (Stolz, 1991; Wälchli, 2000).

6. Further reading

Auer, P., Hinskens, F. and Kerswill, P., eds. (2005). *Dialect Change (Convergence and Divergence in European Languages)*. Cambridge: Cambridge University Press.

This volume has remained among outstanding collections of comprehensive and insightful contributions on convergence (and divergence) in Europe, looked at from a diatopic and/or diastratic perspective. The focus is on Western and Northern Europe. The articles divide into those concentrating on structural aspects and those dealing with the motivations for convergence/divergence from a macro- and from a micro-sociolinguistic point of view.

d’Arcy, A. (2013). Variation and change. In: R. Bayley, C. Richard and L. Ceil, eds., *The Oxford handbook of sociolinguistics*, 1st ed. Oxford: Oxford University Press, pp. 484–502.

This article provides a concise introduction into the basic relations between linguistic variation and change, which, in turn, are essential for an understanding of convergence between speaker groups. Variation, and thus change, is not deterministic, but neither is it random. The relevant tenets are illustrated on the example of so-called new quotatives in different varieties of English.

Glaser, E. (2013). Area formation in morphosyntax. In: P. Auer, M. Hilpert, A. Stukenbrock and B. Szendrői, eds., *Space in language and linguistics: Geographical, Interactional, and cognitive perspectives*, 1st ed. Berlin and Boston: De Gruyter Mouton, pp. 195–221.

This article provides a state-of-the-art report on the account of the spatial distribution, and thus of convergence and areal clines, in dialectography. Possibilities of studying this distribution for syntactic phenomena are discussed mainly on the example of Alemannic dialects in Switzerland. In this context, it ‘re-emphasizes the fundamental issue of feature distributions as a common challenge in geolinguistics’ (from the introduction).

Hansen, B. and Umberto, A. (2016). Areality in modality and mood. In: J. Nuyts and J. van der Auwera, eds., *The Oxford handbook of modality and mood*, 1st ed. Oxford: Oxford University Press, pp. 406–429.

This article discusses general issues of the borrowing and spread of mood and modality markers, with a focus on contact-induced convergence. These classes of markers demonstrate very heterogeneous behaviour (modals are borrowed much more frequently than mood markers), which however correlates well with the borrowing hierarchy in (5b). The exemplary areas in which zones of convergence have been figured out are (Western and Eastern) Europe and mainland Southeast Asia.

Wiemer, B., Wälchli, B. and Hansen, B., eds. (2012). *Grammatical replication and borrowability in language contact*. Berlin and New York: De Gruyter Mouton.

This volume unites papers dealing with central issues of grammatical replication (grammaticalization, polysemy copying, etc.); this includes matter borrowing, whose parallels, but also asymmetries, with pattern borrowing are pointed out. The focus is on convergence in smaller areas and, though with a majority of papers dealing with languages in Eastern Europe, in their entirety the contributions show that the cognitive mechanisms responsible for pattern and matter borrowing, and the problems in distinguishing grammaticalization from related, but different kinds of (contact-induced) change are not restricted to any specific area (or type of language contact), nor to pattern borrowing.

7. Related topics

Balkans, borrowing, linguistic landscape and urban multilingualism, mixed languages, pidgins and creoles, social factors, typological factors, usage-based approaches, variationist methods

Abbreviations

| | |
|-----|------------------|
| L1 | first language |
| L2 | second language |
| M | model language |
| MAT | matter |
| NP | Noun Phrase |
| PAT | pattern |
| R | replica language |

Notes

- 1 I will conceive of dialects as varieties which are ‘used in a geographically limited part of a language area’ in which they are “‘roofed’ by a structurally related standard variety’ (Hinskens et al. 2005, p. 1) by virtue of tight genealogical affiliation. In a looser sense ‘dialect’ can be used to refer to virtually any non-standard variety of a language, with the proviso that there are varieties without standard ‘roofs’ (e.g., Arumanian, Romani, Molise Slavic, Ainu).
- 2 *Convergence* is sometimes employed as a synonym of any sort of *transfer*, or *interference*, or, to the contrary, it is narrowed down. Thus, Matras (2009, p. 236) notes that, after Weinreich (1953), *convergence* ‘has been used almost synonymously with “calque” and “pattern transfer.”’ Usage of this term may be much more theory-dependent, as with Myers-Scotton (2002), from whom I only adopt that convergence can be viewed both as a process and its outcome.
- 3 The distinction of *diatopic*, *diastratic*, and *diaphasic* variation was established by Coseriu (1988). The first two of these terms he himself ‘borrowed’ from Flydal (1951).
- 4 Transitional dialects have been opposed to mixed dialects, which are said to not display a uniform treatment of features but show a much higher frequency of ad hoc-borrowings. The relation between transitional and mixed dialects, in particular their diachronic relation, has remained disputed (Wiemer and Erker, 2012/13). Remarkably, some of the (structural and sociolinguistic) arguments in the discussion on a transitional – mixed dialect contrast are similar to those used in the ‘mixed language debate’ (O’Shannessy, this volume).
- 5 For equivalent terms cf. Höder (2014b, p. 46) and Lindstedt (2014, pp. 171–172).
- 6 For the historical and ethnographical background cf. Winnifrith (1987) and Kahl (1999, 2007).
- 7 On the basilect–mesolect–acrolect distinction cf. Trudgill (1992).
- 8 Older terms are ‘donor’ vs. ‘recipient’ language (Germ. *Geber-* vs. *Nehmersprache*), ‘source’ vs. ‘recipient’ language (Weinreich, 1953).

References

- Adamou, E. (2016). *A corpus-driven approach to language contact: Endangered languages in a comparative perspective*. Berlin and Boston: De Gruyter Mouton.
- Aikhenvald, A.Y. (2002). *Language contact in Amazonia*. Oxford: Oxford University Press.
- Ameka, F.K. (2006). Grammars in contact in the Volta basin (West Africa): On contact-induced grammatical change in Likpe. In: A.Y. Aikhenvald and R.M.W. Dixon, eds., *Grammars in contact. A cross-linguistic typology*, 1st ed. Oxford: Oxford University Press, pp. 114–142.
- Arkadiev, P. (2014). Towards an areal typology of prefixal perfectivization. *Scando-Slavica*, 60(2), pp. 384–405.
- Arkadiev, P.M. (2015). *Areal'naja tipologija prefiksals'nogo perfektiva*. Moskva: Jazyki slavjanskoj kul'tury.
- Auer, P. (2004). Sprache, Grenze, Raum. *Zeitschrift für Sprachwissenschaft*, 23, pp. 149–179.
- Bermel, N. (2000). *Register variation and language standards in Czech*. München: Lincom Europa.
- Bickel, B. (2015). Distributional typology: Statistical inquiries into the dynamics of linguistic diversity. In: B. Heine and H. Narrog, eds., *The Oxford handbook of linguistics analysis*, 2nd ed. Oxford: Oxford University Press, pp. 901–923.
- Bickel, B. and Nichols, J. (2006). Oceania, the Pacific Rim, and the theory of linguistic areas. *Proceedings of the annual meeting of the Berkeley Linguistics Society*, 32. Berkeley: Berkeley Linguistics Society, pp. 3–15.
- Birnbaum, H. (1970). Four approaches to Balto-Slavic. In: V. Rūķe-Draviņa, ed., *Donum Balticum (To Professor Christian S. Stang on the occasion of his seventieth birthday 5 March 1970)*. 1st ed. Stockholm: Almqvist and Wiksell, pp. 69–76.
- Bisang, W. (2006). Contact-induced convergence: Typology and areality. In: K. Brown, ed., *Encyclopedia of language and linguistics, Vol. 3*. 1st ed. Oxford: Elsevier, pp. 88–101.
- Bisang, W. (2010). Areal language typology. In: P. Auer and J.E. Schmidt, eds., *Language and space. An international handbook of linguistic variation, Vol. 1: Theories and methods*, 1st ed. Berlin and New York: De Gruyter Mouton, pp. 419–440.
- Bisang, W. (2015). Problems with primary vs. secondary grammaticalization: The case of East and mainland Southeast Asian languages. *Language Sciences*, 47, pp. 132–147.
- Borin, L. (2013). The why and how of measuring linguistic differences. In: L. Borin and A. Saxena, eds., *Approaches to measuring linguistic differences*, 1st ed. Berlin and Boston: De Gruyter Mouton, pp. 3–25.
- Bowern, C. (2013). Relatedness as a factor in language contact. *Journal of Language Contact*, 6, pp. 411–432.
- Breu, W. (1994). Der Faktor Sprachkontakt in einer dynamischen Typologie des Slavischen. In: H.-R. Mehlig, ed., *Slavistische Linguistik 1993*. 1st ed. München: Sagner, pp. 41–64.
- Breu, W. (2003). Bilingualism and linguistic interference in the Slavic-Romance contact area of Molise (Southern Italy). In: R. Eckardt, K. von Heusinger, and Chr. Schwarze, eds., *Words in time*, 1st ed. Berlin and New York: De Gruyter Mouton, pp. 351–373.
- Breu, W. (2012). The grammaticalization of an indefinite article in Slavic micro-languages. In: B. Wiemer, B. Wälchli, and B. Hansen, eds., *Grammatical replication and borrowability in language contact*, 1st ed. Berlin and New York: De Gruyter Mouton, pp. 275–322.
- Breu, W. (2018). On the influence of Italian on the grammar of Molise Slavic and Italo-Albanian. In: Th. Kahl, I. Krapova, and G. Turano, eds., *Balkan and South Slavic enclaves in Italy. Languages, dialects and identities*, 1st ed. Newcastle upon Tyne: Cambridge Scholars Publishing, pp. 215–236.
- Bubenik, V. (2001). The rise of Koine. In: A.-F. Christidis, ed., *A history of Ancient Greek (from the beginnings to late antiquity)*. 1st ed. Cambridge: Cambridge University Press, pp. 342–345.
- Butt, M., Prokić, J., Meyer, T. and Cysouw, M. (2012). Visualization of linguistic patterns and uncovering language history from multilingual resources. In: *Proceedings of the EACL 2012 joint workshop of LINGVIS & UNCLH*. Avignon, France: Association for Computational Linguistics, pp. 1–6.

- Campbell, L. (2006). Areal linguistics: A closer scrutiny. In: Y. Matras, A. McMahon, and N. Vincent, eds., *Linguistic areas. Convergence in historical and typological perspective*, 1st ed. Houndsmill, New York: Palgrave Macmillan, pp. 1–31.
- Chambers, J.K. and Trudgill, P. (1980/1998). *Dialectology*. Cambridge: Cambridge University Press.
- Coseriu, Eu. (1988). Die Begriffe “Dialekt”, “Niveau” und “Sprachstil” und der eigentliche Sinn der Dialektologie. In: J. Albrecht, ed., *Energeia und Ergon (Sprachliche Variation – Sprachgeschichte – Sprachtypologie)*. *Studia in honorem Eugenio Coseriu, Vol. 1*. 1st ed. Tübingen: Narr, pp. 15–43.
- Dahl, Ö. (2001). Principles of areal typology. In: M. Haspelmath, E. König, W. Oesterreicher, and W. Raible, eds., *Language typology and language universals (an international handbook)*, Vol. 2. 1st ed. Berlin and New York: De Gruyter Mouton, pp. 1456–1470.
- Dahl, Ö. and Wälchli, B. (2016). Perfects and iamitives: Two gram types in one grammatical space. *Lettas de Hoje*, 51(3), pp. 325–348.
- Danylenko, A. (2012). Auxiliary clitics in Southwest Ukrainian: Questions of chronology, areal distribution, and grammaticalization. *Journal of Slavic Linguistics*, 20(1), pp. 3–34.
- De Vogelaer, G. and Seiler, G. (2012). The dialect laboratory: Introductory remarks. In: G. De Vogelaer and G. Seiler, eds., *The dialect laboratory: dialects as a testing ground for theories of language change*, 1st ed. Amsterdam and Philadelphia: John Benjamins, pp. 1–31.
- Dryer, M.S. (1989). Large linguistic areas and language sampling. *Studies in Language*, 13, pp. 237–292.
- Durnovo, N.N., Sokolov, N.N. and Ušakov, D.N. (1915). Opyt dialektologičeskoj karty russkogo jazyka v Evrope. *Russkij filologičeskij vestnik*, 74, pp. 211–344.
- Emeneau, M.B. (1956). India as a linguistic area. *Language*, 32, pp. 3–16.
- Evans, N. and Wilkins, D.P. (2000). In the mind’s ear: The semantic extensions of perception verbs in Australian languages. *Language*, 76, pp. 546–592.
- Flydal, L. (1951). Remarques sur certains rapports entre le style et l’état de langue. *Norsk Tidsskrift for Sprogvidenskap*, 16, pp. 240–257.
- Foley, W.A. (2013). Language contact in the New Guinea region. In: R. Hickey, ed., *The handbook of language contact*, 1st ed. Malden, MA and Oxford: Blackwell, pp. 795–813.
- Gaetà, L. (2018). Im Passiv sprechen in den Alpen. *Sprachwissenschaft*, 43(2), pp. 221–280.
- Gast, V. and van der Auwera, J. (2012). What is ‘contact-induced grammaticalization’? Examples from Mayan and Mixe-Zoquean languages. In: B. Wiemer, B. Wälchli, and B. Hansen, eds., *Grammatical replication and borrowability in language contact*, 1st ed. Berlin and Boston: De Gruyter Mouton, pp. 381–426.
- Giger, M. (1998). Zu Lehnübersetzungen und Entlehnungen deutscher postponierbarer Präverben in sorbischen Dialekten. In: J.P. Locher, ed., *Schweizerische Beiträge zum XII. Internationalen Slavistenkongress in Krakau, August 1998*. 1st ed. Bern: Lang, pp. 129–170.
- Giles, H., Coupland, J. and Coupland, N., eds. (1991). *Contexts of accommodation*. Cambridge: Cambridge University Press.
- Gołąb, Zb. (1990). The ethnic background and internal linguistic mechanism of the so-called Balkanization of Macedonian. *Balkanistica*, 10, pp. 13–19.
- Gumperz, J.J. and Wilson, R. (1971). Convergence and creolization: A case from the Indo-Aryan/Dravidian border in India. In: D. Hymes, ed., *Pidginization and creolization of languages*, 1st ed. Cambridge: Cambridge University Press, pp. 151–167.
- Haig, G. (2017). Western Asia: East Anatolia as a transition zone. In: R. Hickey, ed., *The Cambridge handbook of areal linguistics*, 1st ed. Cambridge: Cambridge University Press, pp. 396–423.
- Haspelmath, M. (2001). The European linguistic area: Standard average European. In: M. Haspelmath, E. König, W. Oesterreicher, and W. Raible, eds., *Language typology and language universals (an international handbook)*, Vol. 2. 1st ed. Berlin and New York: De Gruyter Mouton, pp. 1492–1510.
- Heeringa, W. and Nerbonne, J. (2001). Dialect areas and dialect continua. *Language Variation and Change*, 13, pp. 375–400.
- Heine, B. and Kuteva, T. (2005). *Language contact and grammatical change*. Cambridge: Cambridge University Press.

- Hentschel, G. (2013). Zwischen Variabilität und Regularität, „Chaos“ und Usus: Zu Lautung und Lexik der weißrussisch-russischen gemischten Rede. In: G. Hentschel, ed., *Variation und Stabilität in Kontaktvarietäten Beobachtungen zu gemischten Formen der Rede in Weißrussland, der Ukraine und Schlesien*, 1st ed. Oldenburg: BIS, pp. 63–99.
- Hickey, R. (2003). How do dialects get the features they have? On the process of new dialect formation. In: R. Hickey, ed., *Motives for language change*, 1st ed. Cambridge: Cambridge University Press, pp. 213–239.
- Hilpert, M. (2008). *Germanic future constructions (a usage-based approach to language change)*. Amsterdam and Philadelphia: John Benjamins.
- Hinskens, F., Auer, P. and Kerswill, P. (2005). The study of dialect convergence and divergence: Conceptual and methodological considerations. In: P. Auer, F. Hinskens, and P. Kerswill, eds., *Dialect change (convergence and divergence in european languages)*, 1st ed. Cambridge: Cambridge University Press, pp. 1–48.
- Höder, St. (2014a). Constructing diasystems: Grammatical organisation in bilingual groups. In: T. Áfarli and B. Mæhlum, eds., *The sociolinguistics of grammar*, 1st ed. Amsterdam and Philadelphia: John Benjamins, pp. 137–152.
- Höder, St. (2014b). Convergence vs. divergence from a diasystematic perspective. In: K. Braunmüller, St. Höder, and K. Kühl, eds., *Stability and divergence in language contact (factors and mechanisms)*, 1st ed. Amsterdam and Philadelphia: John Benjamins, pp. 39–60.
- Holzer, G. (1998). Urslavisch und Baltisch. *Wiener Slavistisches Jahrbuch*, 44, pp. 27–56.
- Hopper, P. (1991). On some principles of grammaticization. In: E. Closs Traugott and B. Heine, eds., *Approaches to grammaticalization, Vol. 1*, 1st ed. Amsterdam and Philadelphia: John Benjamins, pp. 17–35.
- Horrocks, G. (1997). *Greek (A history of the language and its speakers)*. London and New York: Longman.
- Huson, D.H. and Bryant, D. (2006). Application of phylogenetic networks in evolutionary studies. *Molecular Biology and Evolution*, 23(2), pp. 254–267.
- Johanson, L. (2002). Contact-induced change in a code-copying framework. In: M.C. Jones and E. Esch, eds., *Language change (the interplay of internal, external and extra-linguistic factors)*, 1st ed. Berlin and New York: De Gruyter Mouton, pp. 285–313.
- Kahl, Th. (1999). *Ethnizität und räumliche Verteilung der Aromunen in Südosteuropa*. Münster: Institut für Geographie der Westfälischen Wilhelms-Universität Münster.
- Kahl, Th. (2007). *Hirten in Kontakt. Sprach- und Kulturwandel ehemaliger Wanderhirten in Epirus und Südalbanien*. Münster, Wien and New York: LIT.
- Keipert, H. (1999). Geschichte der russischen Literatursprache. In: H. Jachnow, ed., *Handbuch der sprachwissenschaftlichen Russistik und ihrer Grenzdisziplinen*, 1st ed. Wiesbaden: Harrassowitz, pp. 727–779.
- Koptjevskaja-Tamm, M. and Wälchli, B. (2001). The Circum-Baltic languages. An areal-typological approach. In: Ö. Dahl and M. Koptjevskaja-Tamm, eds., *Circum-Baltic languages, Vol. 2: Grammar and typology*. Amsterdam and Philadelphia: John Benjamins, pp. 615–761.
- Kouwenberg, S. (2001). Convergence and explanations in creole genesis. In: N. Smith and T. Venstra, eds., *Creolisation and contact*, 1st ed. Amsterdam and Philadelphia: John Benjamins, pp. 219–247.
- Kulkarni-Joshi, S. (2016). Forty years of language contact and change in Kupwar: A critical assessment of the intertranslatability model. *Journal of South Asian Languages and Linguistics*, 3(2), pp. 147–174.
- Kuteva, T. and Heine, B. (2012). An integrative model of grammaticalization. In: B. Wiemer, B. Wälchli, and B. Hansen, eds., *Grammatical replication and borrowability in language contact*, 1st ed. Berlin and New York: De Gruyter Mouton, pp. 159–190.
- LaPolla, R. (2003). The role of migration and language contact in the development of the Sino-Tibetan language family. In: A.Y. Aikhenvald and R.M.W. Dixon, eds., *Areal diffusion and genetic inheritance (problems in comparative linguistics)*. Oxford: Oxford University Press, pp. 225–254.

- Lindstedt, J. (2000). Linguistic Balkanization: Contact-induced change by mutual reinforcement. In: D.G. Gilbers, J. Nerbonne, and J. Schaecken, eds., *Languages in contact*, 1st ed. Amsterdam and Atlanta: Rodopi, pp. 231–246.
- Lindstedt, J. (2014). Balkan Slavic and Balkan Romance: From congruence to convergence. In: J. Besters-Dilger, C. Dermarkar, St. Pfänder, and A. Rabus, eds., *Congruence in contact-induced language change. language families, typological resemblance, and perceived similarity*, 1st ed. Berlin and Boston: De Gruyter, pp. 168–183.
- Lötzsch, R. (1967). Das Tempussystem des Slovinzischen im Vergleich zu dem des Sorbischen und Deutschen. *Létopis Instituta za serbski ludospyt*, 14, pp. 23–46.
- Małeck, M. (2004 [1934]). Do genezy gwar mieszanych i przejściowych (ze szczególnym uwzględnieniem granicy językowej polsko-czeskiej i polsko-słowackiej). In: J. Rusek, ed., *Dialekty polskie i słowiańskie*, 1st ed. Kraków: Wydawnictwo Uniwersytetu Jagiellońskiego, pp. 129–133.
- Matras, Y. (1996). Prozedurale Fusion: Grammatische Interferenzschichten im Romanes. *Sprachtypologie und Universalienforschung (STUF)*, 49(1), pp. 60–78.
- Matras, Y. (2007). The borrowability of grammatical categories. In: Y. Matras and J. Sakel, eds., *Grammatical borrowing in cross-linguistic perspective*, 1st ed. Berlin and New York: De Gruyter Mouton, pp. 31–74.
- Matras, Y. (2009). *Language contact*. Cambridge: Cambridge University Press.
- Matras, Y. (2011). Explaining convergence and the formation of linguistic areas. In: O. Hieda, Chr. König and H. Nakagawa, eds., *Geographical typology and linguistic areas*. Amsterdam: John Benjamins, pp. 143–160.
- Matras, Y. (2013). Contact, convergence, and typology. In: R. Hickey, ed., *The handbook of language contact*, 1st ed. Malden, MA and Oxford: Blackwell, pp. 66–85.
- Matras, Y. and Sakel, J. (2007). Investigating the mechanisms of pattern replication in language convergence. *Studies in Language*, 31(1), pp. 829–865.
- Metslang, H. (2017). Can a language be forced? The case of Estonian. In: D. Van Olmen, H. Cuyckens, and L. Ghesquière, eds., *Aspects of grammaticalization. (inter)subjectification ad directionality*, 1st ed. Berlin and Boston: De Gruyter Mouton, pp. 281–309.
- Myers-Scott, C. (2002). *Contact linguistics: Bilingual encounters and grammatical outcomes*. Oxford: Oxford University Press.
- Nau, N. (2012). Modality in an areal context: The case of a Latgalian dialect. In: B. Wiemer, B. Wälchli, and B. Hansen, eds., *Grammatical replication and borrowability in language contact*. Berlin and New York: De Gruyter Mouton, pp. 465–508.
- Nerbonne, J. (2013). How much does geography influence language variation? In: P. Auer, M. Hilpert, A. Stukenbrock, and B. Szmrecsanyi, eds., *Space in language and linguistics. geographical, interactional, and cognitive perspectives*, 1st ed. Berlin and Boston: De Gruyter Mouton, pp. 220–236.
- Nichols, J. (1992). *Linguistic diversity in space and time*. Chicago: University of Chicago Press.
- Olko, J., Borges, R. and Sullivan, J. (2018). Convergence as the driving force of typological change in Nahuatl. *Sprachtypologie und Universalienforschung (STUF)*, 71(3), pp. 467–507.
- Paul, H. (1968 [1880]). *Prinzipien der Sprachgeschichte*. Tübingen: Niemeyer.
- Rabus, A. (2014). Siblings in contact: The interaction of Church Slavonic and Russian. In: J. Besters-Dilger, C. Dermarkar, St. Pfänder, and A. Rabus, eds., *Congruence in contact-induced language change. language families, typological resemblance, and perceived similarity*, 1st ed. Berlin and Boston: De Gruyter, pp. 337–351.
- Rabus, A. (2019). Vergangenheitsbildung in gesprochenen karpatorussinischen Varietäten: Quantitativ-statistische Perspektiven. *Die Welt der Slaven*, 64(1), pp. 16–34.
- Romaine, S. (1995). *Bilingualism*, 2nd ed. Oxford: Blackwell.
- Ross, M.D. (1996). Contact-induced change and the comparative method: Cases from Papua New Guinea. In: M. Durie and M.D. Ross, eds., *The comparative method reviewed*, 1st ed. New York and Oxford: Oxford University Press, pp. 180–217.

- Ross, M.D. (2001). Contact-induced change in Oceanic languages in North-West Melanesia. In: A.Y. Aikhenvald and R.M.W. Dixon, eds., *Areal diffusion and genetic inheritance: problems in comparative linguistics*. Oxford: Oxford University Press, pp. 134–166.
- Ross, M. (2007). Calquing and metatypy. *Journal of Language Contact*, Thema, 1, pp. 116–143.
- Ross, M. (2008). A history of metatypy in the Bel languages. *Journal of Language Contact*, 2(1), pp. 149–164.
- Sasse, H.-J. (1985). Sprachkontakt und Sprachwandel: Die Gräzisierung der albanischen Mundarten Griechenlands. *Papiere zur Linguistik*, 32, pp. 37–95.
- Schmidt, J. (1872). *Die Verwandtschaftsverhältnisse der indogermanischen Sprachen*. Weimar: Böhlau.
- Schmidt, J.E. and Herrgen, J. (2011). *Sprachdynamik: Eine Einführung in die moderne Regionalsprachenforschung*. Berlin: E. Schmidt.
- Schuchardt, H. (1885). *Ueber die Lautgesetze. Gegen die Junggrammatiker*. Berlin: Oppenheim.
- Siegel, J. (1985). Koines and koineization. *Language in Society*, 14(3), pp. 357–378.
- Siegel, J. (2001). Koine formation and creole genesis. In: N. Smith and T. Venstra, eds., *Creolisation and contact*. Amsterdam and Philadelphia: John Benjamins, pp. 175–197.
- Silva-Corvalán, C. (2008). The limits of convergence in language contact. *Journal of Language Contact*, 2(1), pp. 213–224.
- Stolz, Th. (1991). *Sprachbund im Baltikum? Estnisch und Lettisch im Zentrum einer sprachlichen Konvergenzlandschaft*. Bochum: Brockmeyer.
- Stolz, Th. (2002). No *Sprachbund* beyond this line! On the age-old discussion of how to define a linguistic area. In: P. Ramat and Th. Stolz, eds., *Mediterranean languages (papers from the MEDTYP Workshop, Tirrenia, June 2000)*, 1st ed. Bochum: Brockmeyer, pp. 259–281.
- Stolz, Chr. and Stolz, Th. (1996). Funktionswortentlehnung in Mesoamerika. Spanisch-amerindischer Sprachkontakt: Hispanoindiana II. *Sprachtypologie und Universalienforschung*, 49, pp. 86–123.
- Szmrecsanyi, B. (2012). Geography is overrated. In: S. Hansen, Chr. Schwarz, Ph. Stoeckle, and T. Streck, eds., *Dialectological and folk dialectological concepts of space*. Berlin and Boston: De Gruyter Mouton, pp. 215–231.
- Szmrecsanyi, B. (2013). *Grammatical variation in British English dialects (a study in corpus-based dialectometry)*. Cambridge: Cambridge University Press.
- Thomason, S.G. (2001). *Language contact (an introduction)*. Edinburgh: Edinburgh University Press.
- Thomason, S.G. and Kaufman, T. (1988). *Language contact, creolization, and genetic linguistics*. Berkeley: University of California Press.
- Tosco, M. (2008). What to do when you are unhappy with language areas but you do not want to quit. *Journal of Language Contact*, 2, pp. 112–123.
- Trudgill, P. (1983). Linguistic change and diffusion. Description and explanation in geolinguistics. In: P. Trudgill, ed., *On dialect*, 1st ed. Oxford: Blackwell, pp. 52–87.
- Trudgill, P. (1992). Dialect contact, dialectology and sociolinguistics. In: K. Bolton and H. Kwok, eds., *Sociolinguistics today. International perspectives*, 1st ed. London and New York: Routledge, pp. 71–79.
- Trudgill, P. (2006). *New-dialect formation (the inevitability of colonial Englishes)*. Edinburgh: Edinburgh University Press.
- Tuten, D.N. (2001). Modeling koineization. In: L.J. Brinton, ed., *Historical linguistics 1999*, 1st ed. Amsterdam and Philadelphia: John Benjamins, pp. 325–336.
- Uiboed, K., Hasselblatt, C., Lindström, L., Muischnek, K. and Nerbonne, J. (2013). Variation of verbal constructions in Estonian dialects. *Literary and Linguistic Computing: Journal of Digital Scholarship in the Humanities*, 28(1), pp. 42–62.
- URPP Language and Space. *Language and space lab*. [online] Available at: www.spur.uzh.ch/en/departments.html.
- Uspenskij, B.A. (2002). *Istorija ruskogo literaturnogo jazyka (XI-XVII vv.)*, 3rd ed. Moskva: Aspekt press.
- Versteegh, K. (2014). *The Arabic language*, 2nd ed. Edinburgh: Edinburgh University Press.

- Wälchli, B. (2000). Infinite predication as a marker of evidentiality and modality in the languages of the Baltic region. *Sprachtypologie und Universalienforschung*, 53(2), pp. 186–210.
- Wälchli, B. (2012). Grammaticalization clines in space: Zooming in on synchronic traces of diffusion processes. In: B. Wiemer, B. Wälchli and B. Hansen, eds., *Grammatical replication and borrowability in language contact*, 1st ed. Berlin and New York: De Gruyter Mouton, pp. 233–272.
- Watkins, C. (2003). An Indo-European linguistic area and its characteristics: Ancient Anatolia. Areal diffusion as a challenge to the comparative method? In: A.Y. Aikhenvald and R.M.W. Dixon, eds., *Areal diffusion and genetic inheritance (problems in comparative linguistics)*, 1st ed. Oxford: Oxford University Press, pp. 44–63.
- Weinreich, U. (1953). *Languages in contact*. The Hague: Mouton.
- Weinreich, U. (1958). On the compatibility of genetic relationship and convergent development. *Word*, 14, pp. 374–379.
- Wieling, M. and Nerbonne, J. (2015). Advances in dialectometry. *Annual Review of Linguistics*, 1, pp. 243–264.
- Wiemer, B. (2004). Population linguistics on a micro-scale. Lessons to be learnt from Baltic and Slavic dialects in contact. In: B. Kortmann, ed., *Dialectology meets typology (dialect grammar from a cross-linguistic perspective)*, 1st ed. Berlin and New York: De Gruyter Mouton, pp. 497–526.
- Wiemer, B. (2009). Zu entlehnten Präfixen und anderen morphosyntaktischen Slavismen in litauischen Insel- und Grenzmundarten. In: L. Scholze and B. Wiemer, eds., *Von Zuständen, Dynamik und Veränderung bei Pygmäen und Giganten (Festschrift für Walter Breu zu seinem 60. Geburtstag)*, 1st ed. Bochum: Brockmeyer, pp. 347–390.
- Wiemer, B. (2019). *Matrěška* and areal clusters involving varieties of Slavic: On methodology and data treatment. In: A. Danylenko and M. Nomachi, eds., *Slavic in the language map of Europe*, 1st ed. Berlin and Boston: De Gruyter Mouton, pp. 21–61.
- Wiemer, B. and Erker, A. (2012/13). Übergangs- und Mischdialekte: Eine unnötige begriffliche Differenzierung? *Zeitschrift für slavische Philologie*, 69(1), pp. 1–54.
- Wiemer, B. and Seržant, I.A. (2014). East Slavic dialectology: Achievements and perspectives of areal linguistics. In: I.A. Seržant and B. Wiemer, eds., *Contemporary approaches to dialectology. The area of North, North-West Russian and Belarusian dialects*, 1st ed. Bergen: University of Bergen, pp. 11–80.
- Wiemer, B. and Wälchli, B. (2012). Contact-induced grammatical change: Diverse phenomena, diverse perspectives. In: B. Wiemer, B. Wälchli, and B. Hansen, eds., *Grammatical replication and borrowability in language contact*. Berlin and New York: De Gruyter Mouton, pp. 3–64.
- Winnifrieth, T.J. (1987). *The Vlachs: The history of a Balkan people*. London: Duckworth.
- Ziegeler, D. (2017). Historical replication in contact grammaticalization. In: D. Van Olmen, H. Cuyckens, and L. Ghesquière, eds., *Aspects of grammaticalization. (inter)subjectification and directionality*, 1st ed. Berlin and Boston: De Gruyter Mouton, pp. 311–352.

Creoles and pidgins

Why the latter are not the ancestors of the former

Salikoko S. Mufwene

1. Introduction

In this chapter, I focus on creoles and pidgins lexified by European languages. The primary reason is that it is debatable whether the term *creole* in particular can be extended to other vernaculars not lexified by European languages simply because their recent origins are associated with language contact in colonial settings. It is also purposefully that the essay is titled *creoles and pidgins*, contrary to the traditional phrase *pidgins and creoles*. I question the position that creoles have evolved by nativization from pidgin ancestors, with the latter typically characterized as broken languages that emerged out of sporadic interactions between speakers of European languages and those of non-European languages, in European trade and settlement colonies. To be sure, this tradition explains why in much of the literature expanded pidgins such as Nigerian Pidgin English and Papua New Guinea's Tok Pisin (which have vernacularized and thereby elaborated their structures) have often been lumped together with creoles (Holm, 1988; Thomason and Kaufman, 1988). I show here that the reason for lumping them together has little, if anything, to do with creoles evolving putatively from pidgin ancestors.

Creoles and pidgins first caught the attention of European philologists and historical linguists in the late nineteenth century, in the wake of the exploitation colonization of Africa, Asia, and the Pacific islands. Consistent with one interpretation of the Stammbaum, which suggests uniparental speciation, creoles and pidgins have been treated as anomalous, being the outcomes of language contact and therefore 'mixed languages.' Wedded to the nineteenth-century ideology of language purity, (genetic) linguistics has been little influenced by scholars such as Hugo Schuchardt (1882) who argued that these new colonial language varieties suggest instead how language speciation actually occurs. This is apparently how the Romance languages emerged, out of the contact of Vulgar Latin with Celtic languages in especially the Western Roman Empire (Schlieben-Lange, 1977).

Creoles and pidgins have thus been treated as 'bastard tongues' (Bickerton, 2008) or 'children out of wedlock' (Mufwene, 2001). They continue to be denied genetic connections with their lexifiers and related languages, allegedly because they inherited only most of the vocabularies from their lexifiers, while their grammars have separate origins, in the substrate languages (Thomason and Kaufman, 1988; Thomason, 2001; Lefebvre, 1998) and/or in the

'language bioprogram' (Bickerton, 1981, 1984). As pointed out by DeGraff (2003, 2005), one should not ignore the social philosophy identified as 'mission civilisatrice' in French and as 'white man's burden' in English that was emerging at that time. Assuming that non-Europeans were 'less evolved' and not 'civilized,' European nations invoked this philosophy to justify their 'exploitation colonization' (Mufwene, 2001) of Africa, Asia, and the Pacific.

This philosophy appears to have influenced the accounts provided of the apparent 'deviations' of creoles from their lexifiers, particularly by French philologists Emmanuel Bertrand-Bocandé (1849), Charles Baissac (1880), Julien Vinson (1882, 1888), and Lucien Adam (1883). These pioneers claimed that the (enslaved) Africans were anatomically and mentally inferior and thus unable to learn the sophistications of a European language, such as French, which is presumably more evolved. They also claimed that the structures of creoles reflected those of the 'primitive languages' that their 'creators' had spoken. Characterizing Gullah as the worst English dialect in the world, Gonzales (1922) attributed its peculiarities to racial phenotypes of African slaves. While race is no longer invoked in modern hypotheses, both substratist accounts of the emergence of creoles and those that invoke the role of children in shaping them unfortunately suggest the deep-rooted legacy of these early racist explanations of the divergence of the creoles from their lexifiers, as well pointed out by DeGraff (2003, 2005).

In this chapter, I wish to review the contact ecologies that produced both creoles and pidgins. Invoking the role played by interpreters during both the trade and the exploitation colonization of the relevant territories (Fayer, 2003; Mufwene, 2005), I argue that pidgins in the trade colonies actually may have emerged later than creoles did in the plantation settlement colonies; moreover, they both evolved by basilectalization, diverging away from earlier and closer approximations of their lexifiers (Chaudenson, 1992, 2001). Although speakers of pidgins as L2 varieties must have started with interlanguages cum individual transitional varieties toward closer approximations of the target language, there is no historical documentation of communal 'interlanguages' as incipient pidgins, *pace* Plag (2008, 2009), nor of 'jargons,' *pace* Mühlhäusler (1997). The latter appears to have misinterpreted the historic meaning of the term *jargon* as a language (variety) unintelligible to those reporting about it, especially when they are in a position of power such as the colonizers (Mufwene, 1997).

I argue that the history of European trade colonization from the fifteenth to the nineteenth centuries does not verify the old positions. The relevant history of trade contacts between Europeans and non-Europeans and of the exploitation colonization of Africa, Asia, and the Pacific islands by some European nations is replete with accounts of the service of interpreters, without whom the colonial enterprise may have never succeeded. I show here that the earlier records of non-native varieties of European languages, up to the eighteenth century, evidence closer approximations of the lexifiers rather than the pidgins reported in the nineteenth century or spoken today. I submit that the interpreters delayed the emergence of pidgins up to as late the end of the eighteenth century or perhaps even the early nineteenth century. This is about the same time when the English Pidgin of Canton, to which linguistics owes the term *pidgin* < *business* (Baker and Mühlhäusler, 1990) apparently emerged, with its counterparts of the Pacific, such as Bislama and Tok Pisin, emerging by the middle of nineteenth century.

2. Historical overview

Since the late nineteenth century, pidgins have been assumed to be defective, abnormal languages. Their putative expansion into expanded pidgins, in which case they continue to coexist with their substrate languages in the same territories, can thus be considered as a naturalization process enabled by their respective communities of practice that use them as vernaculars.

In plantation settlement colonies, their alleged nativization into creoles (interpretable also as indigenization, Hall, 1966) putatively had the same effect, except that they displaced the relevant substrate languages while typically continuing to coexist with their lexifiers.

The emergence of creoles and, to some extent, expanded pidgins has been associated with children, who had to use as vernaculars something more elaborate than the pidgins produced by their parents. In its modern form, this position was advocated most strongly by Bickerton (1981, 1984), according to whom the children resorted to the language bioprogram, a blue print for language, to do this. He assumed that the inventors of the antecedent pidgins communicated little with their masters and among themselves, although their approximations of the European colonial languages *did* function as their vernaculars, because they were not able to use their ancestral vernaculars in the exogenous colonies, owing to societal multilingualism in their new communities of practice. I dispute this hypothesis here, in support of other critics in the literature.

An important difference between creoles and pidgins lies in the geographical complementary distribution of the contact settings in which they emerged (Mufwene, 2008). Creoles are products of exogenous colonies, in which the substrate languages died concurrently and in fact the demographic significance of individual substrate languages or clusters thereof (such as the Fon-Gbe family in relation to Bantu languages in Haiti and Suriname) changed over time. By contrast, pidgins have continued to share geographical space with their respective substrate languages and to be influenced by them (Yakpo, 2017). Otherwise, both creoles and pidgins are outcomes of naturalistic language appropriation, although I argue here that this hypothesis of the evolution of both creoles and pidgins from simpler to more complex structures (*viz.*, pidgin > creole or expanded pidgin) is not supported by colonial history. The facts adduced here show that both creoles and pidgins have evolved in ways similar to the Romance languages, by gradual divergence away from their lexifiers.

The contacts leading to the emergence of pidgins have traditionally involved trade. Although this need not take place in a reduced language (as documented later), this is the reason why other contact-based languages with reduced morphosyntaxes and communicative functions have also been identified as pidgins, although their lexifiers are non-European and usually indigenous. The latter include, Tupi *aka* *Lingua Geral*, in Brazil; Chinook Jargon, in the North American Northwest (from Oregon to British Columbia); and Mòbilian, in the southeast of the United States. Many other varieties have been identified in Africa and Asia, such as Sango (in the Central African Republic), Lingala and (Kikongo-)Kituba (in DR Congo), Fanakalo/Fanagalo (in the mining areas of Southern Africa), and Hiri Motu (in Papua New Guinea). It is not evident that in the latter cases the contacts that produced them were as sporadic as between indigenous and European traders around the coast of Africa and Asia, as well as in the Pacific. However, all the pidgins have the peculiarity of serving as non-native lingua francas for the majority, with all parties maintaining their traditional vernaculars for interactions among themselves.

If there were many pidgins lexified by European languages in the late eighteenth and early nineteenth centuries, they must have disappeared when they were no longer needed (Hall, 1966; Mühlhäusler, 1997). The reason is that trade stopped in the relevant territory or the populations in contact shifted to an economically more advantageous or more prestigious language or one of wider communication. This was apparently the case for Chinese Pidgin English; but the history of trade provided here suggests that such cases are rare.

According to traditional accounts, the pidgins that have survived have evolved into ‘expanded pidgins,’ contact-based vernaculars now spoken at home and in other day-to-date interactions, especially in urban centres. They have elaborated their structures accordingly, to

function like ‘normal,’ natural languages. An example of this elaboration involves the transitivity, in Bislama, of some prepositions by the addition of the suffix – *im*, which converts them into verb-like predicates that can also be modified by a tense marker and be followed by a direct object (Jourdan, 2009), as in: *Mami bae insaetim kaletu* (Mother FUTURE take-inside laundry) ‘Mother will take [the] laundry inside.’

There are occasional references to less conventionalized and apparently embryonic stages of pidgins, called *incipient pidgins* or *jargons*. Ironically, the term *jargon* is also used with the meaning of ‘pidgin’ in names such as *Chinook Jargon* and *Mobilian Jargon*. This usage is actually consistent with the older meaning of the term, adopted in colonial discourse, for a language variety unintelligible to the writer or those assumed to be speaking a ‘normal’ language. As noted previously, the selection of the term *jargon* for incipient pidgins, varieties that so far remain hypothetical, also reflects ignorance of the colonial discourse.

The term *lingua franca*, often used to characterize the function of pidgins, is itself a semantic extension of *Lingua Franca*, the name of the mixed language used in medieval trade transactions between Northern, Southern, and Eastern Mediterraneans. Its lexifiers have been identified as Venitian, Portuguese, Catalan, Occitan, and Provençal. According to Dakhli (2008), *Lingua Franca* was Romance-based, with its lexifier assumed to be one or the other of these languages, depending on the part of the Mediterranean where trade was conducted (see also Schuchardt, 1909). The contribution of Arabic, Turkish, and other languages spoken around the Islamic Mediterranean to its vocabulary and structures was putatively marginal.

Lingua Franca is assumed to be the first known pidgin, although similar varieties must have arisen in other interethnic trade settings involving multitudes of languages, such as Songhai between Arab traders and indigenous populations across the Sahara and Baba Malay between Malay speakers and Chinese merchants in Southeast Asia. According to Schuchardt (1909), the name itself is an exonym, reflecting how the Arabs referred to the Northern Mediterranean populations, viz., *al-afrandž* ‘the Franks’ (i.e., Europeans); they assumed they were using the latter’s language in the trade.

Interestingly, some creolists in the 1960s (e.g., Whinnom, 1965, 1971; Thompson, 1961) assumed that, Sabir, a later variety of *Lingua Franca* heavily influenced by Portuguese, is what spread along the African coast and elsewhere around the world. It was allegedly relexified in the process, and eventually evolved into French, English, Dutch, and Portuguese creoles. This hypothesis helped account for the common structural properties of creoles, such as their periphrastic tense-mood-aspect systems and the non-inflectional way of marking the plural of nouns, typically with a third person plural pronoun, around the Atlantic and in the Indian Ocean. For instance, *Im bin fuh come* (in English creoles) and *li te pu vini* (in French creoles) ‘he/she ANTERIOR OBLIGATION come,’ as well as, respectively, *Im bin a come* and *li te ap vini* ‘he/she ANTERIOR DURATIVE come’ in English and French creoles.

In the next section, I show that better knowledge of the history of trade and colonial contacts between Europeans and non-Europeans suggests a different kind of explanation. What is correct about this hypothesis is that Portuguese became an important *lingua franca* after a pope divided the world outside Europe into two halves, according to a line that extended from west of Cape Verde to the middle of modern Brazil, between Castile and Portugal, at the Treaty of Tordesillas in 1494 (Clements, 2014). It was the principal language of trade until the late eighteenth century, if not the early nineteenth century, in coastal territories from Africa to Japan (Huber, 1999; Ostler, 2005; Newitt, 2010). That is when Portugal lost most of its trade colonies to England, France, and Holland, while Spanish was spreading to the territories west of the line that were effectively (being) colonized by Castile. The latter too would lose some of its Caribbean and American colonies to England and France, including Haiti, Jamaica, Trinidad,

Florida, Texas, New Mexico, and other North American territories that it had not colonized yet but could have claimed. The rivaling colonizers then started using their respective languages, except perhaps for the Dutch.

However, there is no evidence of a Portuguese pidgin that developed in these places. Instead, Portuguese indigenized into varieties such as *Papia Kristang* (literally, ‘language of the Christians’) and Macanese spoken by indigenous people who lived in the Portuguese trade forts, converted to Portuguese/Christian religion/culture, and shifted to Portuguese as their vernacular. The settings, in which the Portuguese traders were a small minority and there was relative indigenous ethnolinguistic homogeneity, accounts for significant substrate influence and divergence from the lexifier, especially after the Portuguese lost the trade colonies. This is all consistent with accounts of substrate influence by creolists such as Mufwene (1986), Keesing (1988), Singler (1988), and Kouwenberg (1994), viz., the influence is more extensive when the substrate languages are typologically related. Portuguese was widely used in the Kingdom of the Kongo, by indigenous priests, and in the Kingdom of Dahomey (present-day Benin) until the nineteenth century, when France colonized the latter on the exploitation model, but no Portuguese pidgin has been documented there.

The reason, as I explain later, is that trade between European commercial companies (not individuals!) and indigenous rulers (who could trade in expensive commodities such as gold, ivory, and humans) proceeded through intermediaries, also called ‘linguists’ and ‘interpreters,’ who learned each party’s language by immersion. European *factors* (‘traders’) and *lançados* (‘self-exiled Europeans who escaped religious persecution or evaded tax’) learned indigenous languages, while indigenous rulers dispatched members of their entourage or dispensable individuals (such as slaves) in Europe, especially Portugal, to learn the European language. Even the exploitation colonization of Africa and Asia proceeded through interpreters, which explains why the European colonial languages were not widely learned and have indigenized essentially since Independence (in the second half of the twentieth century). Very few Natives spoke Portuguese or any other European language until the twentieth century.

There are many reasons why one should question the hypothesis that creoles evolved from erstwhile pidgins. First, most creoles emerged probably in the eighteenth century, though some creolists claim that they emerged in the seventeenth century already. Although the territories where they emerged were colonized in the seventeenth century, with homesteads, it typically took much longer, owing to shortage of financial capital (Dunn, 1972), for large plantations to develop and provide the relevant ecology for the emergence of creoles. This scenario included: a significant non-European demographic majority (although this often did not reach the magical 20%/80% ratio often invoked in the literature), a segregated population structure, and, most importantly, rapid population replacement (Chaudenson, 1992, 2001; Mufwene, 2001). If pidgins emerged as late as explained previously, they could not have been the ancestors of creoles.

Second, as noted earlier, there is a geographical complementary distribution between the territories where creoles emerged, in plantation settlement colonies (including Kriol in Queensland, Australia) and those where pidgins emerged, outside settlement colonies (Mufwene, 2001, 2005, 2008). I will not repeat the claim that they emerged in trade colonies, because I argue later that trade was conducted in indigenous varieties and L2 approximations of the European languages, until late in the eighteenth century in Africa and Asia (Mufwene, 2014), and apparently also in the Pacific. According to Drechsel (2014), one of the languages used in the Pacific trade was Maritime Polynesian Pidgin (MPP), which preceded the current English pidgins spoken in the region since the mid-nineteenth century. MPP was used in ‘chain interpretation,’ involving a European language, a pidgin or some other indigenous lingua franca,

and an indigenous language. Keesing (1988) hypothesized that Melanesian Pidgin English(es) originated in the whaling ships, whereas Baker (1993) argued that they originated in Queensland plantations, from which part of the original labourers were recruited. It also appears that Nigerian and Cameroon Pidgin Englishes are offshoots of Krio – just like Pichi (spoken in Bioko, Equatorial Guinea), according to Yakpo (2009) – with some influence from Jamaican Creole (Mufwene, 2014).

Third, as noted previously, the term *pidgin* emerged in the early nineteenth century (Baker and Mühlhäusler, 1990), whereas the term *creole* was coined much earlier, in the late sixteenth century, in reference to non-indigenous people born in the colonies (Mufwene, 1997). It was used for the first time in reference to a language variety, one spoken apparently in present-day Casamance, in Senegal, in the seventeenth century (*Premier voyage du Sieur La Courbe fait à la coste d'Afrique en 1685*). The earliest documentation of any pidgin on the coast of Africa is perhaps from John Matthews (1788, p. 166), a lieutenant in the British royal army stationed in Sierra Leone, and is much closer to nonstandard English than to any modern pidgin:

Well, my friend, you got trade today; you got plenty of slaves?

No, we no got trade yet; by and by trade come. You can't go.

What you go for catch people, you go for make war?

Yes, my brother . . . gone for catch people; or they gone for make war.

3. Critical issues and topics

A striking characteristic about creoles and pidgins is their dominant isolating morphosyntax, which diverges from the fusional morphosyntax of especially nouns and verbs in their lexifiers and the agglutinating or polysynthetic verbal morphosyntax of many of their substrate languages. This has prompted a number of competing hypotheses about their emergence, including the racist views of the nineteenth-century French philologists cited in Section 1. The accounts have typically been grouped into three categories, viz., the substratist, the superstratist, and the universalist hypotheses, although the categories do not do justice to positions that bridge some of these approaches, including what was once identified as the complementary hypothesis. What is not included in this tripartite categorization is the question of whether or not creoles emerged from pidgins, discussed in Section 2. An issue connected to all these competing hypotheses is whether creoles are evolutionarily and typologically an exceptional group of languages, which will be addressed toward the end of this section. Here I will also discuss the relevance of the subject matter of the emergence of creoles and pidgins to that of language birth and death, especially because creoles and pidgins instantiate language speciation, hence language birth, the other side of the coin typically overlooked in the literature on language endangerment and loss (Mufwene, 2008, 2017). Creoles also instantiate a possible outcome of language shift, similarly to the emergence of, for instance, the Romance languages.

Substratist positions are historically related to the *baby talk hypothesis*, which can be traced back to the late nineteenth-century French creolists Bertrand-Bocandé (1849), Baissac (1880), Adam (1883), and Vinson (1882). According to them, the languages previously spoken by the Africans enslaved on New World and Indian Ocean plantations were the primary reason why the European lexifiers they appropriated were restructured into creoles. These early creolists assumed African languages to be 'primitive,' 'instinctive,' in 'natural' state, and simpler than the 'cultivated' European languages with which they came in contact. Creoles' systems were considered to reflect those of the relevant non-European languages, which allegedly reflected the mental inferiority of those who produced them. The enslaved Africans were thus

considered as incapable of learning the putatively ‘more evolved’ structures of the European languages. The ‘baby-talk’ connection is that, in order to be understood, the Europeans supposedly had to speak to the Africans like to children. More or less the same idea is to be found in the ‘foreigner talk’ hypothesis, according to which Europeans reproduced the non-Europeans’ inaccurate approximations of their languages (Bloomfield, 1933).

The revival of the substrate hypothesis (without its racist component) has been attributed to Sylvain (1936). Although she recognizes significant influence from French nonstandard dialects, she concludes her book, surprisingly, with the statement that Haitian Creole is Ewe spoken with a French vocabulary. (Much of the relexification hypothesis, discussed later, is based on this conclusion!) Over two decades later, disputing American dialectologists’ claim that there was virtually no trace of African languages in ‘Black English,’ Turner (1949) highlighted some morphosyntactic similarities between what he called ‘the Gullah dialect’ and some West-African (especially Kwa) languages. He then concluded that ‘Gullah is indebted to African sources’ (1949, p. 254). The work stimulated more research on African substrate influence on African-American English (e.g., Dillard, 1972) and on Caribbean creoles (e.g., Alleyne, 1980; Holm, 1988).

Mufwene (1990, 2010) identifies three main schools of the substrate hypothesis today. The first, led by Alleyne (1980, 1996) and Holm (1988) is closer to Turner’s approach and is marked by what is also its main weakness: invocation of structures from diverse African languages without explaining what kinds of ecological factors account for this seemingly random combination of influences. This approach has been criticized as smacking of the ‘Cafeteria Principle,’ suggesting that these new language varieties were put together in unprincipled ways with materials selected randomly from any language. I argued in Mufwene (2001) that, in a contact setting, materials for communication can in fact be selected from any language. (Interestingly, one can recognize the same idea in the current literature on ‘translanguaging,’ for instance, in MacSwan, 2017.) What is needed to legitimate such diverse substrate influence is a set of ecological factors that can explain why those particular materials or grammatical patterns rather than any other competing with them were selected into the emergent language varieties. Ecology-sensitive markedness considerations are relevant (Mufwene, 1991, 2001) as well as periodization (Chaudenson, 1992, 2001). As noted earlier, the proportions of slaves from different parts of Africa varied over time, which, despite the founder principle (Mufwene, 1996, 2001), could have favoured influence from particular (groups of) languages variably. In other words, the ‘principle’ part of the ‘Cafeteria Principle’ is what has been lacking from these substratists’ accounts.

The second school, identified by its practitioners as the *relexification hypothesis* (RH), is fully articulated by Lefebvre (1998, 2004), who argues that Haitian Creole (HC) consists largely of French lexical entries spoken with the grammar of languages of the Ewe-Fon (or Fon-Gbe) group. Extended to other creoles, the position, has been repeated in some of the contributions to Lefebvre, White and Jourdan (2006), though some others (see especially Aboh, 2006 and Siegel, 2006) dispute it, and almost all of those to Lefebvre (2011) are more cautious. Objections to this phylogenetic hypothesis include the following: (1) RH’s ‘comparative’ approach has not taken into account several features that HC (also) shares with nonstandard French (as already noted by Sylvain, 1936); (2) RH downplays features which HC shares also with several other African languages that were represented in Haiti during the critical stages of its development – thus it is not obvious why the exclusive focus on the Ewe-Fon languages; (3) studies of naturalistic second language acquisition provide no evidence in support of RH, even if the emergence of creoles could at all be associated exclusively with adult L2-learners (Chaudenson, 2001, 2003); and (4) RH does not account for those cases where

HC has selected structural options which are not consistent with those of Ewe-Fon (Mufwene, 2010). Aboh (2015) shows that even if, on the surface, HC appears to replicate a Gbe construction, the details of the pattern and its semantics are not necessarily identical. (5) Moreover, relexificationists assume, disputably, that languages of the Ewe-Fon group are structurally identical in all respects and that no competition of influence was involved among them. The most elaborate critique of RH is DeGraff (2002), which is complemented by various refined analyses of hybridized structures in Haitian by Aboh (2006, 2009, 2015, among many others). For contrary evidence from other creole-like languages, see especially Siegel (2006).

The least disputed version of the substrate hypothesis is provided in Keesing (1988), which shows that substrate languages may indeed prevail when the non-European languages that came in contact with the European lexifier are typologically alike, sharing several structural features. Thus, Melanesian pidgins are like (most of) their substrates in having DUAL/PLURAL and INCLUSIVE/EXCLUSIVE distinctions and in having a transitive marker on the verb. Sankoff and Brown (1976) had shown similar influence with the bracketing of relative clauses with *ia*, as do Sankoff (1993) about the focus marker in Melanesian pidgins and Jourdan (2009) regarding the transitivity of some prepositions into verb-like predicates by the addition of the suffix *-im*, in Solomon Island Pidgin. Keesing argued that even the SVO order of Melanesian pidgins' sentences reflects the order of the argument affixes in the verb complex, which is often used elliptically, although the dominant order of free arguments is VSO in the relevant substrate languages.

This evolution is facilitated by what Singler (1988) calls 'homogeneity of the substrate' (argued for earlier in Mufwene, 1986), if it is not downright its consequence. However, the pidgins have not inherited all peculiarities of Melanesian languages. For instance, they do not have their VSO major constituent order, nor do they have much of a numeral classifying system in the combination of *pela* with quantifiers. For an extensive discussion of substrate influence in Atlantic and Indian Ocean creoles, see Muysken and Smith (1986) and Mufwene (1993; Michaelis, 2008). For similar discussions about creoles and the like in the Pacific, see Lefebvre, White and Jourdan (2006) and Lefebvre (2011).

Competing with the preceding phylogenetic views has been the superstrate hypothesis, which can be lumped together with the old dialectologist views expressed by, e.g., Faine (1937) and Goodman (1964) for French creoles. According to it, the primary, if not the exclusive, sources of creoles' structural features are their nonstandard lexifiers. Regarding African-American English (AAE; not really a creole but with a similar history!), Krapp (1924) and Kurath (1928), for example, claimed that this variety was an archaic retention of the non-standard speech of low-class European colonists with whom the enslaved Africans had been in contact. According to these dialectologists, African substrate influence was limited to some isolated lexical items such as *goober* 'peanut,' *gumbo*, and *okra*. Although substrate influence need not be dismissed offhand, the position is worth considering critically. One must factor in the presence of non-native speakers of English among the European indentured servants with whom the Africans interacted regularly, especially during the early stages of colonization. This is even more significant in the case of Surinam, where the Dutch colonists, replacing their English predecessors, adopted English as the language of communication with the Africans. It must not be assumed that the average Dutch colonist spoke it fluently.

It would take until McDavid (1950) and McDavid and McDavid (1951) before dialectologists made allowance for some African grammatical contributions to AAE, owing largely to the publication of Turner (1949). Otherwise, D'Eloia (1973) and Schneider (1989) invoke several dialectal English models to rebut Dillard's (1972) thesis that AAE developed from an erstwhile West-African Pidgin English (WAPE) brought over by the Africans. Had they

known then that the language of the slave trade was Portuguese, they could have pointed out that WAPE varieties probably emerged later than AAE in the first place (Mufwene, 2014). The scant evidence presented by Dillard (reported about Africans in American colonies!) dates from the early eighteenth century and is closer to nonstandard English than to Pidgin English:

- 1 [. . .] and we nebber see our mudders any more (1732, Dillard, 1992, p. 62)
- 2 By-and-by you die, and go to the bad place, and after a while Cuff die and go and knock at the good gate (mid-eighteenth century, Dillard, 1992, p. 62)

Hancock's (1986) 'Guinea Coast Creole English,' spoken primarily in the mixed households of English 'factors' and indigenous women need not have been as widespread as he claims. It is in fact from the end of the eighteenth century:

- 3 Massa, me been see that white man in me country, in de town where me live, he been come dere one night for sleep, one blacksmith countreyman for me been with him, me been give him rice for supper, and soon, soon in the morning he been go towards the Moors' country (Bolingbroke, 1796, p. 71).

To be sure, this passage contains features that one may consider creole, viz., the use of *been* as a marker of ANTERIOR, of *for* with the meaning of infinitival *to*, and the invariant use of *me* in the subject and possessive function. However, note that *me* in the latter function is not unusual in British nonstandard English. On the other hand, the attestation is not inconsistent with the hypothesis that pidgins emerged much later than has been claimed in the literature, for reasons I articulate in the next section.

Since the late 1980s, Shana Poplack and her associates have shown that African-American Vernacular English (AAVE) shares many features with white nonstandard vernaculars in North America and England, thus it has not developed from an erstwhile creole (Poplack and Tagliamonte, 2001; Poplack, 2000). Because some of the same features are also attested in creoles (Rickford, 1998), we come back to the question of whether many, if not most, features of creoles did not originate in their lexifiers in the first place. The other question is also whether African substrate influence on AAVE must of necessity have come through creoles. History suggests that Gullah and AAVE emerged in a geographic complementary distribution and apparently concurrently, the former on the American southeastern coast and the latter inland (Mufwene, 2015b). During the eighteenth century, when the plantation economy was the most prosperous in English North America, only 15% of the slaves were imported from the Caribbean (Rawley, 1991). The oldest Caribbean English creoles would be just emerging then.

Regarding French creoles, the dialectologist position was first defended by Faine (1937), according to whom HC was essentially Norman French. This position was espoused later by Hall (1958), who argues that,

the 'basic' relationship of Creole is with seventeenth-century French, with heavy carry-overs or survivals of African linguistic structure (on a more superficial structural level) from the previous language(s) of the earliest speakers of Negro Pidgin French; its 'lexical' relationship is with nineteenth- and twentieth-century French.

(1958, p. 372)

Curiously, the 'Negro Pidgin French' can be situated nowhere in colonial history. The so-called français tirrailleur was fabricated in the late nineteenth century by French colonial military

officers, who assumed that the African infantry would learn it more easily than French; but it hardly took off (Vigouroux and Mufwene, 2013). ‘Le petit nègre,’ now known as Abidjanais, is a twentieth-century phenomenon.

Chaudenson (2001, 2003) is more accommodating to substrate influence as a factor that accounts for the more extensive structural divergence of creoles from their lexifiers compared to their non-creole colonial kin. Chaudenson’s allowance for substrate influence is fleshed out especially by Corne (1999), who was the first to articulate the most explicitly how feature selection can be driven by congruence, even if only partial, between the languages in contact. Although, unlike Pacific pidgins, the Atlantic and Indian Ocean French creoles did not typically emerge in settings that satisfied the ‘homogeneity of the substrate’ condition, partial structural congruence between their substrates and nonstandard French favoured the selection of the particular features they have, for instance, in the domain of time reference. No more a dialectologist than Corne, Aboh (2006 ff) has carried this approach farther with detailed analyses that show how structural traits can be hybridized in ways similar to biological gene recombination, more specifically with serial verb constructions and number delimitation. One must then determine whether such substrate influence, which does not boil down to mere introduction of features from substrate languages (identified by Allsopp, 1977 as ‘apports’), was facilitated by the larger numerical proportion of speakers of the relevant languages and/or by the time of the insertion of these in the linguistic feature pool. See Singler (1996, 2009) for such considerations regarding HC.

The *universalist hypotheses* were formulated especially against the substrate hypotheses, in the 1980s and 1990s. They have forerunners in the nineteenth century. For instance, Adolfo Coelho (1880–1886) partly anticipated Bickerton’s (1981, 1984) *language bioprogram hypothesis* in stating that creoles ‘owe their origin to the operation of psychological or physiological laws that are everywhere the same, and not to the influence of the former languages of the people among whom these dialects are found’ (cited in Gilbert, 1986). Bickerton pushed things further in claiming that children, rather than adults, made creoles by setting the parameters of these new language varieties to their unmarked, or default, options as specified in bioprogram. (For a comparison of Bickerton’s ‘language bioprogram’ with Chomsky’s ‘Universal Grammar’ since 1986, see Baptista, 2012.) To account for cross-creole structural differences, Bickerton (1984, pp. 176–177) invoked a ‘Pidginization Index’ (PI) that includes the following factors: (1) the proportion of native to non-native speakers during the initial stages of colonization, (2) the duration of the early (homestead) phase, (3) the rate of increase of the slave population after that initial phase, (4) the kind of social contacts that obtained between the native speakers of the lexifier and the learners, and (5) whether or not regular interactions between the two groups continued after the formation of the new language variety. These factors, on which he did not elaborate in later work, are incorporated in Chaudenson’s (1992 ff) work and in Mufwene’s (1996 ff) ecological approach. One can also use them to counter-argue that the very fact of invoking a PI to account for structural differences among creoles suggests that adults too participated in shaping these vernaculars. As pointed out by, for instance, DeGraff (1999a), creoles were not developed by children alone.

Some nagging questions with Bickerton’s position include the following: is his otherwise intuitively sound PI consistent with his hypothesis that creoles evolved abruptly, in one human generation, by nativization, from a pidgin ancestor? Is the ‘abrupt creolization’ hypothesis consistent with the social histories of the territories where the ‘classic creoles’ of the Atlantic and Indian Ocean developed? How can we explain similarities between ‘abrupt creoles’ and ‘expanded pidgins’ when the stabilization and structural expansion of the latter is not necessarily associated with restructuring by children (Meyerhoff, 2009), *pace* some claims to be

found in references such as Holm (1988) and Thomason (2001)? The scenario articulated in Sections 1 and 2 argue against these claims. Roberts (1998, 2004) shows clearly that even the history of Hawaii, which is so relevant to making sense of Bickerton's evidence, does not support his hypothesis.

To begin with, Hawaii is not representative of the other territories where creoles emerged. It was not a typical plantation settlement colony (Mufwene, 2005, 2008), as its labour force consisted of contract labourers rather than enslaved people. It had less ethnolinguistic diversity, viz., Chinese, Japanese, Koreans, Filipinos, and a few Portuguese, who arrived at different times during the nineteenth century, after the abolition of the slave trade, and lived separately, in different 'houses,' with their families brought from home. The Portuguese assimilated with other Whites on the islands. The other contract labourers maintained their heritage languages (Tagalog in the case of the Filipinos) and they received instructions for work from their foremen, who were bilingual in English. Unlike in the other plantation settlement colonies, Hawaiian Creole emerged in the city, while Pidgin emerged on the plantations, from interethnic interactions. These were not as intense as on the plantations of the Atlantic and Indian Ocean, where interethnic mixing occurred already during the homestead phase and the traditional ethnolinguistic distinctions have not survived. In the latter colonies, no pidgin emerged, except between the European colonists and the Native Americans they traded with; and most of these were lexified by indigenous languages. The colonists referred to them as *baragouins* (Prudent, 1980; Goodman, 1985; Wylie, 1995; Chaudenson, 1992, 2001), with the same meaning as *jargon* (as discussed earlier).

In any case, not all creolists who have invoked universalist explanations have made children critical to the emergence of creoles. For instance, Sankoff (1979) and Mühlhäusler (1981) make allowance for adults to participate in the emergence of creoles, perhaps because they worked on the expanded pidgins of Melanesia, which have often been lumped together with creoles.

Few creolists nowadays subscribe to one exclusive genetic account, as evidenced by the contributions to Mufwene (1993) and implicitly those to Lefebvre (2011). The 'complementary hypothesis' (Baker and Corne, 1986; Corne, 1999; DeGraff, 2009; Hancock, 1986; and Mufwene, 1986 ff) seems to be an adequate alternative, provided we can articulate the ecological conditions under which the competing influences (between the substrate and lexifier languages, and within each group) may either be congruent or prevail upon each other. This position, which can be characterized as uniformitarianist, was well anticipated by Schuchardt (1909, 1914) in his accounts of the genesis of the Mediterranean Lingua Franca and of Saramaccan. More and more research is now underway uncovering the sociohistorical conditions under which different creoles have emerged, for instance, Arends (1989 ff), Baker (1982 ff), Chaudenson (1979 ff), Corne (1999), Mufwene (2001 ff), Arends (1995), Walicek (2007), Aboh and DeGraff (2017). Aboh's (2006 ff) hypothesis of hybridization of features even within properties of lexical items is an improvement not only on Corne's (1999) congruence model but also on Mufwene's (2001 ff) ecology-specific restructuring.

In connection with the preceding discussion, note that the traditional saltationist claim that creoles emerged within one generation, from a pidgin ancestor, has increasingly been questioned by, e.g., Chaudenson (1979 ff), Arends (1986 ff), Singler (1996 ff), Mufwene (1996 ff), and Aboh and DeGraff (2017). Baker (1995) provided part of the evidence against assuming the abrupt emergence of pidgins themselves, pointing out that their features did not all emerge at the same time. What he did not say is whether they emerged by gradual divergence away from their lexifiers, like Chaudenson (1979 ff) hypothesized for creoles, or started from 'jargons' cum unconventionalized pidgins, as claimed by Mühlhäusler (1997). I argue later that the earlier alternative is more likely.

The strongest evidence lies in the fact that the oldest documentary evidence shows more similarity to European nonstandard varieties than to the later or present-day materials. Little has actually been documented about the earlier stages of pidgins lexified by European languages. One may actually argue that in this case the absence of evidence is evidence of absence. What little has been cited comes from territories where creoles emerged, in which case interlanguage utterances produced by Bozal slaves have been invoked, perhaps too hastily, to claim that creoles started from antecedent pidgins.

Documentation by Walter Brasch (1981) shows that in eighteenth-century North America, the most divergent English utterances produced by enslaved Africans are attributed to Bozales, who had arrived recently from Africa, whereas the Creole (i.e., locally born) slaves and the ‘seasoned slaves’ (who have lived longer in the colony) are typically described as speaking (fairly) good English. Likewise, quoting H. Jones (1724/1956), Kulikoff (1986, p. 317) reports that ‘slaves born in Virginia “talk good English, and affect our language, habits, and customs.”’ No historian has ever characterized a whole slave population during the beginnings of any plantation in the Caribbean, North America, and the Indian Ocean as speaking a pidgin understood as a broken language. The following quotations from Brasch (1981) suggest that the English competence of the enslaved Africans varied from one speaker to another, corresponding largely to whether they were locally born or African born, and, in the latter case, how long they have lived in the colony:

4. Ran away . . . the following negroe’s viz. Sambo, a small, thin visaged Fellow, about 30 years of age, speaks *English* so as to be understood . . . Aron . . . can’t even speak *English* . . . Berwick . . . can’t speak English. They have been in above 8 months in the country. (*Virginia Gazette*, Aug. 24, 1751, p. 3)
5. . . . as he was imported very young he speaks very good English. (*Virginia Gazette*, Dec. 12, 1755, p. 4)
6. . . . speaks plain for an *African* born, but avoids looking in the face of them he is speaking to as much as possible. (*Rind’s Virginia Gazette*, Aug. 8, 1788, p. 3)

Bearing in mind that there was plenty of negative stereotyping in the way the speech of slaves was represented, note that the dialogue in (7), from John Leacock’s *The fall of British tyranny* (1776, p. 16), sounds closer to English than the samples in (8–10) from present-day Nigerian and Cameroon Pidgin Englishes, which also support the hypothesis of speciation by basilectalization:

7. *Cudjo*: Disse brak man, disse one, disse one, disse one, disse one, come from Hamton, disse one, disse one, disse one, come from Nawfok, me come from Nawfok too.
Kidnapper: Very well, what was your master’s name?
Cudjo: Me massa name Cunney Tomsee.

Kidnapper: . . . what’s your name?
Cudjo: Me massa cawra me Cudjo.
Kidnapper: Cudjo? – very good – was you ever christened, Cudjo?
Cudjo: No massa, me no crissen.

8. na so a bin don di tel yu (Cameroon PE, Féral, 1989, p. 125)
'This is what I had been telling you'
9. You no see how the title rough like sand-paper for mouth: 'Dr Chief Mrs.' E no catch at all. (Nigerian PE, Chinwa Achebe, *A man of the people*, 1966, p. 19)
'Don't you see how rough the title is, like sand-paper in the mouth: 'Dr Chief Mrs'? It is not catchy/appealing at all.'
10. You see wetin I de talk. How many minister fit hanswer sir to any Tom, Dick and Harry wey senior them for age? (Chinwa Achebe, *A man of the people*, 1966, pp. 11)
'You see what I'm saying? How many ministers can say sir in response to any Tom, Dick and Harry who is older than them?'

The evolution-by-basilectalization scenario is also consistent with the gradual way in which plantation colonies as economic ventures evolved, having started from homestead settings settled by small integrated groups in which the enslaved Africans were in the minority to large plantations on which they became the overwhelming majority. Shortage of money made it difficult to import many slaves during the homestead phase and the colonial populations then grew more by birth than by importation (Mufwene, 2001 ff).

As for pidgins, a number of reasons other than the absence of attestations of such varieties before the late eighteenth century can be adduced to explain why they developed so late. The trade between Europeans and non-Europeans falls in the category of the then emergent worldwide economic globalization (Mufwene, 2014). It involved European mercantile companies (viz., the British, Dutch, and French East India Companies) and non-European rulers. (In the case of the Portuguese, the expeditions were funded by the monarchy.) Transported on fleets rather than isolated ships, the traded commodities were of expensive value, including large loads of gold, textile, ivory, and human beings.

Such transactions could not have taken place in a broken language. Colonial history actually appears to have relied on interpreters from the beginning all the way to the early twentieth century, as is evidenced by sources such as the following: Austen and Derrick (1999), Beechert (1985), Curtin (1984), Drechsel (2014), van Dyke (2005), Gray and Fiering (2000), Kennedy (2013), Lawrence, Osborn and Roberts (2006), Newitt (2010), Northrup (2009), and Wolf (1982). Van Dyke reports that the Chinese were very strict about the service of interpreters, who acted as brokers; they refused to negotiate directly with the Europeans even if the latter had learned to speak Chinese and complained in some cases that their interpreters were not so competent. Newitt (2010) is very clear about the role played by the *Lançados* and *Tangomaos* (self-exiled European traders who escaped jail sentences or religious persecution) and their children from unions with African women, in serving as go-betweens, along with the *Pombeiros* (African merchants), in the slave trade. Even the Mediterranean trade between North African Arabs and Northern Mediterranean Europeans during the Middle Ages apparently depended more on interpreters than on general knowledge of Lingua Franca (Dakhli, 2008).

The practice of interpreters was also the norm in trade between the European colonists and Native Americans, although the lingua francas were indigenous rather than European in this case. (They had been used in pre-colonial trade.) Other than some of the sources cited earlier (among a few others), Lee (2014) and Mello (2014) also cite the use of *factors* (similar to *lançados* on the West African coast) and *linguas* 'interpreters,' who were critical to the trade in the early decades of the colonization of Brazil.

What needs explaining is how the interpreters emerged. Indeed, the initial contacts were very difficult and it is not clear how they took place. No historical account ever mentions

improvised ‘jargons’ of any kind, although there are references to use of gestures. The first attempts by the Portuguese to take with them, during their first expeditions, African slaves in Portugal that had been bought from Arabs who had traded earlier with Europeans did not succeed, owing to extensive multitude of languages on the African coast. However, apparently the Portuguese left sick members of their crews behind, who then learned the local languages by immersion. They also managed to sail back to Portugal with members of the royal entourage of the places they ‘discovered.’ These would learn the European language and would serve as interpreters and brokers. (God knows how the arrangements were made for these exchanges of people!) The *Lançados*, *Tangomaos*, and *factors* themselves may be among the volunteers who participated in the earlier expeditions and decided to stay, or jump ship, and go native. Brooks (2003) reports that in Western Africa even their indigenous wives also worked as intermediaries in the Euro-African trade. Berlin (1998) calls the children from these unions ‘Creoles’ and characterizes them as ‘brokers’ in the trade and in the colonization of Africa. During the exploitation/colonization of the continent, which began in the late nineteenth century, the interpreters would become what Samarin (1989) identifies as ‘colonial auxiliaries.’

In short, there emerged a small class of interpreters, on whom trade relied, while the rest of the commoners either worked for their rulers or were busy protecting themselves from the slavers/traders. In one of his letters, Matthews (1788, p. 170) speaks of Black and Mulatto children that were sent to Europe for their education and learned the relevant European language; otherwise they stuck to indigenous customs and languages (1788, p. 165). In Africa and Asia, the Europeans did not trade in the interior, where the slaves originated; they hired Native ‘middlemen’ who traded for them, including the ‘Creole’ children.

I conjecture that the L2 approximations of the European languages diverged by basilectalization, after trade intensified in the eighteenth century and the need for interpreters increased. More and more people who had worked with the preceding categories of interpreters lent their linguistic service. The divergence was consistent with what Chaudenson (1979 ff) characterizes as ‘approximations of approximations’ in the case of creoles. It is thus not surprising that varieties that struck Europeans as divergent from their languages are reported only in the late eighteenth century, not counting the indigenized Portuguese variety spoken in apparently present-day Casamance (in Senegal) reported by La Courbe in 1685.

It is not evident that this was pidgin or creole, though there is a seemingly race-based tendency in creolistics to hastily characterize any structurally divergent nonstandard variety produced by a non-European population as a creole or pidgin. There is also every reason to ask whether the Portuguese varieties spoken by (descendants of) Christianized Natives in India, Malaysia, and Macao should really be called creoles. They did not emerge under the same conditions of language shift as the creoles around the Atlantic and in the Indian Ocean. The same kind of question arises about those lexified by Spanish in the Philippines. Similarities in the conditions of their genesis lie in the overwhelming indigenous majority and in the homogeneity of the substrate. But then, they are reminiscent of the conditions under which the Romance languages emerged when the indigenous Celts were then Romanizing culturally (Mufwene, 2015a). From this perspective, even Irish English could be called a creole; and the term would become historically less informative (Mufwene, 2005).

4. Current contributions and research

The preceding account has been resisted by especially Bakker et al. (2017) and McWhorter (2012, 2018), who defend what DeGraff (2003) characterized as ‘creole exceptionalism.’ These two have claimed that creoles form a typological category of their own as a result of a

break in the transmission of the lexifier, the formation of an antecedent pidgin, and the subsequent emergence of creole by nativization, which can explain the significant divergence of their structures from those of their lexifiers. The differences can be accounted for by invoking substrate influence or other processes specific to ‘creolization’ interpreted as formation of creoles. This is in contrast with Mufwene’s (2000) uniformitarianist argument that there is no specific language-restructuring process that can be identified as ‘creolization.’

Note that the defenders of creole exceptionalism have never articulated what the process or combination of processes may be. The harsh contact conditions of slavery and rigid race and social class segregation under which creoles emerged are only some of the ecological factors that influenced language transmission. They do not entail a break in the transmission of the lexifier, although the latter was continually being restructured. They have never provided a definition of *pidgin*, which must be different from that of a broken language that arises under ecological conditions of sporadic language contact. These did not obtain during the homestead phase of the settlement plantation colonies, when the enslaved populations were a minority, when the enslaved populations grew more by birth than by importation and the children acquired the colonial language natively, and when the contacts between the European colonists and the enslaved populations were definitely not sporadic (Chaudenson, 1979 ff). Residential segregation was instituted typically during the plantation phase, when the non-European population became the (overwhelming) majority (Wood, 1974). I would be remiss to deny that the abrupt shift of vernaculars must have been a traumatic experience for the enslaved, which I doubt was the case for the Christians of Asia who adopted Portuguese or Spanish as their vernacular.

Plag (2008, 2009) attempts to explain the exceptionality of creoles by suggesting that they must be new languages arrested in the interlanguage stage. How this would happen if creoles have pidgin antecedents is not clear. He would obviously have to drop this traditional assumption in order to support his hypothesis. Another issue is whether there is a particular class of features that can be claimed to be peculiarly interlinguistic, though anybody learning an L2 target must go through an interlanguage stage. He still has to answer the question of what is lacking from creole structures that non-creole languages all have.

McWhorter (1998) had attempted to answer this question by positing a set of three features that would define a ‘creole,’ viz., absence of tones, inflections, and derivations. It turns out that many creoles have some of these features. For instance, Papiamentu has tonal lexical and grammatical contrasts, like some African substrate languages (Kouwenberg, 2004). McWhorter ignores an important fact, viz., none of the European lexifiers of creoles is a tonal language. Therefore, creoles appear to have inherited the lack of tones from their lexifiers, from which they have also inherited various other structural features. These include, but are not limited to, their basic word order, the category of adjectives (which is negligible in most Sub-Saharan African languages), the definite article, the basic structure of relative clauses, preposition stranding in questions and relative clauses in English creoles, in addition to the overwhelming proportions of their lexica. Just because creoles reflect substrate influence from their substrate languages does not mean they have inherited nothing or very little of the grammars from their lexifiers. These retentions actually prove that there was no break in the transmission of the lexifier. Mufwene (2015b) shows that the Romance languages may be more divergent from their Vulgar Latin ancestor than the Romance creoles are from their nonstandard lexifiers.

DeGraff (2001) has also shown that Haitian has a very productive derivational system. While many of the derivative morphemes have been inherited from French itself, there are also a number of them that are Haitian innovations. It is true that creoles are inflection-poor; but so

are the non-standard lexifiers of the creoles themselves, although they are not so depleted of inflections. As observed by Chaudenson (1979 ff), morphological impoverishment has been the natural evolutionary trajectory of the creoles' lexifiers. Modern English is inflectionally very poor compared to Old English; so are the Romance languages compared to Vulgar Latin. Creoles brought this evolutionary trajectory to its ultimate end (Chaudenson, 1979 ff), apparently under the influence of some substrate languages (mostly of the Western Kwa family in the case of Atlantic creoles), which have isolating morphosyntax. This is an interesting case of evolution driven by congruence (Corne, 1999).

McWhorter tries to ward off this issue by claiming that the combination of features is attested only in a small subset of creoles that he calls 'creole prototypes,' similar to what Bickerton (1984) had characterized as 'radical,' i.e., the most divergent structurally from their lexifiers. However, typologists have assumed classical, not fuzzy, categories, which require that all members of the class have the features that justify lumping them together. McWhorter's position is undermined by his own suggestion that membership in the class of 'creoles' is a matter of degree, because the vast majority of these vernaculars do not have the stipulated combination of features.

As already noted accurately by the late nineteenth century pioneer creolists, creoles are related structurally among themselves, however in different ways, on the family resemblance model (Mufwene, 2001), for a number of reasons. These include similarities between their lexifiers and similarities between the substrate languages in particular geographical areas, but also similarities in the restructuring processes that occur during the appropriation of a language by a non-heritage population. This is all consistent with a uniformitarian approach, like the same complex algebraic formula that yields different outputs depending on how the values of the variables are set. In the present case, the values include the particular kinds of languages that came in contact at particular points in time and at particular places.

Inspired by Parkvall (2008) and using data from the *Atlas of Pidgins and Creoles* (APiCs, Michaelis et al., 2013), Bakker et al. (2011) used the features which together would display, in neighbornets, how different creoles around the world cluster typologically together and separately from non-creole languages. Note, however, that computer-based analyses of data from the same APiCs by Blasi, Michaelis and Haspelmath (2017) point, instead, in the direction of the hybridity hypothesized by Mufwene (2001 ff), DeGraff (1999a ff), Aboh (2006 ff), and Aboh and DeGraff (2017), disputing the exceptionalists' claims. Likewise, Sherriah et al. (2018) trace most features of Sranan to nonstandard dialects of its English lexifier.

Bakker et al.'s (2011) neighbornets also show relations between creoles that do not correspond to reality. For instance, Bajan (the creole of Barbados) is presented as more closely related to Hawaiian Creole than it is to its Caribbean neighbours and the French creole of Dominica as more closely related to Tok Pisin (an English expanded pidgin of Melanesia) than to its French Caribbean kin. Tok Pisin is heavily influenced by the Austronesian substrate, whereas Dominican Creole is influenced by African substrate languages. Since the Austronesian languages are typologically quite different from the Sub-Saharan African languages, one should wonder what accounts for the connection suggested by Bakker et al.'s neighbornets. It is unfortunate that they overlook the history of population contacts in which the language contacts that produced creoles and pidgins must be grounded.

5. Future directions

Future research on the development of creoles has some other issues to address. To date, knowledge of the nonstandard lexifiers spoken by the European colonists remains limited,

though more research is now underway and much of the scholarship on the dialectology of the European lexifiers is becoming handy. There are few integrated and relatively comprehensive descriptions of creoles' structures, especially from a diachronic perspective. This limitation makes it difficult to determine globally how the competing influences interacted among themselves and how the features selected from diverse sources became integrated into new systems. Few structural facts have been correlated with the conclusions suggested by the sociohistorical backgrounds of individual creoles.

Other issues remain up in the air, for instance, regarding the markedness model that is the most adequate to account for the selection of features into creoles' systems. Can there really be an ecology-independent, universal scale of markedness, which can account for the selection of particular features into the structures of particular creoles or pidgins? And from the point of view of the restructuring of the lexifier and speciation from it, are there 'semi-creoles,' any more than there can be semi-Mulattos or semi-Mestizos, or, for that matter, semi-offspring from particular parents?

The position developed above is in some ways indebted to pioneers such as Hugo Schuchardt (1882) and Louis Hjelmslev (1938), who argued that all languages are mixed to some extent. It also owes a great deal to Bailey and Maroldt (1977) and Schlieben-Lange (1977), who argued, respectively, that Middle English and the Romance languages were creoles, although I have discouraged this interpretation of the conclusions of a uniformitarian approach to the emergence of creoles (Mufwene, 2005).

However, linguists continue to treat creoles as if they were illegitimate offspring of their lexifiers, thus as evolutionary deviations of some sort, which is still a consequence of the uniparental approach to language speciation inherited from the early nineteenth century. Without branches intersecting at the lower levels, the Stammbaum has been interpreted as an illustration of how speciation occurs rather than as the outcome of various processes that produce it (Joseph and Mufwene, 2008). It so happens that genetic linguistics (including historical dialectology) has been providing more and more information about the sociohistorical conditions that have actuated speciation. This information includes population movements (under colonization or imperial expansion), the ensuing new population structures, and language and/or dialect contact. The research area is becoming more explanatory, recognizing contact under varying population structures as an actuator of structural change (Mufwene, 2018a).

The features of the emergent daughter languages or dialects reflect competition and selection (with modification) in the contact ecologies. It appears that there is something to be learned from the ongoing debate on the emergence of creoles regarding the ecological conditions that shaped their structural divergence from their lexifiers. The same approach can also inform us not only about the emergence of the Romance languages and Middle English, but also about the emergence of Old English itself, Scots, and Irish English, as well as the speciation of Proto-Indo-European and Proto-Bantu into their (grand-)daughter languages, among many, many other cases of language differentiation.

Genetic linguists may have to articulate more clearly what the Stammbaum stands for, bearing in mind issues associated with the interpretation of correspondences produced by the comparative method. According to, for instance, Meillet (1900) and Tremblay (2005), one must distinguish between whether the correspondences are inheritances from a common proto-language, whether they are outcomes of homologous evolutions, or whether they have been borrowed from a third-party language that ancestors of the current languages have been in contact with (Mufwene, 2008). There must be some way of discarding the uniparentalist approach to 'normal' language speciation that the Stammbaum seems to instantiate to some scholars.

It will also be useful to revisit the distinction between externally and internally motivated language change. It is not evident that there are fewer changes that are externally motivated in the evolution of non-creole languages. We must bear in mind the fact that the vast majority of structural changes are actuated by the behaviours of speakers, including the inter-dialectal accommodations they make to each other and the fact that we do not copy from each other faithfully (Mufwene, 2001, 2008). Speakers, who are the most direct ecology to languages and filters of all other ecological pressures external to themselves, are actually the ‘unwitting agents of change’ (Mufwene, 2001). Therefore, the changes they trigger in languages are externally motivated (2018b). Internally motivated changes are those that are consequences of earlier changes, such as the fact that the semi-auxiliary (*be*) *going* for FUTURE can coalesce with the complementizer *to* that follows (viz., historically the GOAL preposition *to*) into *gonna* and, in some varieties (such as AAE), the latter has evolved into *gon*. All the debates on the emergence of creoles should prompt us to ask whether some of the questions asked and issues raised do not also apply to the emergence and evolution of non-creole languages.

6. Conclusions

Most of the hypotheses about what creoles and pidgins are and how they have evolved still bear the legacy of conjectures of philologists in the late nineteenth century. Little of what has been learned from especially theories of language learning has led creolists to question some assumptions that underlie the competing hypotheses on their emergence and how they differ from non-creole languages. Increasing knowledge of the history of the colonization of the world by Europeans since the fifteenth century has led us to re-examine the social interactions that produced creoles and pidgins. What we have learned questions not only assumptions traditionally associated with their uniqueness but also those associated with the normalcy of the evolution of non-creole languages. Genetic creolistics appears to be, after all, a part and parcel of genetic linguistics; practitioners in both research areas will undoubtedly benefit from talking across their professional boundaries.

7. Further reading

DeGraff, M., ed. (1999). *Language creation and language change: Creolization, diachrony, and development*. Cambridge, MA: MIT Press.

DeGraff (1999) connects research on the emergence of creoles to that on language acquisition and the emergence of sign language, as well as various aspects of generative linguistics.

Hymes, D., ed. (1971). *Pidginization and creolization of languages*. Cambridge: Cambridge University Press.

Hymes, (1971) is a classic reading that will help the reader assess how much progress has been made to date regarding the emergence of creoles.

Muysken, P. and Smith, N., eds. (1986). *Substrata versus universals in creole genesis*. Amsterdam: John Benjamins.

Muysken and Smith (1986) has played an important role in narrowing the range of interesting hypotheses to pursue and in highlighting the significance of both good historiography and sociohistorical information.

Kouwenberg, S. and Singler, J.V., eds. (2009). *The Handbook of pidgin and creole studies*. Malden, MA: Wiley-Blackwell.

Kouwenberg and Singler (2009) captures more or less the state of the art about creolistics.

Siegel, J. (2008). *The emergence of pidgin and creole languages*. Oxford: Oxford University Press.

This book presents a survey based mostly on creoles and expanded pidgins of the Pacific, which is a useful counterbalance to the predominance of information about those of the Atlantic.

8. Related topics

Mixed languages, social factors

Abbreviations

| | |
|------|-------------------------------------|
| AAE | African-American English |
| AAVE | African-American Vernacular English |
| APiC | Atlas of Pidgins and Creoles |
| HC | Haitian Creole |
| L1 | first language |
| L2 | second language |
| MPP | Maritime Polynesian Pidgin |
| PE | Pidgin English |
| PI | Pidginization Index |
| RH | relexification hypothesis |
| WAPE | West-African Pidgin English |

References

- Aboh, E.O. (2006). The role of the syntax-semantics interface in language transfer. In: C. Lefebvre, L. White and C. Jourdan, eds., *L2 acquisition and creole genesis: Dialogues*. Amsterdam: John Benjamins, pp. 221–252.
- Aboh, E.O. (2009). Competition and selection: That's all! In: E.O. Aboh and N. Smith, eds., *Complex processes in new languages*, 1st ed. Amsterdam: John Benjamins, pp. 317–344.
- Aboh, E.O. (2015). *The emergence of hybrid grammars: Language contact and change*. Cambridge: Cambridge University Press.
- Aboh, E.O. and DeGraff, M. (2017). A null theory of creole formation based on Universal Grammar. In: I. Roberts, ed., *The Oxford handbook of universal grammar*, 1st ed. Oxford: Oxford University Press, pp. 401–458.
- Achebe, C. (1966). *A man of the people*. London: Heinemann Educational Books.
- Adam, L. (1883). *Les idiomes négro-aryens et malayo-aryens: essai d'hybridologie linguistique*. Paris: Maisonneuve.
- Alleyne, M.C. (1980). *Comparative Afro-American: An historical-comparative study of English-based Afro-American dialects of the New World*. Ann Arbor: Karoma.
- Alleyne, M.C. (1996). *Syntaxe historique créole*. Paris: Karthala.
- Allsopp, R. (1977). Africanisms in the idioms of Caribbean English. In: P.F. Kotey and H. Der-Houssikian, eds., *Language and linguistics problems in Africa*, 1st ed. Columbia, SC: Hornbeam Press, pp. 429–441.
- Arends, J. (1986). Genesis and development of the equative copula in Sranan. In: P. Muysken and N. Smith, eds., *Substrata versus Universals in creole genesis: Papers from the Amsterdam creole workshop, April 1985*. Amsterdam: John Benjamins, pp. 57–70.
- Arends, J. (1989). *Syntactic developments in Sranan: Creolization as a gradual process*. PhD. University of Nijmegen.
- Arends, J., ed. (1995). *The early stages of creolization*. Amsterdam: John Benjamins.
- Austen, R.A. and Derrick, J. (1999). *Middlemen of the Cameroons rivers: The Duala and their Hinterland c. 1600 – c. 1960*. Cambridge: Cambridge University Press.

- Bailey, C.-J.N. and Maroldt, K. (1977). The French lineage of English. In: J. Meisel, ed., *Pidgins – creoles – languages in contact*, 1st ed. Tübingen: Gunter Narr, pp. 21–53.
- Baissac, C. (1880). *Etude sur le patois créole mauricien*. Nancy: Imprimerie Berger-Levrault.
- Baker, P. (1982). *The contribution of non-francophone immigrants to the lexicon of Mauritian creole*. PhD. School of Oriental and African Studies, University of London.
- Baker, P. (1993). Australian influence on Malanesian Pidgin English. *Te Reo*, 36, pp. 3–67.
- Baker, P. (1995). Some developmental inferences from the historical studies of pidgins and creoles. In: J. Arends, ed., *The early stages of creolization*, 1st ed. Amsterdam: John Benjamins, pp. 1–24.
- Baker, P. and Come, C. (1986). Universals, substrata and the Indian ocean creoles. In: P. Muysken and N. Smith, eds., *Substrata versus universals in creole genesis*, 1st ed. Amsterdam: John Benjamins, pp. 163–183.
- Baker, P. and Mühlhäusler, P. (1990). From business to pidgin. *Journal of Asian Pacific Communication*, 1, pp. 87–115.
- Bakker, P., Borchsenius, F., Levisen, C. and Sippola, E., eds. (2017). *Creole studies: Phylogenetic approaches*. Amsterdam: John Benjamins.
- Bakker, P., Daval-Markussen, A., Parkvall, M. and Plag, I. (2011). Creoles are typologically distinct from non-creoles. *Journal of Pidgin and Creole Languages*, 26(1), pp. 5–42.
- Baptista, M. (2012). On universal grammar, the bioprogram hypothesis and creole genesis: An interview with Noam Chomsky. *Journal of Pidgin and Creole Languages*, 27, pp. 351–376.
- Beechert, E.D. (1985). *Working in Hawaii: A labor history*. Honolulu: University of Hawaii Press.
- Berlin, I. (1998). *Many thousands gone: The first two centuries of slavery in North America*. Cambridge, MA: Harvard University Press.
- Bertrand-Bocandé, E. (1849). Notes sur la Guinée portugaise ou Sénégal méridionale. *Bulletin de la Société de Géographie*, 12, pp. 57–93.
- Bickerton, D. (1981). *Roots of language*. Ann Arbor: Karoma.
- Bickerton, D. (1984). The language bioprogram hypothesis. *Behavioral and Brain Sciences*, 7, pp. 173–221.
- Bickerton, D. (2008). *Bastard tongues: A trailblazing linguist finds clues to our common humanity in the world's lowliest languages*. New York: Hill and Wang.
- Blasi, D.E., Michaelis, S.M. and Haspelmath, M. (2017). Grammars are robustly transmitted even during the emergence of creole languages. *Nature Human Behavior*, 1, pp. 723–729.
- Bloomfield, L. (1933). *Language*. New York: Holt, Rinehart and Winston.
- Bolingbroke, H. (1796 [1799–1806]). *A Voyage to Demerara*. London: R. Phillips.
- Brasch, W.M. (1981). *Black English and the mass media*. New York: University Press of America.
- Brooks, G.E. (2003). *Eurafricans in Western Africa: Commerce, social status, gender and religious observance from the sixteenth to the eighteenth century*. Athens: Ohio University Press.
- Chaudenson, R. (1979). *Les créoles français*. Paris: Fernand Nathan.
- Chaudenson, R. (1992). *Des îles, des hommes, des langues: essais sur la créolisation linguistique et culturelle*. Paris: L'Harmattan.
- Chaudenson, R. (2001). *Creolization of language and culture*. London: Routledge.
- Chaudenson, R. (2003). *La créolisation: théorie, applications, implications*. Paris: L'Harmattan.
- Chomsky, N. (1986). *Knowledge of language: Its nature, origin, and use*. New York: Praeger.
- Clements, C. (2014). *The linguistic legacy of Spanish and Portuguese: Colonial expansion and language change*. Cambridge: Cambridge University Press.
- Coelho, A.F. (1880–1886). Os dialectos românicos ou neolatinos na Africa, Asia, ae America. *Bolletim da Sociedade de Geografia de Lisboa*, 2, pp. 129–196 (1880–1881), 3, pp. 451–478 (1882), 6, pp. 705–755 (1886).
- Come, C. (1999). *From French to creole: The development of new vernaculars in the French colonial world*. London: University of Westminster.
- Courbe, S. de la. (1685, publ. 1913). *Premier voyage fait à la coste d'Afrique*. Paris: Edouard Champion and Emile Larose.
- Curtin, P.D. (1984). *Cross-cultural trade in world history*. Cambridge: Cambridge University Press.

- Dakhli, J. (2008). *Lingua Franca: Histoire d'une langue métisse en Méditerranée*. Arles, France: Actes Sud.
- DeGraff, M. (1999a). Creolization, language change, and language acquisition: A prolegomenon. In: M. DeGraff, ed., *Language creation and language change: Creolization, Diachrony, and development*, 1st ed. Cambridge, MA: MIT Press, pp. 1–46.
- DeGraff, M., ed. (1999b). *Language creation and language change: Creolization, diachrony, and development*. Cambridge, MA: MIT Press.
- DeGraff, M. (2001). On the origin of creoles: A Cartesian critique of neo-Darwinian linguistics. *Linguistic Typology*, 5, pp. 213–310.
- DeGraff, M. (2002). Relexification: A reevaluation. *Anthropological Linguistics*, 44, pp. 321–414.
- DeGraff, M. (2003). Against creole exceptionalism: Discussion note. *Language*, 79, pp. 391–410.
- DeGraff, M. (2005). Linguists' most dangerous myth: The fallacy of creole exceptionalism. *Language in Society*, 34, pp. 533–591.
- DeGraff, M. (2009). Language acquisition in creolization and, thus, language change: Some Cartesian-uniformitarian boundary conditions. *Language and Linguistics Compass*, 3–4, pp. 888–971.
- D'Eloia, S. (1973). Issues in the analysis of Negro nonstandard English. Review of Dillard (1972). *Journal of English Linguistics*, 7, pp. 87–106.
- Dillard, J.L. (1972). *Black English: Its history and usage in the United States*. New York: Random House.
- Dillard, J.L. (1992). *A history of American English*. London and New York: Longman.
- Drechsel, E. (2014). *Language contact in the early colonial Pacific: Maritime Polynesian Pidgin before Pidgin English*. Cambridge: Cambridge University Press.
- Dunn, R.S. (1972). *Sugar and slaves: The rise of the planter class in the English West Indies, 1624–1713*. Chapel Hill: The University of North Carolina Press.
- Faine, J. (1937). *Philologie créole: études historiques et étymologiques sur la langue créole d'Haïti*. Port-au-Prince: Imprimerie de l'Etat.
- Fayer, J. (2003). African interpreters in the Atlantic slave trade. *American Anthropologist*, 45, pp. 281–295.
- Féral, C. de. (1989). *Pidgin-English du Cameroun*. Paris: Peteers.
- Gilbert, G. (1986). The language bioprogram hypothesis: Déjà vu? In: P. Muysken and N. Smith, eds., *Substrata versus universals in creole genesis*, 1st ed. Amsterdam: John Benjamins, pp. 15–24.
- Goodman, M. (1964). *A comparative study of Creole French dialects*. The Hague: De Gruyter Mouton.
- Goodman, M. (1985). Review article on Bickerton's (1981) roots of language. *International Journal of American Linguistics*, 51, pp. 109–137.
- Gonzales, A.E. (1922). *The black border: Gullah stories of the Carolina coast (with a Glossary)*. Columbia, SC: The State Company.
- Gray, E.G. and Fiering, N., eds. (2000). *The language encounter in the Americas 1492–1800*. New York: Berghahn Books.
- Hall, R.A. Jr. (1958). Creole languages and genetic relationships. *Word*, 14, pp. 367–373.
- Hall, R.A. Jr. (1966). *Pidgin and creole languages*. Ithaca, NY: Cornell University Press.
- Hancock, I.F. (1986). The domestic hypothesis, diffusion and componentiality: An account of Atlantic Anglophone creole origins. In: P. Muysken and N. Smith, eds., *Substrata versus universals in creole genesis*, 1st ed. Amsterdam: John Benjamins, pp. 71–102.
- Hjelmlev, L. (1938). Etudes sur la notion de parenté linguistique. Première étude: Relations de parenté de langues créoles. *Revue des Etudes Indo-Européennes*, 1, pp. 271–286.
- Holm, J. (1988). *Pidgins and creoles. Vol. 1: Theory and structure*. Cambridge: Cambridge University Press.
- Huber, M. (1999). Atlantic creoles and the Lower Guinea Coast: A case against Afrogenesis. In: M. Huber and M. Parkvall, eds., *Spreading the word: The issue of diffusion among the Atlantic creoles*, 1st ed. London: University of Westminster Press, pp. 81–110.
- Hymes, D., ed. (1971). *Pidginization and creolization of languages*. Cambridge: Cambridge University Press.
- Jones, H. and Morton, R.L. ([1724] 1956). *The present state of Virginia: From whence is inferred a short view of Maryland and North Carolina*. Chapel Hill: University of North Carolina Press.

- Joseph, B. and Mufwene, S.S. (2008). Parsing the evolution of language. *Science*, 320, p. 446.
- Jourdan, C. (2009). Complexification or regularization of paradigms: The case of prepositional verbs in Solomon Islands Pijin. In: E.O. Aboh and N. Smith, eds., *Complex processes in new languages*, 1st ed. Amsterdam: John Benjamins, pp. 159–170.
- Keesing, R.M. (1988). *Melanesian Pidgin and the Oceanic substrate*. Stanford: Stanford University Press.
- Kennedy, D. (2013). *The last blank spaces: Exploring Africa and Australia*. Cambridge, MA: Harvard University Press.
- Kouwenberg, S. (1994). *A grammar of Berbice Dutch Creole*. Berlin: De Gruyter Mouton.
- Kouwenberg, S. (2004). The grammatical function of Papiamentu tone. *Journal of Portuguese Linguistics*, 3, pp. 55–69.
- Kouwenberg, S. and Singler, J.V., eds. (2009). *The handbook of pidgin and creole studies*. Malden, MA: Wiley-Blackwell.
- Krapp, G.P. (1924). The English of the Negro. *The American Mercury*, 2, pp. 190–195.
- Kulikoff, A. (1986). *Tobacco and slaves: The development of southern cultures in the Chesapeake, 1680–1800*. Chapel Hill: University of North Carolina Press.
- Kurath, H. (1928). The origin of dialectal differences in spoken American English. *Modern Philology*, 25, pp. 385–395.
- Lawrence, B., Osborn, E.L. and Roberts, R.L., eds. (2006). *Intermediaries, interpreters, and clerks: African employees in the making of colonial Africa*. Madison: University of Wisconsin Press.
- Lee, K.M. (2014). Language and conquest: Tupi-Guarani expansion in the European colonization of Brazil and Amazonia. In: S.S. Mufwene, ed., *Iberian imperialism and language evolution in Latin America*. Chicago: University of Chicago Press, pp. 144–167.
- Lefebvre, C. (1998). *Creole genesis and the acquisition of grammar: The case of Haitian Creole*. Cambridge: Cambridge University Press.
- Lefebvre, C. (2004). *Issues in the study of pidgin and creole languages*. Amsterdam: John Benjamins.
- Lefebvre, C., ed. (2011). *Creoles, their substrates, and language typology*. Amsterdam: John Benjamins.
- Lefebvre, C., White, L. and Jourdan, C., eds. (2006). *L2 acquisition and creole genesis*. Amsterdam: John Benjamins.
- MacSwan, J. (2017). A multilingual perspective on translanguaging. *American Educational Research Journal*, 54, pp. 167–201.
- Matthews, J. (1788). *A voyage to the river Sierra-Leone on the coast of Africa*. London: B. White and Son; J. Sewell.
- McDavid, R. Jr. (1950). Review of Lorenzo Dow Turner's Africanisms in the Gullah dialect. *Language*, 26, pp. 323–333.
- McDavid, R. Jr. and McDavid, V. (1951). The relationship of the speech of the American Negroes to the speech of Whites. *American Speech*, 26, pp. 3–17.
- McWhorter, J.H. (1998). Identifying the creole prototype: Vindicating a typological class. *Language*, 74, pp. 788–818.
- McWhorter, J.H. (2012). Case closed? Testing the feature pool hypothesis. *Journal of Pidgin and Creole Languages*, 27, pp. 171–182.
- McWhorter, J.H. (2018). *The creole debate*. Cambridge: Cambridge University Press.
- Meillet, A. (1900). Note sur une difficulté générale de la grammaire comparée. Reprinted in (1982). *Linguistique historique et linguistique générale*. Genève, Slakline, and Paris: Champion, pp. 36–43.
- Mello, H. (2014). African descendants' rural vernacular Portuguese and its contribution to understanding the development of Brazilian Portuguese. In: S.S. Mufwene, ed., *Iberian imperialism and language evolution in Latin America*, 1st ed. Chicago: University of Chicago Press, pp. 168–185.
- Meyerhoff, M. (2009). Forging Pacific pidgin and creole syntax: Substrate, discourse, and inherent variability. In: S. Kouwenberg and J.V. Singler, eds., *The handbook of pidgin and creole studies*, 1st ed. Malden, MA: Wiley-Blackwell, pp. 48–73.
- Michaelis, S.M., ed. (2008). *Roots of creole structures weighing the contribution of substrates and superstrates*. Amsterdam: John Benjamins.

- Michaelis, S.M., Maurer, P., Haspelmath, M. and Huber, M., eds. (2013). *Atlas of pidgin and creole languages, Vols. 1–4*. Oxford: Oxford University Press.
- Mufwene, S.S. (1986). The universalist and substrate hypotheses complement one another. In: P. Muysken and N. Smith, eds., *Substrata versus universals in creole genesis*, 1st ed. Amsterdam: John Benjamins, pp. 129–162.
- Mufwene, S.S. (1990). Transfer and the substrate hypothesis in creolistics. *Studies in Second Language Acquisition*, 12, pp. 1–23.
- Mufwene, S.S. (1991). Pidgins, creoles, typology, and markedness. In: F. Byrne and T. Huebner, eds., *Development and structures of creole languages: Essays in honor of Derek Bickerton*. Amsterdam: John Benjamins, pp. 123–143.
- Mufwene, S.S., ed. (1993). *Africanisms in Afro-American language varieties*. Athens: University of Georgia Press.
- Mufwene, S.S. (1996). The founder principle in creole genesis. *Diachronica*, 13, pp. 83–134.
- Mufwene, S.S. (1997). Jargons, pidgins, creoles, and koinés: What are they? In: A.K. Spears and D. Winford, eds., *Pidgins and creoles: Structure and status*, 1st ed. Amsterdam: John Benjamins, pp. 35–70.
- Mufwene, S.S. (2000). Some sociohistorical inferences about the development of African-American English. In: S. Poplack, ed., *The English history of African-American English*, 1st ed. Oxford: Blackwell, pp. 233–263.
- Mufwene, S.S. (2001). *The ecology of language evolution*. Cambridge: Cambridge University Press.
- Mufwene, S.S. (2005). *Créoles, écologie sociale, évolution linguistique*. Paris: L'Harmattan.
- Mufwene, S.S. (2008). *Language evolution: Contact, competition and change*. London: Continuum.
- Mufwene, S.S. (2010). SLA and the emergence of creoles. *Studies in Second Language Acquisition*, 32, pp. 359–400.
- Mufwene, S.S. (2014). Globalisation économique mondiale des XVIIe – XVIIIe siècles, émergence des créoles, et vitalité langagière. In: A. Carpooran, ed., *Langues créoles, mondialisation et éducation*, 1st ed. St. Louis, Mauritius: Creole Speaking Unit, pp. 23–79.
- Mufwene, S.S. (2015a). L'émergence des parlers créoles et l'évolution des langues romanes: Faits, mythes et idéologies. *Etudes Créoles*, 33, pp. 1–29.
- Mufwene, S.S. (2015b). The emergence of African American English: Monogenetic or polygenetic? Under how much substrate influence? In: S. Lanehart, ed., *The Oxford handbook of African American language*, 1st ed. Oxford: Oxford University Press, pp. 57–84.
- Mufwene, S.S. (2017). Language vitality: The weak theoretical underpinnings of what can be an exciting research area. *Language*, 93, e202–e223.
- Mufwene, S.S. (2018a). Language contact and evolutionary linguistics: An African(ist)'s and creolist's perspective. In: R. Mesthrie and D. Bradley, eds., *Dynamics of language: Plenary and focus lectures 20th international congress of linguists*, Cape Town, South Africa: University of Cape Town Press, pp. 36–51.
- Mufwene, S.S. (2018b). Language evolution from an ecological perspective. In: A. Fill and H. Benz, eds., *The handbook of ecolinguistics*, 1st ed. London: Routledge, pp. 73–88.
- Mühlhäusler, P. (1981). The development of the category of number in Tok Pisin. In: P. Muysken, ed., *Generative studies on creole languages*, 1st ed. Dordrecht: Foris, pp. 35–84.
- Mühlhäusler, P. (1997). *Pidgin and creole linguistics (revised and expanded edition)*. London: University of Westminster Press.
- Muysken, P. and Smith, N., eds. (1986). *Substrata versus universals in creole genesis*. Amsterdam: John Benjamins.
- Newitt, M. (2010). *A history of Portuguese overseas expansion: 1400–1668*. London: Routledge.
- Northrup, D. (2009). *Africa's discovery of Europe 1450–1850*, 2nd ed. New York: Oxford University Press.
- Ostler, N. (2005). *Empires of the word: A language history of the world*. New York: Harper Collins.
- Parkvall, M. (2008). The simplicity of creoles in a cross-linguistic perspective. In: M. Miestamo, K. Sinnemäki, and F. Karlsson, eds., *Language complexity*, 1st ed. Amsterdam: John Benjamins, pp. 265–285.

- Plag, I. (2008). Creoles as interlanguages: Syntactic structures. *Journal of Pidgin and Creole Languages*, 23, pp. 307–328.
- Plag, I. (2009). Creoles as interlanguages: Phonology. *Journal of Pidgin and Creole Languages*, 24, pp. 119–138.
- Poplack, S., ed. (2000). *The English history of African-American English*. Oxford: Blackwell.
- Poplack, S. and Tagliamonte, S. (2001). *African American English in the diaspora*. Malden, MA: Blackwell.
- Prudent, L-F. (1980). *Des baragouins à la langue antillaise: Analyse historique et sociolinguistique du discours sur le créole*. Paris: Editions Caribéennes.
- Rawley, J.A. (1991). Slave trade. In: E. Foner and J.A. Garraty, eds., *The reader's companion to American history*, 1st ed. Boston: Houghton Mifflin, pp. 994–995.
- Rickford, J.R. (1998). The creole origins of African American Vernacular English: Evidence from copular absence. In: S.S. Mufwene, J.R. Rickford, G. Bailey, and J. Baugh, eds., *African-American English: Structure, history, and use*, 1st ed. London: Routledge, pp. 154–200.
- Roberts, S.J. (1998). The role of diffusion in the genesis of Hawaiian creole. *Language*, 74, pp. 1–39.
- Roberts, S.J. (2004). *The emergence of Hawai'i creole English in the early 20th century: The sociohistorical context of creole genesis*. PhD. Stanford University.
- Samarin, W.J. (1989). *The Black man's burden: African colonial labor on the Congo and Ubangi Rivers, 1880–1900*. Boulder, CO: Westview Press.
- Sankoff, G. (1979). The genesis of a language. In: K.C. Hill, ed., *The genesis of language*, 1st ed. Ann Arbor: Karoma, pp. 23–47.
- Sankoff, G. (1993). Focus in Tok Pisin. In: F. Byrne and D. Winford, eds., *Focus and grammatical relations in creole languages*, 1st ed. Amsterdam: John Benjamins, pp. 117–140.
- Sankoff, G. and Brown, P. (1976). The origins of syntax in discourse: A case study of Tok Pisin relatives. *Language*, 52, pp. 631–666.
- Schlieben-Lange, B. (1977). L'origine des langues romanes: Un cas de créolisation? In: J. Meisel, ed., *Pidgins – creoles – languages in contact*, 1st ed. Tübingen: Narr, pp. 81–101.
- Schneider, E.W. (1989). *American earlier Black English: Morphological and syntactic variables*. Tuscaloosa: University of Alabama Press.
- Schuchardt, H. (1882). Zur afrikanischen Sprachmischung. *Das Ausland*, 55, pp. 867–869.
- Schuchardt, H. (1909). Die lingua franca. *Zeitschrift für Romanische Philologie*, 33, pp. 441–461.
- Schuchardt, H. (1914). *Die Sprache der Saramakkaneger in Surinam*. Amsterdam: Johannes Müller.
- Translated in Schuchardt, H. (1979). *The ethnography of variation: Selected writings on pidgins and creoles*, trans. & ed. by T.L. Markey. Ann Arbor: Karoma, pp. 73–108.
- Sherriah, A.C., Devonish, H., Thomas, Ewart A.C. and Creanza, N. et al. (2018). Using features of a creole language to reconstruct population history and cultural evolution: Tracing the English origins of Sranan. *Philosophical Transactions of the Royal Society B*, 373, 20170055.
- Siegel, J. (2006). Links between SLA and creole studies: Past and present. In: C. Lefebvre, L. White and C. Jourdan, eds., *L2 acquisition and creole genesis*, 1st ed. Amsterdam: John Benjamins, pp. 15–46.
- Siegel, J. (2008). *The emergence of pidgin and creole languages*. Oxford: Oxford University Press.
- Singler, J.V. (1988). The homogeneity of the substrate as a factor in pidgin/creole genesis. *Language*, 64, pp. 27–51.
- Singler, J.V. (1996). Theories of creole genesis, sociohistorical considerations, and the evaluation of evidence: The case of Haitian Creole and the relexification hypothesis. *Journal of Pidgin and Creole Languages*, 11, pp. 185–230.
- Singler, J.V. (2009). The sociohistorical context of creole genesis. In: S. Kouwenberg and J.V. Singler, eds., *The handbook of pidgin and creole studies*, 1st ed. Malden, MA: Wiley-Blackwell, pp. 332–358.
- Sylvain, S. (1936). *Le créole haïtien: morphologie et syntaxe*. Wetteren, Belgium: Imprimerie De Meester.
- Thompson, R.W. (1961). A note on some possible affinities between the creole dialects of the old world and those of the new. In: R.B. LePage, ed., *Creole language studies II*, 1st ed. London: Macmillan, pp. 107–113.

- Thomason, S.G. (2001). *Language contact: An introduction*. Washington, DC: Georgetown University Press.
- Thomason, S.G. and Kaufman, T. (1988). *Language contact, creolization, and genetic linguistics*. Berkeley: University of California Press.
- Tremblay, X. (2005). Grammaire comparée et grammaire historique: quelle réalité est reconstruite par la grammaire comparée? In: G. Fussman, J. Kellens, H-P. Francfort and X. Tremblay, eds., *Aryas, Aryens et Iraniens en Asie Centrale*, 1st ed. Paris: Edition-Diffusion de Boaccard, pp. 33–180.
- Turner, L.D. (1949). *Africanisms in the Gullah Dialect*. Chicago: University of Chicago Press.
- Van Dyke, P.A. (2005). *The canton trade: Life and enterprise on the China coast, 1700–1845*. Hong Kong: Hong Kong University Press.
- Vigouroux, C.B. and Mufwene, S.S. (2013). “Le français tirailleurr” (FT): A historical assessment of its emergence. Paper presented at *The international conference on colonial and postcolonial linguistics*, University of Bremen, Germany.
- Vinson, J. (1882). Créole. *Dictionnaire des Sciences Anthropologiques et Ethnologiques*. Paris.
- Vinson, J. (1888). La linguistique. *La grande encyclopédie*, 22. Paris: Lamirault et cie, pp. 286–296.
- Walicek, D.E. (2007). Farther South: Speaking American, the language of migration in Samaná. In: J. Adams, M.P. Bibler and C. Accilien, eds., *Just below South: Intercultural performance in the Caribbean and the US South*, 1st ed. Charlottesville, VA: New World Studies, pp. 95–120.
- Whinnom, K. (1965). The origin of the European-based pidgins and creoles. *Orbis*, 14, pp. 509–527.
- Whinnom, K. (1971). Linguistic hybridization and the ‘special case’ of pidgins and creoles. In: D. Hymes, ed., *Pidginization and creolization of languages*, 1st ed. Cambridge: Cambridge University Press, pp. 91–115.
- Wolf, E.R. (1982). *Europe and the people without history*. Berkeley: University of California Press.
- Wood, P. (1974). *Black Majority: Negroes in colonial South Carolina from 1670 through the Stono Rebellion*. New York: Alfred A. Knopf.
- Wylie, J. (1995). The origins of Lesser Antillean French creole: Some literary and lexical evidence. *Journal of Pidgin and Creole Languages*, 10, pp. 77–127.
- Yakpo, K. (2009). Complexity revised: Pichi (Equatorial Guinea) and Spanish in contact. In: N. Faraclas and T.B. Klein, eds., *Simplicity and complexity in creoles and pidgins*. 1st ed. London: Battlebridge Publications, pp. 1184–1216.
- Yakpo, K. (2017). Unity in diversity: The homogeneity of the substrate and the grammar of space in the African and Caribbean English-lexifier creoles. In: C. Cutler, Z. Vrzic and P. Angermeyer, eds., *Language contact in Africa and the African diaspora in the Americas: In honor of John V. Singler*, 1st ed. Amsterdam: John Benjamins, pp. 226–251.

Mixed Languages

Carmel O'Shannessy

1. Introduction and definitions

Mixed Languages combine a significant portion of lexical and/or grammatical material from more than one source language, largely intact, in a conventionalized systematic way, and therefore do not unambiguously have only one parent language, but two or more (Thomason and Kaufman, 1988, p. 12; Matras and Bakker, 2003b). Mixed Languages are distinguished from other types of contact languages, e.g., creoles, by retaining subsystems of the source languages largely intact (Matras and Bakker, 2003b; Matras, 2003a; Thomason, 1997b, 2003). They are stable, highly conventionalized ways of speaking that show a very consistent way of combining source language elements across speaker, time and context. For example, a Mixed Language may combine the grammar from one language with the lexicon from another (Bakker, 1997). There is often regularization, adaptation, and innovation of features from the source languages also, to differing extents, yet other parts of the grammatical and/or lexical subsystems remain mostly unchanged. Many are or were one of the first languages of at least one generation of their speakers, or a register of the speakers' first language (Matras and Bakker, 2003b, p. 7; Mous, 2003b). While most Mixed Languages combine features from two source languages, there are instances of Mixed Languages with significant influence from more than two source languages (Ansaldò, 2009; O'Shannessy, 2013), but in each case two of the languages pattern together in some way, making the dominant split between two sets of patterning, if not two languages. Language mixing combinations that do not conventionalize and diffuse through a community of speakers are not part of the category of Mixed Languages, nor are languages that can still be classified in a genetic tree as deriving from a single parent (Matras and Bakker, 2003b; Thomason, 2003), but there can be disagreement as to how to make these distinctions, and the difference may be a matter of degree (Thomason, 1995, 2003; Auer, 2014; Myers-Scotton, 2003).

Mixed Languages have different social histories from creoles. Yet the social histories and mechanisms that led to many of the Mixed Languages may be similar to those in some other contexts of bilingualism and multilingualism that do not have such systematized outcomes, or to the same degree. For example, situations of heavy borrowing or convergence, or language mixing by bilinguals, may appear similar to Mixed Languages, but if researchers agree that

a single parent language can be identified then they are not included in the Mixed Language category in this chapter, interesting though they are.

Common to many Mixed Languages is an intensive language contact situation at the time of genesis. Bilingualism and/or language endangerment or shift have played a role in each at some point in their histories. A Mixed Language emerges within a single community of speakers who already have a shared speech repertoire. This is in contrast to creoles, which emerge in situations where speakers need to communicate with others who do not share their language repertoire. The number of source languages is a heuristic, but not a defining difference between a creole and a Mixed Language, rather the difference is in the type of combination of subsystems of the sources in the new language, as well as the social history of their emergence.

Mixed Languages are important because they lead to richer understandings of the mechanisms and outcomes of contact-induced language change in different types of context, and they allow us to examine whether unusual outcomes result from unusual or commonly occurring processes.

Terms for Mixed Languages in the literature include bilingual mixed languages (Thomason, 2001, p. 157), stable mixed languages (Matras and Bakker, 2003b), and split languages (Myers-Scotton, 2003). These terms refer to an overarching type that contains several different patterns of language combination, the outcomes of differing language contact configurations. Specific types of combination have been called intertwined languages or Lexicon-Grammar (L-G) mixed languages (Bakker and Mous, 1994a; Bakker, 2003), Noun-Verb Mixed Languages (Matras, 2003b; Muysken, 2007) and converted languages (Bakker, 2003).

Distinctions are made on the basis of structure, social history, and function, both between Mixed Languages and other types of contact language, and between Mixed Languages and other languages that are categorized as having a single genetic parent but have undergone heavy borrowing, a mechanism for bringing lexicon from one language into another. Matras (2009, p. 291) suggests that Mixed Languages perform different functions, and can be positioned on a 'functional cline' ranging from secret languages that deliberately exclude hearers other than those who know the language, to languages used for everyday communication. In every case, the new language functions as a way of signalling the distinct identity or in-group solidarity of its speakers, often in contrast to speakers of a socially dominant language.

2. Historical overview

Until recently, Mixed Languages received much less attention than other types of contact language such as pidgins and creoles, but over the last four decades, research on Mixed Languages has brought forward increasing amounts of documentation and theorizing about their identification, their status as a typological class, and the mechanisms that lead to their emergence. The languages show diversity in the ways they combine their sources to form a new way of speaking. For many of them the sociolinguistic situation at the time of emergence was not well known, making it difficult to trace the exact social and linguistic mechanisms involved, but for others there is a considerable amount of information.

In the late nineteenth century there was a prevailing view that it was not possible for a language to have the kind of contact-induced influence seen in Mixed Languages (Müller, 1873, p. 86, Bakker and Mous, 1994a, p. 5). Some of the history of thinking about contact influences is given in Thomason and Kaufman (1988). Case studies of some of the languages detailed in this chapter provided evidence that languages could combine a significant amount of elements from more than one source, but with mixing patterns and social histories of speakers that differed from, for instance, creole languages, e.g., Bakker and Mous (1994b), Thomason (1997a) and Matras and Bakker (2003a).

This chapter will not include every candidate language that could be argued to fit the Mixed Language category, but those for which there is the most documentation and discussion. There are several fairly current reviews of Mixed Languages (e.g., Bakker and Mous, 1994b; Bakker, 2003; Bakker and Matras, 2013; Matras, 2009; Thomason, 1997a; Auer, 2014; Muysken, 2007; Meakins, 2013, 2016; Muysken, 2013). This chapter only includes discussion of lexical and morphosyntactic properties. However, there is a growing body of work on the phonologies of Mixed Languages (e.g., Jones and Meakins, 2013; Jones, Meakins and Buchan, 2011; Stewart, 2014, 2015, 2018a, 2018b; Stewart et al., 2018; Rosen, 2006, 2007; Bundgaard-Nielsen and O'Shannessy, 2015, 2018; van Gijn, 2009; Stewart and Meakins, Forthcoming). In phonology, also, the languages show both similarities and differences in the ways the source languages combine in the new system.

3. Critical issues and topics

The most critical issue is whether a way of speaking is best categorized as a Mixed Language, in keeping with the preceding definition. This is tightly bound to the questions of the types of language combination that are attested, and the mechanisms of their emergence. A challenge is that languages that show a significant portion of their subsystems from two or more languages, largely intact, combine them in diverse ways, as a result of different kinds of social motivations and language typologies, yet a theory needs to identify what is common to them all. Nevertheless, most researchers agree on the main types attested (see Section 4).

Questions about the mechanisms of emergence of Mixed Languages include whether the processes are similar to those of other language contact phenomena, with considerable discussion of whether they can emerge from the bilingual processes of code-switching. Most researchers agree that at least some of the processes also occur in other language contact settings. Some theories view both the outcomes and mechanisms as distinct (Matras, 2003b, 2009; Bakker, 2003). In contrast, the continuum or gradient view sees them as the extreme result of processes that appear in other contact contexts, differing not in quality but in degree (Auer, 1999, 2014; Thomason, 1995, 1997b, 2001, 2003; Myers-Scotton, 2003). A related issue is about the autonomy of a language – whether it functions independently from its sources, which may still be spoken.

Questions embedded in these theories are the degree to which the processes are deliberate or conscious at some level on the part of the speakers (Matras, 2003b; Thomason, 2003), and the possibility that different processes may operate at different times in the emergence of a Mixed Language (Mous, 2003a). Linked to these is the question of the role of different age groups of speakers at important times in the development of a language, for example, the role that adults or children play (O'Shannessy, 2012, Forthcoming).

A final issue is the functions that a Mixed Language performs for its speakers, which may range from being secret, exclusive languages, to those of everyday communication, where they typically also indicate the identity of their speakers (Matras, 2009, p. 291).

These issues will be further elaborated in the following section for each language described.

4. Current contributions and research

4.1 *Types of language combination*

The languages are grouped here based on the type of language combination, even though the languages in each type of combination may have different social histories, indicating some

of the complexity of categorization. For each language brief details of the circumstances of its emergence are given. A detailed discussion of mechanisms of emergence that draws on all types is given in Section 4.2.

4.1.1 *Lexicon-Grammar (L-G) Mixed Languages, or intertwining languages*

The most common pattern of combination is where the lexicon or word shapes are sourced from one language and the grammatical system from another, and within this category the combinations and mechanisms involved vary. The combination has been called 'language intertwining,' and the language type 'Lexicon-Grammar (L-G) mixed languages' by Bakker (2003) and Bakker and Mous (1994a). This pattern has been called a 'prototype' because it is the most common pattern, but see Matras (2003b) and Section 3 for discussion of the notion of a prototype.

Michif

An example of a Lexicon-Grammar mixed language is Michif (Bakker, 1994, 1997; Bakker and Papen, 1997), spoken in parts of Canada and North Dakota in the USA by members of the Métis nation, who have Native American and European ancestry. Michif emerged when French-speaking fur traders and Cree (Algonquian) women married in the early 1800s. Their children grew up bilingually in Cree and French, and presumably they formed Michif, although chronological details are not known.

Michif combines mostly Cree grammar with mostly Canadian French lexicon and nominal morphology, but demonstratives and interrogatives are Cree (Bakker and Papen, 1997, p. 295). Cree is retained partly because it is a polysynthetic language, with many grammatical functions realized in verbal morphology, so that it would be difficult to segment the verbal structure and insert words from another language. Bakker argues that another reason that Cree verb grammar is retained in Michif is that at the time of emergence, Cree-speaking mothers would have spoken predominantly Cree to their children, who were then Cree dominant. In example (1), elements from Cree are in italics and from French in plain font.

1. *maci-kîsikâ-w* pas moyên *si-misk-ahk* son shack *waisi-n*
 bad-weather-3SG.INAN no way COMP-find-he.it his cabin be.lost-he
 'A storm came up, he got lost, he couldn't find his way back to the cabin.'

(Adapted from Bakker, 1997, p. 6)

Michif is now spoken by about 450 people and is highly endangered (Gillon and Rosen, 2018). For many years now, most speakers have spoken Michif regardless of whether they also speak French and/or Cree independently.

Media Lengua

Media Lengua, developed by young Quechua men who moved to Spanish-dominant cities in Central Ecuador for work in the 1920–1940s, combines Spanish word-shapes with Quechua grammar (Muysken, 1981, 1994, 1997). In example (2) elements from Quechua are in italics

and Spanish in plain font. Line a) gives the Media Lengua forms, line b) the Quechua forms, and line c) the Spanish forms.

2. a. unu fabur-ta pidi-nga-bu bini-xu-ni. (Media Lengua)
 one favour-ACC ask-NOM-BEN come-PROG-1
 ‘I come to ask a favour.’
 b. *Shuk fabur-da mana-nga-bu shamu-xu-ni.* (Quechua)
 c. Vengo para pedir un favor. (Rural Spanish)
 (Adapted from Muysken, 1997, p. 365)

Muysken (1981, 1994, 1997) identifies the process involved in the emergence of Media Lengua as one of conscious relexification, in which speakers replaced lexical items from Quechua with Spanish word shapes, but retained the semantic and grammatical specifications from Quechua, and also integrated the Spanish words with Quechua phonology. The Spanish vocabulary was brought in almost completely, including basic vocabulary – 90% of words are from Spanish. But the Spanish words did not replace Quechua words, rather Spanish ‘phonological shapes’ were ‘grafted onto Quechua lexical entries’ (Muysken, 1997, p. 379). In addition, some parts of the Quechua grammatical system have undergone ‘compromise restructuring’ under the influence of Spanish, e.g., for deictic pronouns (Muysken, 1997, p. 393). Media Lengua emerged because Quechua speakers working in the city and acculturating to Spanish no longer identified completely with Quechua culture nor urban Spanish culture (Muysken, 1997, p. 376), and Media Lengua signalled an identity distinct from both of those. Media Lengua also has some structural features that are neither from Quechua nor Spanish.

Para-Romani varieties

Anglo-Romani is spoken by the Roma in Britain as an indicator of community belonging and shared culture and experience (Matras, 2009, p. 395). It consists of Romani words inserted into an English grammatical frame, taking English affixes (Matras, 2009, p. 295; Bakker, 2003, p. 112). In example (3) elements from Romani are in italics and from English in plain font.

3. Don’t *pukker* your *nav*, *mong* the *gaera* how much *luvva* *duvya* is.
 don’t say your name, ask the man how much money this is
 ‘Don’t say your name, ask the man how much money this is.’
 (Adapted from Matras, 2010, p. 14)

It appears that there was a time when Anglo-Romani was spoken in passages of substantial length in conversations, as in an example from Sampson 1930 (Bakker, 2003, p. 112). But now it has a more restricted function, and fewer Romani words may be inserted into an other-wise English clause to achieve the purposes of indicating an in-group tone, or a specific Romani orientation to the discourse. Matras (2010, p. 134) calls this an ‘emotive mode’ that expresses an emotionally engaged orientation to the topic, drawing on shared socio-cultural history and experiences. There is a clear link between a more even division of Romani and English elements in earlier ways of speaking, and the more emblematic use of Romani more recently. Speakers in recent times may not know Romani beyond the lexicon they use in Anglo-Romani.

Matras (2010) contains a detailed history of Romani languages and varieties and the development of mixed languages such as Anglo-Romani.

There is documentation of similar types of combination of Romani with the majority languages in areas where Roma live, collectively referred to as Para-Romani varieties, for a detailed overview see Matras (2009, p. 294, 2010, pp. 9–12 and references therein).

Lekoudesch

In south-west Germany and surrounding areas, Jewish cattle-traders inserted vocabulary from Ashkenazic Hebrew into a rural Judeo-German variety, resulting in a variety known as Lekoudesch (which also has other names, see (Matras, 2009, p. 292)), which they used as a secret, exclusionary language. Despite the genocide in Germany during World War II, knowledge of Lekoudesch was retained by some non-Jewish men who had worked with Jewish farmers as young men. Example (4) is from that group. Elements from Hebrew are in italics and from Swabian German in plain font.

4. Der *schäfft* de ganze *Jomm* im *Uschpiss*, un duat immer
he sits the whole day in.DEF pub, and does always
harme schasskenna und *meloucht lou*
much drinking and works not
'He sits all day in the pub, and drinks a lot, and doesn't work.'

(Adapted from Matras (2009, p. 292))

Lekoudesch came to indicate a shared group identity among the cattle-traders, and to disguise some elements from those not part of the group. The speakers might not have been fluently bilingual in both source languages, but knew a solid store lexical material, or 'lexical reservoir' (Matras, 2009, p. 292) to use for these functions. Lekoudesch could be thought of as a specific register rather than a language used for a large range of functions.

Jenisch

Also in southwest Germany, Switzerland and Austria, Jenisch is spoken by travelling families in service occupations. Probably formed in the seventeenth century, in Jenisch lexical elements from Romani languages and Hebrew are inserted into German varieties. In one village 30% of the lexicon is from Romani and 10–15% from Hebrew (Matras, 2009, p. 292). Jenisch is a vocabulary of several hundred lexical items that speakers use to make meaning opaque to others and indicate in-group solidarity (Matras, 2009, p. 292). Bilingualism plays only a minor role for Jenisch speakers, but the lexical reservoir is transmitted across generations. In example (5) Jenisch elements are in italics and German in plain font.

5. *Nasch* zum *Ruach* und *mang* mir *Maro*
Go to.DEF farmer and ask me.DAT bread
'Go to the farmer and get me some bread.'

(Matras, 2009, p. 292)

Both Lekoudesch and Jenisch, and to some extent Anglo-Romani, function as strategies for disguising meaning from others, as a kind of secret language, as well as indicating in-group

solidarity and perspective, and operate at the level of the individual utterance (Matras, 2009, p. 294).

Ma'á

Ma'á, or Inner Mbugu, is spoken in Tanzania. It consists of Mbugu (Bantu) grammar with lexicon derived from Cushitic languages, some Bantu languages, and also from Mbugu, by changing the phonology and tone patterns (Mous, 2003a, pp. 214–215). Historically, the cattle-herding Mbugu endured conflict with other groups but consistently resisted subjugation and assimilation (Thomason, 1997c; Mous, 2003b). Today, speakers of Ma'á are multilingual, speaking Shambaa, Pare and Swahili (Mous, 2003b, p. 7). In example (6) a sentence in Ma'á is given in Line a, and in Mbugu in Line b.

- | | | | | | | |
|----|----|-----------------------------|---------|------|-------|---------|
| 6. | a. | áa-té | mi-hatú | kwa | choká | (Ma'á) |
| | | 3SG:PST-cut | 4-trees | with | axe | |
| | b. | áa-tema | mi-tí | kwa | izoka | (Mbugu) |
| | | 3SG:PST-cut | 4-trees | with | axe | |
| | | 'He cut trees with an axe.' | | | | |

(Adapted from Mous, 2003a, p. 212)

Mous (2003a) calls the Ma'á lexicon a 'parallel' lexicon, because the semantics and morphological properties of each word are the same as those for the equivalent items in Mbugu. He names the process lexical manipulation, essentially the same process as Matras' lexical re-orientation, and sees it as a conscious process, and the core mechanism for Mixed Languages of the Lexicon-Grammar type. In Mous' view conscious lexical manipulation is also the process of formation of other types of registers and varieties, including, e.g., secret, respect and initiation registers, with some similarity to Para-Romani varieties (Mous, 2003a, p. 222). Thomason (2001, p. 205, 1995) argues for a slow process of language formation over several centuries, in which the grammar of Cushitic languages was replaced by Bantu grammar and some lexicon through heavy borrowing, by bilingual speakers wanting to preserve their separate identity from other, powerful, groups, but under significant pressure to acculturate and shift. Mous (2003a, pp. 74–75) argues for a combination of processes. First, a slow process of gradual Bantuisation took place that eventually led to a shift to Mbugu from Cushitic. During the language shift, speakers maintained their lexicon and expanded it with words from Maasai (Nilotic) and Gorwaa (South Cushitic) (Mous, 2003a, p. 213). Researchers agree that the function of Ma'á is to indicate an identity that differs from that of the dominant cultural groups (earlier Pare, and now Shambaa) (Mous, 2003a; Thomason, 1997c).

Petjo and Javindo

Petjo (Van Rheedeen, 1994) and Javindo (De Gruiter, 1994) are grouped together here because they were formed in the former Dutch East Indies (now Indonesia), with similar language combinations. Both languages were formed in the context of Malay or Javanese-speaking women (respectively) marrying Dutch men, and their children interacting with each other in casual environments. Dutch was taught in some schools but not all children attended school and Dutch input for the Javindo-speaking children was limited (Van Rheedeen, 1994, p. 234). The (mostly) male children played together and inserted Dutch lexical items into a Malay

or Javanese morphosyntactic frame. In the case of Javindo, Javanese-speaking women married to Dutch men are reported to have first produced the mixed language, and their children to have learned it from them (De Gruiter, 1994, p. 152). In school playgrounds and in other play contexts, peer-group pressure meant that all of the children interacting together spoke Javindo (De Gruiter, 1994, p. 153). It seems possible, then, that the language emerged through a combination of parent-child and child-child interaction. Van Rheedeen (1994, p. 234) reports that Petjo was spoken into adulthood, but once Indonesia gained political independence Petjo did not remain. In example (7) Petjo is given in line a, Malay in line b and Dutch in line c.

7. a. ik haat noh lopen (Petjo)
 b. saja lagi djalan (Malay)
 c. ik loop/ik ben aan het lopen (Dutch)
 'I am walking.'

(Adapted from van Rheedeen, 1994, p. 231)

Although the two languages are no longer spoken, they are included here because of the question of the role of child speakers in their formation and continuation.

4.1.2 *Noun-Verb split languages*

While Bakker (1994, 1997) categorizes Michif as a Lexicon-Grammar type, Matras (2003b) and Muysken (2007) observe that the distinction in Michif is also between the language of the noun phrase versus the verb phrase, which can be called a Noun-Verb (N-V) split. In this split one language supplies the verb morphology, and the other the nominal morphology, creating a structural split between the verbal and nominal subsystems. Michif can be categorized this way because the Cree grammar in Michif is mostly verbal morphology, which is largely intact from Cree. The French nominals in Michif to a large extent have French nominal morphology. For some time Michif was believed to be the only language in the world with this type of Noun-Verb source mixture (Bakker and Papen, 1997). However Gillon and Rosen (2018) question both classifications of Michif, and argue that it is better characterized as Cree with French insertions.

Light Warlpiri

Recently two languages in Australia have been shown to have a similar noun-verb split to that of Michif, showing that the type can occur in other environments. Light Warlpiri, spoken by young Warlpiri in a remote Australian community, combines Warlpiri nominal morphology with Kriol verbal structure, and has significant innovations in the verbal system that draw on Warlpiri, Kriol and English (O'Shannessy, 2005, 2012, 2013). Kriol is an English-lexified Creole spoken across much of north Australia by Indigenous speakers (Schultze-Berndt, Meakins and Angelo, 2013; Sandefur, 1979, 1991). Warlpiri provides over 90% of the nominal morphology in Light Warlpiri, and Kriol and English are drawn on in an innovative structure in the verbal complex. However, since Kriol and English are both English-lexified and isolating in structure, they can in a sense be grouped together in supplying verbal forms and verbal morphology, making the split one between English-lexified languages vs. Warlpiri, which is agglutinating. In example (8) elements from Warlpiri are in italics and

those from Kriol and English are in plain font. The underlined -m is an innovative element explained later.

8. a. wi-m get-im *nyampu* *jarntu* *waja*
 1PL.NFUT get-TRANS DET dog EMPH
- b. *kuuku-ng* na i-m stii-im from us *junga* *mayi*
 monster-ERG FOC 1PL-NFUT steal-TR from us true Q
 ‘We got our dog. The monster stole it from us, didn’t it?’
 (ERGstoryLA92_2015, The Language Archive)

While many Mixed Languages have some degree of re-structuring of source language elements, Light Warlpiri is unique in having dramatic re-structuring of the verbal auxiliary system. In Light Warlpiri, transitive verbs are mostly derived from Kriol and English, and they and Warlpiri-derived transitive verbs take the Kriol transitive marker *-im* ‘TRANSITIVE.’ However, an innovative auxiliary structure *-m* ‘NONFUTURE’ combines pronoun forms from English and Kriol, as in *wi* ‘1PL’ in (8), with *-m*, a form common to Kriol pronouns (*im* ‘3SG,’ *dem* ‘3PL’) and English *I’m* ‘1SG-1SG.PRES.’ Where Kriol indicates past tense through the pre-verbal word *bin* ‘PAST,’ Light Warlpiri takes the ‘PAST’ semantic element of *bin* and combines it with the temporal semantics of English *I’m* ‘PRES,’ to create a separate morpheme meaning present and past, or nonfuture. The new auxiliary structure and semantics draw on Warlpiri, English and Kriol, but differ from all of them. Also noteworthy is that Kriol verb forms select ergative case marking from Warlpiri, as in (8), indicating that they combine features from more than one source (O’Shannessy, 2013).

Light Warlpiri has been shown to have emerged in a two-stage process. First, adult speakers code-switched between the source languages in a specific pattern when talking to very young children as part of a baby talk register (O’Shannessy, 2012). The patterning is so similar to that in the newly emerged system that it is clearly a forerunner to the new system. The children then internalized the code-switching as a single linguistic system, but also understood that one of the sources, Warlpiri, was itself a separate linguistic system, i.e., their two languages became Light Warlpiri and Warlpiri. The children continued to speak the new system, Light Warlpiri, as they grew up, and at the same time added the radical structural innovations in the verbal auxiliary system (O’Shannessy, 2012, 2013).

Gurindji Kriol

Gurindji Kriol, spoken in a remote community neighbouring that of Light Warlpiri speakers, combines nominal morphology from Gurindji and verbal grammar from Kriol (McConvell and Meakins, 2005). In example (9) elements from Gurindji are in italics, and from Kriol in plain font.

9. det *warlaku-ngku* i bin bait-im det *marluka* leg-*ta-rni*
 the dog-ERG 3SG.S PST bite-TR the old.man leg-LOC-ONLY
 ‘The dog bit the old man bang on the leg.’
 (O’Shannessy and Meakins, 2012, p. 8)

Gurindji Kriol transitive verbs also select Gurindji ergative case-marking, as in (9), but the ergative case-marking system has undergone considerable re-structuring (Meakins, 2015).

In addition to main verbs being from Kriol, Gurindji Kriol also has uninflected verbs from Gurindji. It has also been shown to be the result of conventionalization of code-switching processes by adults, internalized by children (McConvell and Meakins, 2005).

Light Warlpiri and Gurindji Kriol arose in situations of partial language shift and also reflect partial language maintenance in the face of intense pressure to shift. The ancestral languages of the speakers are Warlpiri and Gurindji, respectively. The wider society demands English for education and most administrative functions, and Kriol is spoken as a first or second language by many of the Indigenous people in nearby areas, with whom the Warlpiri and Gurindji are in contact. The new languages are a way of retaining the ancestral languages to a significant extent, and the identity that belongs with them (Meakins, 2008; O'Shannessy, 2013, 2015b).

Mednyj Aleut

On the other side of the world, a different type of split is seen in Mednyj Aleut, which was spoken by people of Aleut-Russian heritage on one of the Commander Islands in the Bering Strait. Mednyj Aleut might not be spoken any more as in the 1990s there were few speakers remaining, but it is well-known for its unusual type of language combination. It retains mostly Aleut morphological structure, but the entire finite verb inflection is Russian (Thomason, 1997d, p. 450; Golovko, 1994; Vakhtin, 1998). In example (10), elements from Russian are in italics, and from Aleut in plain font.

10. *on ivo* hunguu-*l*, husa-qayaam aalu-qalii-*l*
 he him push-PST roll-after.3SG smile-begin-3.SG
 'He pushed him, turned and laughed.'

(Golovko, 1994, p. 115)

Mednyj Aleut speakers positioned themselves as socially distinct from both Aleut and Russian speakers, and several researchers agree that they consciously developed their way of speaking to signal this, through lexical re-orientation (Thomason, 1997d, p. 463; Golovko, 1994, p. 113).

4.1.3 Converted languages

The term converted language was coined by Bakker (2003) who also uses the term Form-Semantics (F-S) mixed language, meaning one that combines the semantics of one language with the form and morpheme position (pre- or post-position) of another.

Sri Lanka Malay and Sri Lanka Portuguese

Sri Lanka Malay combines word shapes and lexicon from Javanese and other languages of the Indonesian archipelago (Ansaldò, 2009, p. 122) with grammar from Sinhala and Tamil (Bakker, 2003; Ansaldò, 2009). The term 'Malay' has been used since British rule to refer to peoples from the Indonesian archipelago, regardless of their cultural and linguistic diversity (Ansaldò, 2009, p. 122). 'Malay' varieties are not agglutinating, but both Sinhala and Tamil are. The nominal marking in Sri Lanka Malay combines 'Malay'-derived word shapes with Sinhala and Tamil agglutinative morphological structure. Sinhala and Tamil have undergone convergence through intensive and extended contact (Ansaldò, 2009; De Silva, 1974) so that it is difficult to tease apart specific features that derive from each language. In this sense, in

terms of influence on Sri Lanka Malay, Sinhala and Tamil can be thought of as one type, and ‘Malay’ varieties as another, making a two-way split.

11. Ni anak-nang baik buku-yang attu-aada
 This student-DAT good book-ACC one-have
 ‘This student has a good book.’

(Ansaldo, 2009, p. 129)

There is debate as to the social and linguistic history of Sri Lanka Malay, including, for example, whether its history includes an earlier creolized variety, and the relative influence of Tamil and Sinhala. Ansaldo (2009) argues for the gradual emergence of Sri Lanka Malay, while Bakker (2000) argues for rapid emergence. For details see, for example, Ansaldo (2009), Bakker (2000), Smith, Paauw and Hussainmiya (2004) and Smith and Paauw (2006).

Sri Lanka Portuguese is characterized similarly as having Portuguese lexicon, but having become an agglutinative language under the influence of Tamil (Bakker, 2003, p. 117; Smith, 1979).

4.2 Definitions and mechanisms of emergence

This section elaborates on the critical issues presented in Section 3, outlining major theories that attempt to capture and unify the language combinations present in Mixed Languages and the mechanisms leading to them.

Matras (2000, 2003b, 2009) defines a structural prototype of Mixed Languages, where one language provides the finite verb inflection and the ‘anchoring of predications,’ called the INFL-language, and the other supplies ‘most unbound, potentially autonomous content words, especially nouns’ (Matras, 2003b, p. 155). The INFL-language determines word order in the verb phrase and provides the structures for combining clauses. The split is argued to hold across all types of Mixed Languages, regardless of whether they are of the Lexicon-Grammar split type. The INFL-language could be the ancestral language, as in *Media Lengua*, where Quechua is the INFL-language, or could be the non-ancestral language, in situations of language shift, as in Ma’a, Mednyj Aleut, Para-Romani varieties, and the Australian mixed languages (Matras, 2003b, 2009). Nominal grammatical morphology need not pattern with the INFL-language. If the speakers retain the ancestral language as the INFL-language, they insert lexical items through a process of ‘lexical re-orientation’ (Matras, 2003b, p. 155). If the speakers have shifted to the incoming, socially dominant language as the INFL-language (the Functional Turnover Hypothesis (Matras, 2000, p. 87)), they retain some of the ancestral language through ‘selective replication’ (Matras, 2003b, p. 155), deliberately retaining some lexicon as part of an exclusive way of speaking. In this theory the components of the two source languages do not have equal grammatical status – ‘the INFL-language is the base onto which lexifier language material is incorporated’ (Matras, 2003b, p. 165). In Matras’ view the grammatical make-up of Mixed Languages is distinct from that of languages that have undergone heavy borrowing (Bakker and Matras, 2013, p. 2). Matras views the dichotomy between the material anchoring the predication – finite verbal morphology – and lexical material as a psycholinguistically natural one, also seen in child bilingual acquisition (Matras, 2009, p. 288). The predication material ‘reflects the overall social and communicative orientation of the speaker’ (Matras, 2009, p. 305), while the referential material is more open to conscious choice. Matras views Mixed Languages as a product of conscious decision.

Myers-Scotton takes the continuum view that the linguistic processes leading to Mixed Languages are not unique. She employs the term 'split language' and defines one as a way of speaking where there is 'systematic evidence that some of the abstract grammatical structure underlying its morphosyntactic frame comes from a source other than the major source of its lexicon' (Myers-Scotton, 2003, p. 73). She further qualifies the definition, adding that 'a language's morphosyntactic frame must have some abstract grammatical structure underlying the morphosyntactic frame from both participating languages (i.e., the type of structure relevant to late system morphemes' (Myers-Scotton, 2003, p. 91)).

Her Matrix Language Frame (MLF) and 4-M models (referring to four types of morpheme) identify a specific type of morphosyntactic structure that is sourced from a language that does not provide a significant amount of lexicon (Myers-Scotton, 2003, 2004). The models propose that split languages are preceded by the bilingual practice of code-switching, and that the structure of language combination in bilingual speech shows a systematic asymmetry in how elements from the two languages are combined. Both languages provide surface morphemes, but only one language provides the underlying morphosyntactic frame of a clause, and this language is the matrix language (Myers-Scotton, 2004, p. 106). The other language is the embedded language and elements from it are inserted into the matrix language or can occur as Embedded Islands; these are phrases with full morphology but embedded into a larger matrix language frame. The matrix language determines the order of morphemes (regardless of which language they are from), and also supplies specific types of grammatical morpheme, called late outsider system morphemes. In this theory morphemes are categorized as content or system morphemes; system morphemes are further grouped into early or late system morphemes, and late system morphemes are of two types, bridge or late outsider morphemes. Late outsider morphemes co-index larger constituents, and include, for example, subject agreement markers on verbs, or nominal case-markers (Myers-Scotton, 2004, p. 111).

Myers-Scotton (2003, p. 88) argues that in bilingual contexts there can be a matrix language turnover, in which the former embedded language gradually becomes the matrix language, as 'the morphosyntactic frame changes from that derived from one language to that derived from the other' (Myers-Scotton, 2003, p. 88). But a turnover can be arrested midway, in which case the matrix language only partly changes, so the way of speaking then has abstract grammatical material from both languages, or a composite matrix language frame. In Myers-Scotton's theory this is the mechanism of emergence of a split language. A context in which a matrix language turnover might not be completed would be when the socially dominant language in a community shifts from being one language to the other. If a matrix language turnover has been arrested, the result is structure in at least one component of the morphosyntactic frame sourced from both contributing languages, which is Myers-Scotton's definition of a split language. By abstract material she means lexical-conceptual structure, predicate-argument structure, and morphological realization (Myers-Scotton, 2003, p. 85). Different combinations of abstract material at the three levels result in different surface forms.

Auer (1999, 2014) also takes the gradient view of Mixed Language emergence, arguing that Mixed Languages are not unique in terms of the grammatical processes leading to them, but that their social histories lead to extreme outcomes from processes that occur in other language contact situations. Auer views grammaticization of insertional bilingual mixing patterns and structural conventionalization as central processes in what he terms 'fused lects.' Auer (1999) proposes that code-switching practices, that first perform a locally meaningful discourse contextualization function, where specific choices of lexicon signal something about the context or the speaker, can conventionalize to the extent that the practice loses the discourse function and becomes the expected way to structure utterances, with little variation. At this point the

language combining takes on a new function, as a whole linguistic system, of expressing a specific ethnic or group identity, under pressure from a majority group to shift to their language (Auer, 2014, p. 297).

Also taking the gradient view, Thomason (1995, 2003) argues that Mixed Languages are the ‘extraordinary results’ of ‘ordinary processes’ of language contact. In terms of ‘strictly linguistic processes,’ there is ‘no sharp divergence between ordinary contact-induced changes and the combinations of features found in bilingual mixed languages: the difference is in degree, not in kind’ (Thomason, 2003, p. 26). She categorizes Mixed Languages (which she calls bilingual mixed languages, as opposed to other languages that are mixed – i.e., pidgins and creoles) into two groups – the language of a persistent ethnic group that develops slowly over generations, in contrast to the language of a new ethnic group that emerges relatively rapidly, sometimes in a single generation. Thomason positions Ma’a in the slowly emerging category, and Michif, Mednyj Aleut and Media Lengua in the abruptly emerging category (Thomason, 1995, pp. 18–20). However, the recently documented Mixed Languages in Australia, Light Warlpiri (O’Shannessy, 2005, 2013) and Gurindji Kriol (McConvell and Meakins, 2005), challenge this dichotomy, as they emerged relatively abruptly, in two generations, but are not spoken by new ethnic groups. Rather, part of the motivation for these Mixed Languages is the retention of heritage identity through language expression.

Muysken (2007, pp. 333–334) suggests a competition model in which linguistic processing principles operating in situations of language contact apply differentially in different contexts, leading to the variety of types of Mixed Languages. The principles include, for example, ‘verbal elements are retained more frequently than nominal elements,’ and ‘functional elements are frequently taken from the same language as lexical elements in their immediate environment.’ Muysken (2007) explains how combinations of the principles play out in each Mixed Language.

With Lexicon-Grammar Mixed Languages in mind, Mous (2003a, p. 230) views deliberate lexical manipulation as the core of this type of Mixed Language. However, in his view, ‘lexical manipulation occurs not only in Mixed Languages but in a variety of situations in which speakers attempt to exercise conscious control over their language.’ Mous (2003a) outlines several lexical manipulation strategies and the motivations and criteria, social and linguistic, for each.

Autonomy

Most Mixed Languages, perhaps all except Michif, are spoken alongside at least one of the source languages, and are known as symbiotic Mixed Languages (Smith, 1995). Nevertheless they are independent, autonomous systems, used by their speakers for different functions (Bakker, 2003, p. 125). A major indicator of autonomy is when the new Mixed Language is learned by children as (one of) their first language(s), and they might not receive input in both of the source languages. For example Light Warlpiri-speaking children hear both Light Warlpiri and Warlpiri from birth, but only learn English and some Kriol as they grow up (O’Shannessy, 2012). Their Light Warlpiri, then, is clearly not code-switching, as they do not know all of the source languages well enough to switch between them. Another indicator of autonomy is that a change that takes place in one of the sources need not take place in the Mixed Language, and vice versa. Independence in the Mixed Language is seen in, for example, further change and regularisation of morphology over time (e.g., Light Warlpiri, Gurindji Kriol). Light Warlpiri speakers choose different words from pairs of homonyms, show more regularization of nominal morphology, and choose more consonant-final variants of some

suffixes, when they are speaking the mixed language Light Warlpiri, as opposed to the heritage language, Warlpiri (O'Shannessy, 2015a). The independent changes make the distinction between the languages even greater, as the type and amount of shared portions diverge. Other examples of a change in a source language or a Mixed Language, but not both, are given in Bakker (2003, pp. 126–128). However, a minority language may be influenced by other languages with which it is in contact, regardless of the mechanisms of its origin, and even though it is an autonomous linguistic system (Bakker, 2003, p. 136). For example, Light Warlpiri speakers speak Warlpiri so are still in contact with Warlpiri speakers, and this has been argued to assist in the amount of lexicon and nominal morphology in Light Warlpiri that is sourced from Warlpiri (O'Shannessy, 2015a; O'Shannessy and Meakins, 2012).

4.3 The outcome of a mixed language compared to bilingual speech practices

Since Mixed Languages most often arise in situations of bilingualism and multilingualism, it is interesting to ask how bilingual practices contribute to the resulting structures of Mixed Languages. Are the speech practices of those who first bring the languages together in a specific structure similar to speech practices of bilinguals in other situations? As mentioned previously, several researchers see Mixed Languages as an extreme result differing only in degree from the practices of bilinguals in other contexts, while others see the process as specific to Mixed Languages.

4.3.1 The role of code-switching

Before the two Australian Mixed Languages had been identified, bringing empirical evidence to show that Mixed Languages can arise from the conventionalization of community code-switching, researchers had been divided as to whether community code-switching could lead to the outcome of a Mixed Language. Some researchers observed that (writing in 2002–2003) there had been no direct observation of the emergence of a Mixed Language (Thomason, 2003; Matras and Bakker, 2003b). A call was made for the role of conventionalized code-switching practices to be evidenced by longitudinal data showing that there was code-switching in a community before a Mixed Language conventionalized, and that the structural patterns of language combining in the code-switching were the same as those seen in the Mixed Language (Matras and Bakker, 2003b), and that there was evidence of a transitory phase between code-switching and Mixed Language emergence (Bakker, 2003). Researchers drew on Muysken's (2000) dichotomy of types of code-switching, distinguishing insertional code-switching, where items from one language are inserted into the frame of another, including within a phrase, from alternational code-switching, where both the lexicon and grammar change from one language to the other, often at clausal or phrasal boundaries.

Backus (2003) and Auer (2014) explain how alternational code-switching patterns could not lead to the outcome of a Mixed Language because the structures of mixing differ, and alternational patterns of mixing are highly variable. Bakker (2003) argues that insertional code-switching cannot lead to a Mixed Language either, because in code-switching only a small quantity of items are inserted, and the code-switched lexical items are typically not from everyday vocabulary. In contrast, in Mixed Languages, there is wholesale replacement or retention of lexicon, including everyday basic vocabulary. Bakker (2003) also argues that combining elements from two languages within a word, e.g., a stem and suffix from different languages, is not common. Although he describes several configurations of Mixed Languages,

his arguments against code-switching pertain only to the intertwining category. Thomason (2003, pp. 28–29) sees a partial role for code-switching in Michif and Media Lengua, along with deliberate decision, but not in Mednyj Aleut, because the type of language combination does not reflect patterns of code-switching seen in other contexts (Thomason, 2003, p. 30).

In contrast, Myers-Scotton's models propose that the structural split of Mixed Languages largely echoes that of code-switching, and that the Mixed Language is brought about by a halting of the turnover of the matrix language. Insertions of elements of one language into the morphosyntactic frame of the other is therefore an essential component. Auer (1999, 2014) argues that two types of insertional language mixing lead to fusion, or Mixed Languages. Mous (2003a, p. 217) suggests that code-switching may play a role in Mixed Languages. Muysken (1997) sees a role for code-switching as a mechanism as well as relexification in Media Lengua, and suggests that the Lexicon-Grammar split could derive from 'a grammaticalised and regularised result of code-switching practices' (Muysken, 2007, p. 331). Thomason (2003, p. 27) claims that code-switching may have been an important factor in the formation of Michif. Bakker (1997) and Bakker and Papen (1997) report that code-switching was present in the multilingual Métis community at the time of the emergence of Michif, but do not see it as a mechanism leading to Michif.

Recently the documentation of Light Warlpiri and Gurindji Kriol have provided the data for the transitory phase between code-switching practices and the emergence of a Mixed Language, answering Matras and Bakker's (2003b) call for evidence. McConvell (1988) had published code-switching data from speakers switching between Gurindji and Kriol in the 1970s. Data in McConvell and Meakins (2005) show that the mixing patterns in the 1970s data are the same as those in the conventionalized way of speaking now known as Gurindji Kriol. Similarly, O'Shannessy (2012) shows code-switching practices of adults older than the Light Warlpiri-speaking cohort directed to young children, and shows that these are the same pattern of language combination as in Light Warlpiri (with the absence of innovations that occurred later). The data from both mixed languages confirm that code-switching patterns of language combination can conventionalize to become the regular, everyday way of speaking in a community. These cases show the language mixing processes before and after they were systematized to the extent of being Mixed Languages, involving subsystems of the source languages appearing largely intact in the new way of speaking, and the mixed clauses being the default manner of expression, reducing variation in how concepts are expressed.

Other evidence of the transitory phase is seen in cases where speakers' language mixing has the properties of Mixed Languages to some extent, but does not affect whole subsystems, and has not continued to conventionalize as much as in Mixed Languages. Thrace Romani and Finnish Romani combine Turkish or Finnish verbal morphology, respectively, with otherwise Romani strings, but Romani remains the grammatically dominant language, and can be seen to be the genetic parent (Adamou and Granqvist, 2015, p. 526). The authors see both varieties of Romani as 'early stages of the formation of Mixed Languages,' resulting from an arrested matrix language turnover (cf. Myers-Scotton, 2003). Winford (2013, p. 369) suggests that urban youth varieties in Africa (cf. Kießling and Mous, 2004) could also be viewed as examples of the transitory phase, as speakers use both code-switching and lexical manipulation to create new varieties for the purpose of signalling a distinct identity, but they are not systematized to the extent of Mixed Languages.

It is not the case, of course, that code-switching in bilingual communities always leads to a Mixed Language, as there are many communities where code-switching is a stable, everyday practice, and is structurally patterned (e.g., Poplack, 1980; Myers-Scotton, 2004), but not in the systematic, relatively invariable way of Mixed Languages, and does not conventionalize

to become a Mixed Language. Similarly, language mixing or fusion may lead to a fused lect, but not necessarily to the extreme result of a Mixed Language (Adamou and Granqvist, 2015, p. 526).

4.3.2 How conscious or deliberate are the processes?

The role of code-switching practices preceding the emergence of a Mixed Language is related to the question of deliberate decision as a mechanism in Mixed Language emergence. There is likely to be an inverse relationship between the degree to which code-switching practices play a role in Mixed Language formation and the degree of deliberate decision involved. Code-switching practices in bilingual communities are often not conscious choices (Poplack, 1980; Zentella, 1997), and their conventionalization probably is not either. But where lexical manipulation (Mous, 2003a) or re-orientation (Matras, 2003b, 2009) is a major factor, then conscious decision may come to the fore.

Several researchers (Bakker, 1997; Bakker and Matras, 2013; Thomason, 2003; Muysken, 1997; Golovko, 2003) see deliberate decision as a major factor in the motivation for some Mixed Languages. Referring to mixed languages with a Lexicon-Grammar split, Mous (2003a, p. 217) states that 'lexical manipulation is the key process in the emergence of mixed languages' and that this is a conscious act by speakers. He allows, though, that code-switching 'may very well lead to the emergence of a mixed language, and in particular such a development is conceivable through the conventionalisation of the switches' (Mous, 2003a, p. 217). For Mednyj Aleut, Golovko (2003) argues that speakers consciously manipulated their way of speaking.

For some languages, for instance current Para-Romani varieties, speakers have consciously tried to bring back parts of a language that was spoken very little at the time (Matras, 2010). Thomason (2001, pp. 149–152, 204) also sees 'change by deliberate decisions' as a major factor in several Mixed Languages but argues that both code-switching and deliberate decision played a role in Michif, and code-switching may have played a role in Media Lengua (Thomason, 2003, p. 28). Similarly, Muysken (1981) views relexification in Media Lengua as deliberate manipulation by its speakers in a situation of partial acculturation to a dominant language and culture.

4.3.3 The roles of adults vs. children

Linked to the notion of deliberate language change is the idea that the changes are brought about by bilingual adult speakers, or bilingual speakers old enough to have already learned their two languages very well, since they retain the morphosyntax of one of the languages largely intact (Thomason, 2003, p. 32).

But a role for young children in the process of first language acquisition as well as of adults has also been seen. O'Shannessy (2012, 2013) has shown that young children played a role of pattern conventionalization and innovation in the second stage of the process, when receiving code-switched input from adults. The type of language processing they undertook was the same as that of children everywhere (O'Shannessy, Forthcoming), but the differing sociolinguistic environment meant that the innovations remained in the children's speech and became part of their linguistic system, rather than disappearing as children's innovations often do as they grow older. This evidence supports hypotheses made by Aboh (2009, 2015), Aboh and Ansaldo (2006) and Mufwene (2001), who argue that altered replication of the input to children is likely to occur in contexts of multilingual first language acquisition when there

is considerable variation in the input. It also supports the hypothesis of McConvell (1991, 2008), that children may model their speech styles on each other's speech, rather than on that of adults. For Gurindji Kriol, McConvell and Meakins (2005) argue that adult speakers developed the code-switching patterns, and the next generation of children conventionalised them further. It is shown that once the system was conventionalized, children and teenagers added innovations (Meakins, 2011; van den Bos, Meakins and Algy, 2017).

In cases where there is a two-stage process in the emergence of the new language, and the second, consolidation, stage is undertaken by children, it is not likely that the young children deliberately decide to speak in a new way. It is more likely that the children receive very consistently patterned input, and that factors such as role alignment with other children (McConvell, 2008; Agha, 2007), speech accommodation (Giles, Coupland and Coupland, 1991) and priming (Pickering and Garrod, 2004) all play a role in the continuation of the new speech style.

Data from bilingual children's speech that does not conventionalize into a new system nevertheless supports the hypothesis that in some specific sociolinguistic circumstances children could play a role in the development of a Mixed Language, and also support Matras' and others' ideas of the psycholinguistic naturalness of a Lexicon-Grammar split, or an INFL-other-language split. A two-year-old child learning Garo (spoken in Assam, India) and English at one stage produced English words with Garo morphology (Burling, 1959). The patterning was very like that of a Mixed Language, with a Lexicon-Grammar split. Similarly, bilingual children in the early stages of language development often produce lexical stems from one language with morphology from the other (Lanza, 1992, 1997; Comeau, Genesee and Lapaquette, 2003; Genesee, Nicoladis and Paradis, 1995). Usually the patterning in these cases is not like that of a Mixed Language, because the lexical stems are drawn from one language in some contexts and from the other in other contexts, and likewise the morphology, yet the Lexicon-Grammar combinations hint at the possibility of the children's combinations being like those of Mixed Languages in specific situations. However, the evidence suggests that young bilingual children would not develop a Mixed Language without receiving Mixed Language-like input from those speaking to them. A two-stage process such as that described for Light Warlpiri and Gurindji Kriol would be required, where the children received systematic, patterned code-switching input, and then further conventionalized that input.

5. Main research methods

To discuss the research methods for identifying a Mixed Language I draw on Thomason's (2001, pp. 93–94) criteria for recognizing contact influence in a language, and adapt them to focus on Mixed Languages. First, examine all of the subsystems in the language. Second, identify the possible source(s) by understanding the sociolinguistic situation and its dynamics over time, making sure that the speakers were bilingual in the relevant languages, and identifying any social pressures on them. Does the socio-historical situation have elements in common with those of documented Mixed Languages? Third, find the sub-systems that are shared between the sources and putative Mixed Language, looking for systematicity of proposed source elements, and whether the subsystem is present in its entirety (or almost). Fourth, ensure that the proposed shared features were not present in the putative Mixed Language before the time of language contact; and finally, be sure that they were present in the sources. As a further guide, identify whether elements in the language pattern according to a Mixed Language type already documented. Examine whether features of the sources have changed while being retained in the Mixed Language, and in what way. These may be qualitative

changes, such as re-lexification (cf. Muysken, 1994), or for instance, changes in the degree of use of an element (cf. Meakins and O'Shannessy, 2010).

Speech production data are essential, because key to categorization is how systematically each source language is present in the new way of speaking. Spontaneous and naturalistic data are the most useful, where people speak in their most vernacular style (cf. Labov, 1963). This is especially important where speakers also speak one of the sources and where their Mixed Language may be stigmatized. Sometimes part of a sub-system might not occur frequently in everyday speech and creative methods are required to record naturalistic data. Examples include creating picture books as stimuli for speakers to use to tell their own stories, where the pictures prompt the more frequent production of some structures (cf. O'Shannessy, 2009). Similar results can be obtained through other creative games and activities.

Another important perspective is that of the social functions of the language – who uses it, and for what purposes? Is it the everyday language of the speakers? Are children learning the language as (one of) their first language(s)? Do they also speak one or more of the sources? The methods require detailed investigation into the socio-historical situation of the speakers, regardless of whether the time of change was long ago or more recent.

Although production data are the basis for categorization, experimental methods can shed light on complementary questions. Acceptability judgments of derivational nominal affixes by speakers of Philippine Hybrid Hokkien (combining Hokkien, Tagalog and English), showed preferences for one- or two-syllable Tagalog prefixes occurring on Hokkien stems (Gonzales, 2018). These data suggest that Philippine Hybrid Hokkien may fit the category of a Mixed Language, with Tagalog affixes occurring systematically on Hokkien stems, but production data are needed to confirm this.

To examine the motivations for preferring some kinds of language combination over others in intra-sentential code-switching data, bilingual speakers of Quichua and Media Lengua in Ecuador gave grammaticality judgments about Quichua, Media Lengua and mixed Quichua-Media Lengua utterances (Lipski, 2019). They preferred unmixed utterances, reflecting that the two languages are not often mixed in everyday interactions. In a memory-loaded repetition task bilingual speakers of the same two languages showed that language switches between subject pronouns and verbs, and after interrogatives, are accepted, contrary to much of the earlier literature on the question (Lipski, 2019).

A sentence processing experiment was used to test whether the cognitive costs of code-switching vary according to the degree of code-switching frequency. Romani-Turkish data in Greek Thrace shows the use of Turkish verbs and verbal morphology in otherwise Romani utterances (Adamou and Granqvist, 2015; Adamou, 2010). Speakers heard sentences with verb and nonverb combinations that either reflected everyday use in the community, or did not. Reaction time results showed that mixed Romani-Turkish utterances were processed as quickly as unilingual utterances when the Turkish verbs present were those that are frequently used in everyday Romani interactions in the community. Costs of switching to a Turkish verb were observed when that verb is not the most commonly used (Adamou and Shen, 2019). The study shows that this type of experiment needs to take into account the actual languages configurations used by speakers in their everyday lives.

Sentence processing was also used in a study of young children's language development in two mixed languages, Light Warlpiri and Gurindji Kriol (O'Shannessy and Meakins, 2012). Both languages have variable word order and optional ergative case-marking on overt subjects of transitive verbs. Young adults and children aged five to nine years old heard transitive sentences with manipulations of whether ergative case-marking was present, and of word order;

the participants pointed to one of two possible animations depicted by each sentence. Results show that while speakers of both languages used both word order and ergative case-marking to identify agents of transitive events, Light Warlpiri speakers relied on ergative marking more often, and on word order less often, than Gurindji Kriol speakers did. The differences in processing are likely due to different sociolinguistic contexts – the Light Warlpiri speakers also learn traditional Warlpiri (in which ergative case-marking is close to obligatory) and hear Warlpiri spoken regularly, but the Gurindji Kriol speakers have much less access to traditional Gurindji.

6. Future directions

This chapter has outlined research and critical questions about definitions of, and mechanisms leading to, the emergence of Mixed Languages. Many questions remain, including whether there are significantly more Mixed Languages spoken than those currently documented, and whether the types of Mixed Language attested so far exhaust the types that are possible. As more candidate Mixed Languages are documented, yet more nuance about the mechanisms of emergence will likely be found. Where possible, documenting this type of language from close to the time of emergence allows us to better understand the sociolinguistic situations involved. Details of how the languages develop over time would add greatly to our understandings. Another fruitful perspective is whether young children's bilingual speech production resembles that of the combinations found in Mixed Languages. Continuing research will assist in illuminating the contexts and mechanisms of this fascinating group of languages.

7. Further reading

Bakker, P. (1997). *A language of our own: The genesis of Michif, the mixed Cree-French language of the Canadian Metis*. Oxford: Oxford University Press.

A detailed examination of Michif, including historical sociolinguistic factors and variation in Michif.

Bakker, P. and Mous, M. (1994). *Mixed languages and language intertwining*. Amsterdam: IFOTT.

The volume presents a brief overview of several case studies of Mixed Languages. It is an important volume in the history of the research.

Matras, Y. (2010). *Romani in Britain*. Edinburgh: Edinburgh University Press.

The book examines the history of, and social and psycholinguistic motivations for, the emergence of Para-Romani varieties.

Matras, Y. and Bakker, P., eds. (2003). *The mixed language debate*. Berlin and New York: De Gruyter Mouton.

The volume provides differing perspectives by leading researchers on the processes leading to Mixed Languages, including socio-historical and psycholinguistic perspectives. It is a foundational reading in this area.

Thomason, S.G. (1997). *Contact languages: A wider perspective*. Amsterdam: John Benjamins.

This volume is not restricted to Mixed Languages only, but presents discussions of five Mixed Languages.

8. Related topics

Pidgins, creoles, language creation, language shift, language evolution, language change

Abbreviations

| | |
|------|-----------------------|
| 1 | 1st person |
| 3 | 3rd person |
| 4 | noun class 4 |
| ACC | accusative |
| BEN | benefactive |
| COMP | complementizer |
| DAT | dative |
| DEF | definite |
| DET | determiner |
| EMPH | emphatic |
| ERG | ergative |
| F-S | Form-Semantics |
| FOC | focus |
| INAN | inanimate |
| L-G | Lexicon-Grammar |
| LOC | locative |
| MLF | Matrix Language Frame |
| N | noun |
| NFUT | nonfuture |
| NOM | nominative |
| PL | plural |
| PROG | progressive |
| PST | past |
| Q | question |
| SG | singular |
| TR | transitive |
| V | verb |

References

- Aboh, E.O. (2009). Competition and selection. That's all! In: E. Aboh and N. Smith, eds., *Complex processes in new languages*, 1st ed. Amsterdam: John Benjamins.
- Aboh, E.O. (2015). *The emergence of hybrid grammars: Language contact and change*. Cambridge: Cambridge University Press.
- Aboh, E.O. and Ansaldo, U. (2006). The role of typology in language creation. In: U. Ansaldo, S. Matthews and L. Lim, eds., *Deconstructing creole*, 1st ed. Amsterdam: John Benjamins.
- Adamou, E. (2010). Bilingual speech and language ecology in Greek Thrace: Romani and Pomak in contact with Turkish. *Language in Society*, 39, pp. 147–171.
- Adamou, E. and Granqvist, K. (2015). Unevenly mixed Romani languages. *International Journal of Bilingualism*, 19, pp. 525–547.
- Adamou, E. and Shen, X.R. (2019). There are no language switching costs when codeswitching is frequent. *International Journal of Bilingualism*, 23, pp. 53–70.
- Agha, A. (2007). *Language and social relations*. Cambridge: Cambridge University Press.
- Ansaldo, U. (2009). *Contact languages: Ecology and evolution in Asia*. Cambridge: Cambridge University Press.
- Auer, P. (1999). From code-switching via language mixing to fused lects: Toward a dynamic typology of bilingual speech. *The International Journal of Bilingualism*, 3, pp. 309–332.

- Auer, P. (2014). Language mixing and language fusion: When bilingual talk becomes monolingual. In: J. Besters-Dilger, C. Dermarkar, S. Pfänder and A. Rabus, eds., *Congruence in contact-induced language change: Language families, typological resemblance and perceived similarity*, 1st ed. Berlin: De Gruyter.
- Backus, A. (2003). Can a mixed language be conventionalized alternational code-switching? In: Y. Matras and P. Bakker, eds., *The mixed language debate*, 1st ed. Berlin and New York: De Gruyter Mouton.
- Bakker, P. (1994). Michif, the Cree-French mixed language of the metis buffalo hunters in Canada. In: P. Bakker and M. Mous, eds., *Mixed languages: 15 Case studies in language intertwining*, 1st ed. Amsterdam: IFOTT.
- Bakker, P. (1997). *A language of our own: The genesis of Michif, the mixed Cree-French language of the Canadian Metis*. Oxford: Oxford University Press.
- Bakker, P. (2000). Rapid language change: Creolization, intertwining, convergence. In: C. Renfrew, A. McMahon and L. Trask, eds., *Time depth in historical linguistics*, 1st ed. Cambridge: McDonald Institute for Archeological Research.
- Bakker, P. (2003). Mixed languages as autonomous systems. In: P. Bakker and Y. Matras, eds., *The mixed language debate*, 1st ed. Berlin and New York: De Gruyter Mouton.
- Bakker, P. and Matras, Y. (2013). Introduction. In: P. Bakker and Y. Matras, eds., *Contact languages: A comprehensive guide*, 1st ed. Berlin and New York: De Gruyter Mouton.
- Bakker, P. and Mous, M. (1994a). Introduction. In: P. Bakker and M. Mous, eds., *Mixed languages and language intertwining*, 1st ed. Amsterdam: IFOTT.
- Bakker, P. and Mous, M. (1994b). *Mixed languages and language intertwining*. Amsterdam: IFOTT.
- Bakker, P. and Papen, R.A. (1997). Michif: A mixed language based on Cree and French. In: S.G. Thomason, ed., *Contact languages: A wider perspective*, 1st ed. Amsterdam: John Benjamins.
- Bundgaard-Nielsen, R. and O'Shannessy, C. (2015). A marriage of necessity: Light Warlpiri obstruent perception. In: *Presented at the Australian linguistic society annual conference*. University of Western Sydney, Sydney.
- Bundgaard-Nielsen, R. and O'Shannessy, C. (2018). Obstruent distinctions in Light Warlpiri. *Presented at the Australian linguistics society annual conference*. University of South Australia, Adelaide.
- Burling, R. (1959). Language development of a Garo and English-speaking child. *Word*, 15, pp. 45–68.
- Comeau, L., Genesee, F. and Lapaquette, L. (2003). The modelling hypothesis and child bilingual code-mixing. *International Journal of Bilingualism*, 7, pp. 113–126.
- De Gruiter, M. (1994). Javindo, a contact language in pre-war Semarang. In: P. Bakker and M. Mous, eds., *Mixed languages: 15 Case studies in language intertwining*, 1st ed. Amsterdam: IFOTT.
- De Silva, M.W.S. (1974). Convergence in diglossia: The Sinhalese situation. *International Journal of Dravidian Linguistics*, 3, pp. 60–91.
- Genesee, F., Nicoladis, E. and Paradis, J. (1995). Language differentiation in early bilingual development. *Journal of Child Language*, 22, pp. 611–631.
- Giles, H., Coupland, N. and Coupland, J. (1991). Accommodation theory: Communication, context and consequence. In: H. Giles, J. Coupland and N. Coupland, eds., *Contexts of accommodation: Developments in applied sociolinguistics*, 1st ed. Cambridge: Cambridge University Press.
- Gillon, C. and Rosen, N. (2018). *Nominal contact in Michif*. Oxford: Oxford University Press.
- Golovko, E.V. (1994). Mednij Aleut or Copper Island Aleut: An Aleut-Russian mixed language. In: P. Bakker and M. Mous, eds., *Mixed languages: 15 case studies in language intertwining*, 1st ed. Amsterdam: IFOTT.
- Golovko, E.V. (2003). Language contact and group identity: The role of “folk” linguistic engineering. In: Y. Matras and P. Bakker, eds., *The mixed language debate*, 1st ed. Berlin and New York: De Gruyter Mouton.
- Gonzales, W.D.O.W. (2018). *Philippine hybrid Hokkien as a postcolonial mixed language: Evidence from nominal derivational affixation mixing*. MA thesis. National University of Singapore.
- Jones, C. and Meakins, F. (2013). Variation in voice onset time in stops in Gurindji Kriol: Picture naming and conversational speech. *Australian Journal of Linguistics*, 33, pp. 196–220.

- Jones, C., Meakins, F. and Buchan, H. (2011). Comparing vowels in Gurindji Kriol and Katherine English: Citation speech data. *Australian Journal of Linguistics*, 31, pp. 305–326.
- Kiessling, R. and Mous, M. (2004). Urban youth languages in Africa. *Anthropological Linguistics*, 46, pp. 303–341.
- Labov, W. (1963). The social motivation of a sound change. *Word*, 19, pp. 273–309.
- Lanza, E. (1992). Can bilingual two-year-olds code-switch? *Journal of Child Language*, 19, pp. 633–658.
- Lanza, E. (1997). *Language mixing in infant bilingualism: A sociolinguistic perspective*. Oxford: Oxford University Press.
- Lipski, J.M. (2019). Field-testing code-switching constraints: A report on a strategic languages project. *Languages*, 4, p. 7.
- Matras, Y. (2000). Mixed languages: A functional-communicative approach. *Bilingualism: Language and Cognition*, 3, pp. 79–99.
- Matras, Y. (2003a). Mixed languages: Re-examining the structural prototype. In: Y. Matras and P. Bakker, eds., *The mixed language debate*, 1st ed. Berlin and New York: De Gruyter Mouton.
- Matras, Y. (2003b). Mixed languages: Re-examining the structural prototype. In: Y. Matras and P. Bakker, eds., *The mixed language debate: Theoretical and empirical advances*, 1st ed. Berlin and New York: De Gruyter Mouton.
- Matras, Y. (2009). *Language contact*. Cambridge: Cambridge University Press.
- Matras, Y. (2010). *Romani in Britain*. Edinburgh: Edinburgh University Press.
- Matras, Y. and Bakker, P., eds. (2003a). *The mixed language debate: Theoretical and empirical advances*, 1st ed. Berlin and New York: De Gruyter Mouton.
- Matras, Y. and Bakker, P. (2003b). The study of mixed languages. In: Y. Matras and P. Bakker, eds., *The mixed language debate*, 1st ed. Berlin and New York: De Gruyter Mouton.
- McConvell, P. (1988). Mix-im-up: Aboriginal code-switching, old and new. In: M. Heller, ed., *Codeswitching: Anthropological and sociolinguistic perspectives*, 1st ed. Berlin: De Gruyter Mouton.
- McConvell, P. (1991). Understanding language shift: A step towards language maintenance. In: S. Romaine, ed., *Language in Australia*, 1st ed. Cambridge: Cambridge University Press.
- McConvell, P. (2008). Language mixing and language shift in Indigenous Australia. In: J. Simpson and G. Wigglesworth, eds., *Children's language and multilingualism: Indigenous language use at home and school*, 1st ed. London: Continuum Press.
- McConvell, P. and Meakins, F. (2005). Gurindji Kriol: A mixed language emerges from code-switching. *Australian Journal of Linguistics*, 25(1), pp. 9–30.
- Meakins, F. (2008). Land, language and identity: The socio-political origins of Gurindji Kriol. In: M. Meyerhoff and N. Nagy, eds., *Social lives in language*, 1st ed. Amsterdam: John Benjamins.
- Meakins, F. (2011). *Case-marking in contact: The development and function of case morphology in Gurindji Kriol*. Amsterdam: John Benjamins.
- Meakins, F. (2013). Mixed languages. In: P. Bakker and Y. Matras, eds., *Contact languages: A comprehensive guide*, 1st ed. Boston and Berlin: De Gruyter Mouton.
- Meakins, F. (2015). From absolutely optional to only nominally ergative: The life cycle of the Gurindji ergative suffix. In: F. Gardani, P. Arkadiev and N. Amirdzhe, eds., *Borrowed morphology*, 1st ed. Berlin and New York: De Gruyter.
- Meakins, F. (2016). Mixed languages. In: A. Pine, M. Turin and M. Aronoff, eds., *Oxford research encyclopedia of linguistics*, 1st ed. Oxford: Oxford University Press.
- Meakins, F. and O'Shannessy, C. (2010). Ordering arguments about: Word order and discourse motivations in the development and use of the ergative marker in two Australian mixed languages. *Lingua*, 20, pp. 1693–1713.
- Mous, M. (2003a). The linguistic properties of lexical manipulation and its relevance for Ma'a. In: Y. Matras and P. Bakker, eds., *The mixed language debate*, 1st ed. Berlin and New York: De Gruyter Mouton.
- Mous, M. (2003b). *The making of a mixed language: The case of Ma'a/Mbugu*. Amsterdam and Philadelphia: John Benjamins.
- Mufwene, S. (2001). *The ecology of language evolution*. Cambridge: Cambridge University Press.

- Müller, F.M. (1873). *Lectures on the science of language*. London: Longman, Green & Co.
- Muysken, P. (1981). Halfway between Quechua and Spanish: The case for relexification. In: A. Valdman and A. Highfield, eds., *Historicity and variation in creole studies*, 1st ed. Ann Arbor: Karoma.
- Muysken, P. (1994). Media Lengua. In: P. Bakker and M. Mous, eds., *Mixed languages: 15 case studies in language intertwining*, 1st ed. Amsterdam: IFOTT.
- Muysken, P. (1997). Media Lengua. In: S.G. Thomason, ed., *Contact languages: A wider perspective*, 1st ed. Amsterdam and Philadelphia: John Benjamins.
- Muysken, P. (2000). *Bilingual speech: A typology of code-mixing*. Cambridge: Cambridge University Press.
- Muysken, P. (2007). Mixed codes. In: P. Auer and L. Wei, eds., *Handbook of multilingualism and multilingual communication*. Berlin and New York: De Gruyter Mouton.
- Muysken, P. (2013). Language contact outcomes as the result of bilingual optimization strategies. *Bilingualism: Language and Cognition*, 16, pp. 709–730.
- Myers-Scotton, C. (2003). What lies beneath: Split (mixed) languages as contact phenomena. In: P. Bakker and Y. Matras, eds., *The mixed language debate*. Berlin and New York: De Gruyter Mouton.
- Myers-Scotton, C. (2004). Precision tuning of the matrix language frame (MLF) model of codeswitching. *Sociolinguistica: International Yearbook of European Sociolinguistics*, 18(1).
- O'Shannessy, C. (2005). Light Warlpiri: A new language. *Australian Journal of Linguistics*, 25, pp. 31–57.
- O'Shannessy, C. (2009). Language variation and change in a North Australian Indigenous community. In: D. Preston and J. Stanford, eds., *Variationist approaches to indigenous minority languages*, 1st ed. Amsterdam: John Benjamins.
- O'Shannessy, C. (2012). The role of code-switched input to children in the origin of a new mixed language. *Linguistics*, 50, pp. 328–353.
- O'Shannessy, C. (2013). The role of multiple sources in the formation of an innovative auxiliary category in Light Warlpiri, a new Australian mixed language. *Language*, 89, pp. 328–353.
- O'Shannessy, C. (2015a). Multilingual children increase language differentiation by indexing communities of practice. *First Language*, 35, pp. 305–326.
- O'Shannessy, C. (2015b). Typological and social factors influencing a new mixed language, Light Warlpiri. In: *Code-switching between structural and sociolinguistic perspectives*. Berlin, Munich and Boston: De Gruyter, pp. 289–304.
- O'Shannessy, C. (Forthcoming). How ordinary child language acquisition processes can lead to the unusual outcome of a mixed language. *International Journal of Bilingualism*.
- O'Shannessy, C. and Meakins, F. (2012). Comprehension of competing argument marking systems in two Australian mixed languages. *Bilingualism: Language and Cognition*, 38, pp. 763–792.
- Pickering, M. and Garrod, S. (2004). Toward a mechanistic psychology of dialogue. *Behavioral and Brain Sciences*, 27, pp. 169–226.
- Poplack, S. (1980). Sometimes I'll start a sentence in Spanish *y termino en español*: Toward a typology of code-switching. *Linguistics*, 18, pp. 581–618.
- Rosen, N. (2006). Language contact and Michif stress assignment. *Sprachtypologie und Universalienforschung (STUF)*, 59(2), pp. 170–190.
- Rosen, N. (2007). *Domains in Michif phonology*. PhD. University of Toronto.
- Sandefur, J. (1979). *An Australian creole in the northern territory: A description of Nukurr-Bamyili dialects (Part I)*. Darwin: Summer Institute of Linguistics, Australian Aborigines and Islanders Branch.
- Sandefur, J. (1991). A sketch of the structure of Kriol. In: S. Romaine, ed., *Language in Australia*, 1st ed. Cambridge: Cambridge University Press.
- Schultze-Berndt, E., Meakins, F. and Angelo, D. (2013). Kriol. In: S. Michaelis, P. Maurer, M. Haspelmath and M. Huber, eds., *English-based and Dutch-based languages*, 1st ed. Oxford: Oxford University Press.
- Smith, I. (1979). Convergence in South Asia: A creole example. *Lingua*, 48, pp. 193–222.
- Smith, I. and Paauw, S. (2006). Sri Lanka Malay: Creole or convert? In: A. Deumert and S. Durrleman, eds., *Structure and variation in language contact*, 1st ed. Amsterdam and Philadelphia: John Benjamins.

- Smith, I., Paauw, S. and Hussainmiya, B.A. (2004). Sri Lanka Malay: The state of the art. In: R. Singh, ed., *Yearbook of South Asian languages*, 1st ed. Berlin and New York: De Gruyter Mouton.
- Smith, N. (1995). An annotated list of creoles, pidgins, and mixed languages. In: J. Arends, P. Muysken and N. Smith, eds., *Pidgins and creoles: An introduction*, 1st ed. Amsterdam: John Benjamins.
- Stewart, J. (2014). A comparative analysis of Media Lengua and Quichua vowel production. *Phonetica*, 71, pp. 159–182.
- Stewart, J. (2015). Intonation patterns in Pijal Media Lengua. *Journal of Language Contact*, 8, pp. 223–262.
- Stewart, J. (2018a). Voice onset time production in Ecuadorian Spanish, Quichua, and Media Lengua. *Journal of the International Phonetic Association*, 48, pp. 173–197.
- Stewart, J. (2018b). Vowel perception by native Media Lengua, Quichua, and Spanish speakers. *Journal of Phonetics*, 71, pp. 177–193.
- Stewart, J. and Meakins, F. Forthcoming. Advances in mixed language phonology: An overview of three case studies. In: M. Mazzoli and E. Sippola, eds., *New perspectives on mixed languages: From core to fringe*, 1st ed. Berlin: De Gruyter Mouton.
- Stewart, J., Meakins, F., Algy, C. and Joshua, A. (2018). The development of phonological stratification: Evidence from stop voicing perception in Gurindji Kriol and Roper Kriol. *Journal of Language Contact*, 11, pp. 71–112.
- The Language Archive. *The Max Planck insitute for psycholinguistics*. Available at: <https://tla.mpi.nl>.
- Thomason, S.G. (1995). Language mixture: Ordinary processes, extraordinary results. In: C. Silva-Corvalán, ed., *Spanish in four continents: Studies in language contact and bilingualism*, 1st ed. Washington, DC: Georgetown University Press.
- Thomason, S.G. (1997a). *Contact languages: A wider perspective*. Amsterdam: John Benjamins.
- Thomason, S.G. (1997b). Introduction. In: S.G. Thomason, ed., *Contact languages: A wider perspective*, 1st ed. Amsterdam and Philadelphia: John Benjamins.
- Thomason, S.G. (1997c). Ma'a (Mbugu). In: S.G. Thomason, ed., *Contact languages: A wider perspective*. Amsterdam and Philadelphia: John Benjamins.
- Thomason, S.G. (1997d). Mednyj Aleut. In: S.G. Thomason, ed., *Contact languages: A wider perspective*. Amsterdam and Philadelphia: John Benjamins.
- Thomason, S.G. (2001). *Language contact: An introduction*. Edinburgh: Edinburgh University Press.
- Thomason, S.G. (2003). Social factors and linguistics processes in the emergence of stable mixed languages. In: Y. Matras and P. Bakker, eds., *The mixed language debate*. Berlin and New York: De Gruyter Mouton.
- Thomason, S.G. and Kaufman, T. (1988). *Language contact, creolization, and genetic linguistics*. Berkeley: University of California Press.
- Vakhtin, N. (1998). Copper Island Aleut: A case of language “resurrection”. In: L.A. Grenoble and L.J. Whaley, eds., *Endangered languages*, 1st ed. Cambridge: Cambridge University Press.
- Van Den Bos, J., Meakins, F. and Algy, C. (2017). Searching for “agent zero”: The origins of a relative case system. *Language Ecology*, 1, pp. 4–24.
- Van Gijn, R. (2009). The phonology of mixed languages. *Journal of Pidgin and Creole Languages*, 24, pp. 91–117.
- Van Rheedeen, H. (1994). Petjo: The mixed language of the Indos in Batavia. In: P. Bakker and M. Mous, eds., *Mixed languages: 15 case studies in language intertwining*, 1st ed. Amsterdam: IFOTT.
- Winford, D. (2013). Social factors in contact languages. In: P. Bakker and Y. Matras, eds., *Contact languages: A comprehensive guide*, 1st ed. Boston and Berlin: De Gruyter Mouton.
- Zentella, A.C. (1997). *Growing up bilingual*. Malden, MA and Oxford: Blackwell.

Linguistic landscape and urban multilingualism

Carla Bagna, Monica Barni, and Martina Bellinzona

1. Introduction

Weinreich (1953), one of the pioneers of the study of ‘modern’ language contact, stated that two or more languages are in contact if used alternately by the same people. As the studies progressed, a more nuanced and articulated concept of language contact emerged (see Winford, 2003; Matras, 2009 for a state-of-the-art description of the field), the study of which can be centred both on individual bilingualism, and therefore from the point of view of speakers, and on societal multilingualism, with the focus shifted on languages and linguistic systems (Berruto, 2009). The manifestations of language contact can be found in a wide variety of domains and, in fact, numerous disciplines and fields of study have dealt with the theme, from acquisitional linguistics to typology, from sociolinguistics to historical linguistics.

This chapter intends to make a further contribution to the study and analysis of language contact, observing it from the point of view of linguistic systems and their coexistence in society, adopting the definition provided by Thomason (2001, p. 1), namely that ‘language contact is the use of more than one language in the same place at the same time.’ In particular, the aim is to analyze the link between urban multilingualism and the so-called Linguistic Landscape (LL), i.e., the written language visible in public space (Gorter, 2006): large modern urban contexts, in fact, are magnets for people and business and represent the most interesting and significant sources for examining and interpreting societal multilingualism. Cities are increasingly becoming ‘showcases’ where signs display users’ choices of languages not only at an individual level, but also in terms of language policy/management/marketing/education, etc. Consequently, we believe that the study of urban multilingualism is fundamental for a greater understanding of the process and the effects of language contact and that LL is one of the most obvious areas, and yet until a few decades ago most ignored, where to observe said contract. LL is used both as an element capable of defining the linguistic arrangement of a district, but also as a ‘tool’ adopted by various subjects (citizens, bloggers, experts, policymakers) to change attitudes, perceptions, expectations towards the neighbourhood, in a kind of continuous narrative.

If the co-presence of different languages in a territory is certainly not a new phenomenon, the superdiversity (Vertovec, 2007) that characterizes urban centres today presents different characteristics (Extra and Yağmur, 2004; Duarte and Gogolin, 2013), also from a linguistic

point of view, with new forms of contact and new linguistic spaces to be explored. The concept of linguistic space can be interpreted both in a synchronic perspective, meaning the set of linguistic varieties forming the linguistic reality of a given area, and in diachronic perspective, that is, the set of linguistic varieties which, over the centuries, have characterized its linguistic-cultural history (De Mauro, 1980). Any speaker can move within this linguistic space in a variety of ways, choosing based on the needs of expressive immediacy or formality, of adhesion to uses and local forms or of wider and general circulation, with motivations different from time to time. In other words, multilingual speakers have at their disposal a complex repertoire of linguistic forms, which are associated, through a process of linguistic socialization, with a range of social activities. It is important to underline that the selection of the appropriate resources does not imply a separation of languages but, on the contrary, the latter interact dynamically in the learning process (Canagarajah, 2011; Creese and Blackledge, 2010; García, 2009). Our hypothesis is that what has been said so far can be observed simply by looking around, walking through the streets of any city in the world, that is, by paying attention to the LL which, according to this perspective, becomes a privileged space in which to analyze multilingualism and superdiversity, processes and consequences of language contact in all its forms.

Before proceeding by retracing the main stages of the expansion of the field of LL and its link with urban multilingualism and language contact, it is important to provide a definition of the concept. The first and one of the most cited definition is that of Landry and Bourhis (1997), which refer to the LL saying:

The language of public road signs, advertising billboards, street names, place names, commercial shop signs, and public signs on government buildings combines to form the linguistic landscape of a given territory, region or urban agglomeration.

(Landry and Bourhis, 1997, p. 25)

Starting from this initial definition, the concept of LL has been reviewed and expanded by scholars and researchers. New objects of study, methods and tools of analysis have emerged and have developed over the last 20 years to satisfy ever-new research objectives and to try to interpret public space. Thus, the definition of LL has also changed, and researchers have opened a long and still unresolved debate on what should actually be considered the object of analysis. While Gorter (2006, p. 2) wanted to emphasize the fact that LL research should focus only on the use of the language in its written form in the public sphere, Ben-Rafael et al. (2006, p. 14) define the LL as any sign or announcement outside or inside a public body or a private company in a specific geographical area. In this way, signs located inside buildings and not only those visible from the street are also included as an object of analysis. Itagi and Singh (2002: xi) have further expanded the concept, stating that LL should not necessarily be interpreted as having a tendency towards written language, but it is to be understood as any linguistic practice taking place in a given context, therefore also oral. However, we agree with Backhaus (2007, p. 4), who writes:

The study object of LL research should be confined to language on signs, since an expansion to other forms of language use in public sphere would water down the usefulness of the concept as a whole.

This position does not intend exclude the relationship between LL and other forms of language use because, as will be seen, the field of LL is closely connected to various disciplines and opens the door to very different perspectives of analysis.

2. Historical overview

The term Linguistic Landscape, as mentioned in the previous section, began to assert itself in 1997, when Landry and Bourhis published the article entitled ‘Linguistic Landscape and Ethnolinguistic Vitality,’ in which the two Canadian researchers gave evidence of the importance of the languages used on signage for the evaluation of the ethnolinguistic vitality of the different ethnic groups present in a territory. In reality, the notion and relevance of LL had previously emerged in the field of Language Policy and Planning (LPP), particularly in territories in which the presence of different linguistic communities had created tensions and conflicts. In Belgium, for example, where there is a total division between Wallonia and Flanders from the point of view of the linguistic management of public services and administration, it is immediately possible to identify the border between the two territories through road signs. It is not surprising, therefore, that Belgium was the scene of one of the first researches on LL. Tulp (1978), in fact, had conducted a study on the LL of Brussels, analyzing the large billboards visible along the urban routes of the most frequented trams and buses and the major metro stations, as her purpose was to demonstrate how the use of French had contributed, over time, to the gradual Francization of the city. With her study, therefore, emerged the concept that the visibility of a language in public space is essential for its ethnolinguistic vitality, real and perceived.

Among the first contributions of LL, prior to the work of Landry and Bourhis, also Rosenbaum et al. (1977) and Monnier (1989) deserve a mention. Both were interested in addressing the phenomenon of multilingualism in terms of languages visible on signs in urban contexts, albeit the purpose of their studies was very different. Rosenbaum et al. (1977) aimed to detect the presence of English in the public sphere in Jerusalem, to understand to what extent and why it was used. From their study, the tendency to exploit the so-called snob appeal of English emerged: the greater the cost of the goods or the service offered, the greater the likelihood of encountering signs in this language. Monnier (1989), on the other hand, focused on the study of the languages used on commercial signs in Montréal, Québec. His goal was to evaluate to what extent the signs posted by individuals were in line with the provisions of the Charter of the French Language of 1977 (Bill 101). From his analysis, it emerged that, although French was predominant in quantitative terms, the type of store and the geographical distribution could be decisive in the choice of the employed language. He also observed the existence of a possible relationship between the language of the sign and the language spoken in the shop: he realized, in sum, how the signs did not merely serve to convey a linguistic content, but also contained a message about the author.

Another example worthy of mention is the research carried out by Spolsky and Cooper (1991) in the Old City of Jerusalem in which the two scholars propose a model of the rules that determine linguistic choices, which are rooted in a social context. In this sense, through the study of Arabic, Hebrew, and English toponymy of the streets, it was possible to identify the various occupations and dominions that have marked the history of the city in the last century. This is a relevant study for the treatment proposed here, as it highlights the strong relationship between LL and power relations between communities (and therefore languages) in competition, if not in conflict, among themselves in a territory. As will be shown in the next sections, in fact, power relations are a recurring and central topic in LL studies, whose implications are not limited to the linguistic field.

The researches mentioned so far have traced the path to a development of the LL approach as it is understood today, whose centrality in the studies on urban multilingualism is, however, only to be found with the publication of the Special Issue of the International Journal of

Multilingualism, edited by Gorter (2006). In this collection, entitled 'Linguistic Landscape: a new approach to multilingualism,' four case studies are proposed, conducted in various parts of the world, from Israel to Thailand, from Japan to Spain, passing through the Netherlands. It is interesting to note that one of the essays in this collection, Huebner (2006), is expressly oriented to the analysis of the effects of language contact (language mixing and language dominance) on the LL of Bangkok, a strongly multilingual contest. From the analysis emerges, on the one hand, the shift from English to Chinese, intended as wider communication languages, and on the other it is shown how English itself influences Thai both from a lexical point of view, with various examples of borrowings, in spelling, pronunciation and syntax. This and the others are researches conducted according to a quantitative/distributive approach: after collecting the data, consisting of photographs of linguistic signs, the analysis consisted of counting which languages appeared on the signs, and how frequently, so as to obtain a macro-overview of the languages used for public written communication in a territory and the functions attributed to them. In particular, the distinction is proposed between top-down, i.e., public signs (for example road signs, toponyms, inscriptions on government buildings) and bottom-up signs, i.e., the private ones (for example billboards, shop signs, and later also graffiti). Looking at the differences in the distribution of languages between these two types of signs, in fact, it is possible to investigate the actual role of language policies and their scope within a community of speakers, power relations, as well as linguistic ideologies and attitudes.

Equally important is the taxonomy proposed by Reh (2004), who distinguishes multilingual signs based on translation strategies or, better, typology of language arrangement in four categories: duplicating, fragmentary, overlapping and complementary. This taxonomy is particularly relevant as it allows to observe language contact both from a formal perspective, from the structural properties of languages, and from a functional perspective, investigating in which contexts, with what modalities and why a language be used on multilingual signs.

Already from these first studies the importance of LL emerges as a fundamental tool for a better understanding of urban multilingualism. Subsequently, further collections and volumes were published that dealt with the LL and the relationship with urban multilingualism, with descriptions and analyses of the phenomenon according to different perspectives. Among these, it seems important to mention the collection of Shohamy and Gorter (2009), in which the discussion on the definition of LL and the boundaries of this field of study is summarized, making explicit its link with many disciplines, from history to sociology, from linguistics to the economy, and the topic of research methodology is addressed. The different chapters look at the link between LL and linguistic ideologies, language policies and language awareness in the most diverse contexts, opening up to many avenues for further study and research.

Shohamy, Ben-Rafael and Barni (2010), as the title of the collection suggests, 'Linguistic Landscape and the city,' is centred on urban multilingualism and the various chapters highlight how LL is the scene in which the public space is symbolically constructed. Jaworski and Thurlow (2010) shift the focus to the intrinsic semiotic aspect of the study of signs in public space, dealing with issues related to tourist flows and urban geography. As stated by the authors, the goal of the collection is to put in the foreground 'the interplay between language, visual discourse, and spatial practices and dimensions of culture, especially the textual mediation or discursive construction of space and use of space as a semiotic resource in its own right' (Jaworski and Thurlow, 2010, p. 1). As regards semiotic landscape, seminal is the work of Scollon and Scollon (2003), 'Discourses in place: language in the material world,' a fundamental text for anyone involved in LL, language and communication, in which the authors developed the 'first systematic analysis of the ways we interpret language as it is

materially placed in the world' (Scollon and Scollon, 2003: xi) and introduce the concept of geosemiotics, understood as the relationship between space and social meaning, captured by the interplay of an interactional order (action), visual semiotics (the appearance of sign) and place semiotics (location of sign at a given point in space).

Other collections of essays and case studies have opened new LL study perspectives. Among these, Gorter, Marten and Van Mensel (2012), in which the focus is shifted on the visibility of minority languages in the LL, investigating the complex relations between languages, society and identity, making explicit the role of LL in spreading ideologies and strengthening power relations. Hélot et al. (2012) aim to investigate in a deeper way the processes of language contact and multilingualism in local and global contexts, again through the study of LL.

Since 2015, in order to cope with a rapidly expanding area of study, a scientific journal dedicated to the LL has been promoted: 'Linguistic Landscape. An international journal.' Finally, a review of LL studies, published in English can be consulted on Zotero (2013; see reference list).

3. Critical issues and topics

As highlighted in the previous paragraph, in the last decades the theme of LL and of urban multilingualism, understood as a space in which the signs of language contact manifest themselves more extensively, has aroused a growing interest in research.

The great versatility of the LL represents an element at the root of the increase in research in this regard. In fact, urban LL appear to be able to provide indications on multiple aspects of multilingual environments starting from the ethnolinguistic vitality of resident communities (the 1997 Landry and Bourhis study aimed to investigate this aspect) to issues related to identity, from ideologies promoted through signs to relations of power between languages (and therefore communities) competing in a territory.

Based on the corpus of articles and books available on Zotero (see the previous section) and to the interventions presented in international conferences, it is possible to identify some of the themes most investigated and addressed. Researches on urban multilingualism through the LL has allowed to explore in more depth the entity and the underlying motivations of the diffusion of English and other languages on a global level, the effects of language policies, reactions (both top-down and bottom-up) to the presence of minority and lesser-used languages, immigrant languages and the differences between the linguistic choices promoted. Globalization and tourism, the intensification in migration and the processes of revitalization of minority languages have made increasing numbers of people, goods, ideas, cultures, and, of course, languages come into contact, meeting and clashing in increasingly multilingual cities. This superdiversity (Vertovec, 2007) finds space, is reflected and multiplied in the LL, whose observation can contribute to the study of the phenomenon.

A relevant aspect emerged from the study of urban multilingualism through the LL is the visibility of lesser-used languages (LUL). Thanks to the commitment of the European Union and the Council of Europe to the promotion of languages and linguistic diversity, also and above all in reference to minority languages, as well as thanks to the renewed global interest in these issues, attempts to revitalize and disseminate LUL are growing. LL appears not only as one of the first areas of intervention of language policies oriented in this direction, but also as the privileged space in which individuals, groups and communities can make their identity manifest, first of all their linguistic identity. The presence of LULs in the LL, with the aim of highlighting the identity and vitality of a community, is, in fact, identified as one of the main functions of their use (Agnihotri and McCormick, 2010). At the same time, however,

in several contexts (see for example the work of Coupland and Garrett (2010) on Welsh in Patagonia) a commodification of language for commercial purposes was noted. In this sense, the presence of a local variety or a minority language aims to give a taste of authenticity, can transmit the values of the rustic, of the genuine, of simplicity, of orientation towards the past and of local tradition, to assert uniqueness.

The reasons, instead, for the great visibility of English, understood as *lingua franca*, the language of international communication, are different. The presence of this language in the LL of almost all the cities of the world is due, in the first place, to the informative function of public signage: texts primarily serve to communicate, to provide an indication of the place, of the goods sold or of the service offered, to give instructions or indicate a ban. English, see Bruyèl-Olmedo and Juan-Garau (2009) in this regard, is used to address tourists or speakers of languages other than the national one. Numerous studies (e.g., Griffin, 2004), however, have shown that its large use is not due solely to purely informative needs, but must be sought in its symbolic function. Choosing the name of a business or product in English and using it on signs and billboards activates values such as an orientation which is international, towards the future, of success, leisure, and refinement (Piller, 2003). If we take Griffin's study (2004), conducted in Rome, the presence of English, in 75% of cases, results as limited to a few words, which means that the transmission of the most important messages is left to Italian. In most of the examples, English was found to be 'a device to establish a trendy cosmopolitan image to native Italian speakers' (Griffin, 2004, p. 7). It is therefore not its function as a *lingua franca* to decree its strong presence on the signs, but the fact that it is seen as a fashionable and attractive language.

With regard to LPP actions, it has already been highlighted how important the LL can be in terms of both information and practicality, and above all, symbolic. Language policy interventions on signage have always primarily regarded top-down signs. Especially where minority languages exist or where there are language conflicts between different communities, governments have included signposting among the priorities of their policy (consider the case of Belgium, already discussed by Landry and Bourhis, 1997). This is particularly evident in the case of toponyms, i.e., the proper names of geographical places, which have always been at the centre of tensions in territories inhabited by linguistic minorities. As for Italy, think for example what happened in South Tyrol at the time of Fascism, where all toponyms of German origin underwent a process of Italianization to justify first the military occupation and then the political annexation to the Kingdom of Italy. Or consider Lombardy, where the Lega Nord (Northern League) party fought a battle in the 1970s and 1980s to have toponyms translated into the local language (for more information see Puzey, 2012). In some cases, however, actions have been taken to regulate the use of some languages even on bottom-up signs: staying in Italy, in different municipalities (for example, Prato and Rome), with the aim of regulating the presence of some businesses, the use of the language on signs has also been controlled, prohibiting the exclusive use of languages other than the national one on signs (for more details see Barni and Bagna, 2010). From these examples it is clear how the LL plays a key role in the political debate, becoming itself an arena in which it is possible to observe language conflicts which, in turn, imply conflicts and interests of a completely different nature.

In the face of a growing body of research, also confirmed by global workshops on the theme and an increasing number of presentations in various conferences (LL Workshops, Sociolinguistic Symposium, AILA, AAAL, etc.) there remains open a question related to the specific object of the LL: is it a method or an instrument, a framework or a field of study? As expressed by Spolsky (2009, p. 25): 'Is linguistic landscape a phenomenon calling for a theory, or simply

a collection of somewhat disparate methodologies for studying the nature of public written signs?’

From a theoretical point of view, it is necessary to consider the multidisciplinary nature of the study of LL: being used in different fields, it can be linked to sociolinguistic theories, language policy, theories related to language contact, as well as urban planning. Some scholars have applied existing theories, as in the case of Ben-Rafael et al. (2006) who, in their analysis of the linguistic choices of the LL actors, adapt the principles already proposed in the sociological field by Goffman (1963, 1981) and Boudon (1990). In recent years, moreover, more scholars have linked the study of LL to interactional pragmatics, considering linguistic signs as speech acts, i.e., as localized communicative events (Kallen, 2009) inserted in a discursive context. This is the case of Collins and Slembrouck (2007) who, after analyzing a variety of data from an ethnographic study of language contact in Ghent (Belgium), define signage as a contextual act. Interviewed readers, in fact, interpret the multilingual choices on signs in various ways, making use of their repertoire of linguistic forms and cultural knowledge, giving to those linguistic choices an ‘indexed order,’ a hierarchical structure. Malinowski (2009), in turn, combines the theory of linguistic acts with the notion of language as symbolic capital, proposed by Bourdieu (1991), identifying in signs a performative power and an intended effect, whose success depends on the reader’s ability to decode. Matras, Gaiser and Reershemius (2018), analyzing the presence of the Yiddish language in Manchester, join this panorama, to propose a typology of signs that can be placed along a continuum, which represents different levels of authority and strength through which the sign producer controls the effect and the receiver actions.

What has been said so far makes it clear how the inhabitants of an area, the potential speakers or readers, initially perceived as almost absent or as viewers rather than authors of the LL, have taken an increasingly active role in the analysis of the LL, becoming involved as its first interpreters (Pappenhagen, Scarvaglieri and Redder, 2016; Shohamy and Waksman, 2012). This greater attention to the role of the LL actors, whether they are sign makers, authorities who commission them, customers, passers-by or can also be seen from the methodological developments that have occurred in recent years. In fact, there is a rise in research conducted according to mixed method, using not only visual data taken from the LL, but also interviews (Papen, 2012), participatory observations (Lou, 2016) and questionnaires (Bagna, Gallina and Machetti, 2018), useful for obtaining information on the perception, awareness, attitudes and language ideologies of those who live the LL daily (see Section 5 for further information on methodology and research tools).

4. Current contributions and research

As already underlined in Barni and Bagna (2015), ever since the first surveys, it has been clear that the LL is a complex, multifaceted phenomenon and the potential of LL studies has been increasingly highlighted, leading to developments in a number of different areas, related to the object and subjects involved (as pointed out in the previous section), the relationship with other disciplines and the choice of the most suitable methodologies.

At this point, it is useful to present two studies, the first conducted by Backhaus (2006) and the second by Blommaert (2013) which, in our opinion, can be taken as paradigmatic respectively of the first and second wave of research on LL and that, therefore, can serve as a demonstration of how and in what direction the concept of LL has evolved. Backhaus’ research, which he presents in the special issue of the *Journal of Multilingualism* edited by Gorter (see Section 2), aims to investigate multilingualism in Tokyo, a metropolis described by

the media as essentially monolingual. To verify if this is true, the scholar chooses to conduct the investigation in the streets immediately adjacent to the 28 stations on the Yamanote train loop in the city centre and collects 11,834 signs, defined as any piece of written text within a spatially definable frame, of which he analyzes the multilingual ones (2,444). The analysis that he leads then is of a statistical nature, with the focus on official/non-official status of the signs, code choice and preference, with the aim of highlighting the differences and the relationship between the two types of multilingual signs, i.e., those top-down and those bottom-up. His interpretation of the results sees the linguistic choices on the top-down signs, on which only Japanese, English, Chinese and Korean appear, as an expression and an attempt to strengthen the power relations in Japan: he argues, in fact, that the mandatory inclusion of Japanese on official signs is ‘asserting the power of the national language over the designated space,’ while the widest range of languages on bottom-up signs would be oriented to expressing solidarity (Backhaus, 2006, p. 62). This is especially true as regards English (and to a lesser extent French), whose presence ‘can be interpreted as a symbolic expression by Japanese sign writers to join the English language community and to associate with [its] values’ (Backhaus, 2006, p. 63). This study, although taken as a reference, must be traced back to the context of the research in which it is carried out, given that in other areas of the city of Tokyo, such as Ginza, in the same years the Italian language was the most visible, alongside Japanese and English, as evidenced by the collections on Italianisms carried out (Bagna and Barni, 2007).

While the study of Backhaus presents a purely quantitative research, in which signs are seen as static objects, the actors of the LL as somehow passive agents and the approach of the LL itself as synchronic, different is the case of Blommaert’s (2013) work, in which the theoretical framework of ethnography, or rather a ‘historically sensitive sociolinguistics ethnography,’ is applied to the LL study. In the book, entitled ‘Ethnography, superdiversity and linguistic landscapes: chronicles of complexity,’ the analysis of the LL of a district of Antwerp (Belgium) is conducted not in quantitative terms (although the scholar has been collecting data since 2007), but qualitative and through a strongly interdisciplinary approach, inspired among others by the work of semiotics, nexus analysis (Scollon and Scollon, 2003, 2004) and anthropology (Fabian, 1983). The aim of the research, in fact, is not to describe the multilingualism of the city, but to demonstrate firstly how multilingual signs can be read as *chronicles* that document the complex histories of a place and, secondly, that *complexity* itself can be seen as *order*, an order that is multiscale and dynamic. To do this, for example, Blommaert compares three posters, the first in Turkish, produced in a professional manner, the other two in Polish and Spanish and done in a ‘homemade’ manner. Analyzing the placement, the style and the information contained therein, Blommaert traces the story of the communities themselves and the intended public: while the first sign indicates a community of migrants, well-connected and financed and residing in the territory for a long time, the other two publicize groups of local communities of recent formation and also underline their modest resources. As already highlighted by Pavlenko (2010, p. 133), ‘LL is not a state, but a diachronic process and the meaning of the present day’s arrangements cannot be fully understood without considering those of the past’ and the work of Blommaert makes evident both the complexity of the LL, and the usefulness of its analysis.

As shown by the examples just reported, the range of different observable signs in the LL and the boundaries of what can be defined as LL or as a part of it are much broader than the six linguistic objects identified by Landry and Bourhis (see Introduction). LL is the arena and area where the negotiation of rights and identities is wrought by way of mutable and not merely static linguistic choices within a territory. Bodies, objects, and symbols are more fully incorporated into the analysis of LL. Peck and Stroud’s study (2015) is a demonstration of this

because, through a ethnographic investigation (which includes participant observation, interviews, field notes collected in six months at a tattoo shop in Cape Town) into the affective and performative means by which bodies and places write each other, the scholars ask how ‘bodies that matter’ in historically and racially divided places are created through the semiotics of and on the skin. From their results it emerges how tattooing, as bodily inscription, joins participants’ lived experience of time and place with social discourses of race, gender, class, occupation, and other areas of identity formation. It is interesting to note that, for the analysis, the two scholars have used not only linguistic elements, but also postural, gaze, tactile, and affective features. If the so-called Skinscape (Peck and Stroud, 2015) is undoubtedly a new frontier of LL, equally innovative, albeit with a more consolidated tradition, is the study of the semiotic dimension of the so-called Graffscape (Bilkić, 2018), that is a landscape constituted by graffiti, usually not authorized or seen as transgressive. Pennycook (2009, p. 302), in an essay which has become a classic in the LL literature, states that ‘an understanding of graffiti as *transgressive urban semiotics* opens up important directions for an understanding of linguistic landscapes (LL) more generally’ (our italics). Interpreting graffiti, and the act of writing itself, as an act of appropriation of space, of identity and narration, pride and rebellion, allows to take a step forward in understanding the LL itself, which can be thought of in terms of the five I’s: integration, identity, imagination, illocution, interpellation (Pennycook, 2009, p. 309).

It is therefore clear how, from 1997 to the present, the approach to LL, the methodologies adopted, the role of people and texts involved, in addition to the impact of the LL, has become increasingly wide-ranging and has acquired greater breadth and depth. As has been stated, the topics covered have also changed over time. During the 10th Linguistic Landscape International Workshop (Bern, 2–4 May 2018) the most recent research lines emerged, with a special space dedicated to the theme of urban transformations and the effects of gentrification processes in different cities of the world (Papen, 2012; Leeman and Modan, 2009; Trinch and Snajdr, 2016). Gentrification is defined as a set of urban and social transformations of such impact to make the area in which they occur a new, more expensive and exclusive one (Semi, 2015). Gentrification involves changes not only in terms of economic and social capital, but also on the linguistic level. Different languages, referring to opposing symbolic capitals, for example English, which evokes globalization, and dialects, which enhances the local cultural heritage, meet and give rise to language contact phenomena that have not been studied so far.

In addition to the urban LL, to the transformations that involve the city and therefore its linguistic aspect, increasing attention is paid to the LL of semi-public and private spaces, such as offices, hospitals, and schools. If the multilingualism deriving from globalization and migration flows manifests itself on the signs visible in the city, it cannot help but be reflected also within the school walls, nursery and universities, and therefore today scholars discuss the Linguistic Schoolscape, understood as ‘the school-based environment where place and text, both written (graphic) and oral, constitute, reproduce and transform language ideologies’ (Brown, 2012, p. 282). In other words, the ‘visual and spatial organization of educational spaces, with special emphasis on inscriptions, images and the arrangement of the supplies’ (Szabò, 2015, p. 24) becomes an area of research and study closely connected to the dynamics of urban LL. Research in this sense has been oriented to the study of the multilingual dimension of schoolscape related to geographic and political areas already involved in specific processes of management of linguistic minorities (Gorter and Cenoz, 2014; Brown, 2012) or to monitor the presence of immigrant languages (Bellinzona, 2018). At the same time, numerous studies have demonstrated the potential of urban LL as a tool for the development of language awareness and for the learning of a second language. Many of these studies were mainly carried out

at a theoretical level (Cenoz and Gorter, 2008; Chern and Dooley, 2014; Malinowski, 2015) with the aim of highlighting the skills developed by exposing learners to urban multilingualism. In addition to the possibility of incidental learning, students, observing and analyzing the LL, could increase their pragmatic, multimodal and symbolic competence. However, practical applications were not lacking (Rowland, 2013; Chesnut, Lee and Schulte, 2013), which highlighted the advantages to which the use of LL lead for second language learning (Sayer, 2009), for the development of intercultural competence or for the increase of awareness of language diversity (Dagenais et al., 2009; Clemente, Andrade and Martins, 2012; Hancock, 2012).

5. Main research methods

In the previous paragraph it has been shown how, in recent years, there has been a ‘critical turning point’ in LL research, with a shift from purely descriptive/quantitative studies to more qualitative surveys. This obviously involved a change also in terms of methodology and tools for data collection and analysis. The main purpose remained, however, that of providing an analysis of how languages (or other semiotic systems) are used visually in multilingual urban contexts, and thus the data to be collected consist primarily of written texts present in the public space. These, in most cases, are obtained from surveys in the field and ‘linguistic walks,’ which allow the researcher to immerse herself or himself completely in the area under study. Data collection takes place today with the use of digital tools, such as smartphones or specific apps, which have replaced manual transcription, camera and other devices.

But what signs constitute the research unit? The large number of occurrences and texts close to each other (think of the shop window for example), the dynamic nature of LL (some signs are fixed, others change from day to day, others are even moving such as advertisements on buses, vans, and cars) and the lack of visible borders (for example in graffiti) make it difficult to establish what is meant by sign unit. Is it necessary to consider the sum of all the texts that are on the same surface (a shop window, a wall) or in the same place (a whole street, the inside of a building) as a single occurrence or should each text be considered separately? Each of these options has advantages and disadvantages depending on the context of the investigation and, in fact, studies has sometimes been conducted in one way, sometimes in the other.

As for the research method, one of the first attempts to give scientific grounding to the method is identifiable in the work of the Centre of Excellence in Research – Permanent Linguistic Observatory of the Italian Language among Foreigners and of Immigrant languages in Italy of the University for Foreigners of Siena. This centre has been engaged since 2004 in the analysis and mapping of the languages of immigrant groups based in Italian territory. What is called the ‘Esquilino method,’ since it was used for the first time in this district of Rome (Bagna and Barni, 2006), consists in the combined use of cameras and laptops connected directly to software for geo-referencing and immediately processing some of the linguistic data collected on site. This system (MapGeoLing 1.0.0 and 1.0.1 extension described in Barni and Bagna, 2009), allowed to collect a large amount of data in a very short period of time and facilitated the creation of linguistic maps, useful both for synchronic and diachronic studies.

The tools currently used for data collection have advantages: easy use; ease of integration and implementation of databases even by non-experts; possibility of obtaining a larger amount of data. Equally, the creation of special apps, able to immediately geolocate the sign once the photograph has been taken, increased the number of people interested (even for a

short time) in experimenting with micro-activities to include ethnographic linguists, geomatic linguists, etc. Among the apps we mention the one developed by the University of Manchester, *LinguaSnapp*, within the broader project ‘Multilingual Manchester,’ whose purpose is to study urban multilingualism and develop awareness of the linguistic diversity present in that and other context. In fact, although it was initially used in Manchester, later versions have been adopted for many other cities, both European and non-European, as can be seen from the interactive map available on the website. Another app is *Lingscape*, developed by the University of Luxembourg, whose purpose, as reported by the site, is ‘to analyse the diversity and dynamics of public writing.’ Both give the opportunity to take pictures directly from the app, geolocate and edit them, add notes, tag the languages and even the alphabet in the detected texts.

The discussion about the use of different methodologies is central: which methods for which goals (Gorter, 2006; Shohamy and Gorter, 2009)? It has become clear that in the LL approach the methods used to collect and analyze data can offer different interpretations for different disciplines. For sociological analysis, interviews are more pertinent than purely static LL items; for language policy, the quality of the LL items found can determine other consequences for decisions makers; in ethnographic studies, it is necessary to combine qualitative and quantitative data in order to understand the appearance and development of certain phenomena. In the evolution of LL research, the discussion about methods is still central, due both to the possibility of employing the innovative tools offered by developments in technology, often highly sophisticated or capable of recording all the LL data, as well as to the widening use of LL as an approach for recording and then analyzing different phenomena. It is therefore necessary to reflect on the need to achieve a balance when engaging with an LL that is constantly accessible, subject to endless comment (just think of social networks) and used without filters.

6. Future directions

As mentioned in the Introduction, cities are increasingly becoming ‘showcases’: walking through the streets of any city in the world, it is easy to realize how many texts are visible. There is a predominance of visual information, which is increasingly multilingual: it is now almost impossible to find a monolingual LL, even if only for ubiquitous international brands, ethnic shops and restaurants and tourist information. How much individual citizens (in their capacity as residents or temporary local users) are aware of this linguistic diversity that surrounds us is unclear: studies related to (critical) language awareness regarding LL are still very limited. However, in the latest edition of the *Encyclopedia of Language and Education*, particularly in the volume dedicated to *Language Awareness and Multilingualism*, an entire chapter is dedicated to this topic (Gorter and Cenoz, 2017). It seems, therefore, that there is an increase in interest in this direction, as is also evidenced by the *Multilingual Manchester* project, and studies on the impact of LL in educational contexts.

The theme of teaching and the link between LL and educational contexts is one of the topics that will receive evermore attention, given the potential of using urban multilingualism as a stimulus for language learning and the development of skills, both linguistic and non-linguistic.

As has been seen, most of the LL studies have focused so far on the analysis of certain neighbourhoods and cities and have consisted of a mapping of the detected texts, mostly static elements. It should be remembered, however, that the LL is also formed by moving subjects

and texts: placards posted on buses or trams, written on taxis but also on t-shirts or on the skin of passers-by, displays that change the promotional message continuously or devices that make use of augmented reality. This dynamic nature of the LL is a challenge, and, in the future, researchers will be increasingly involved in the creation of appropriate methodological models for the study of this LL dimension. In general, the development of the LL as a research field consisted of an increase in the number of descriptive researches, of case studies of individual contexts. This entailed the definition of theoretical and methodological frameworks which were always innovative and different according to the research objectives and multiple aspects of urban multilingualism. If on the one hand this characteristic of LL is undoubtedly an advantage and a strength, on the other hand it could constitute a limitation, as it allows researchers too much freedom to manoeuvre and there is a risk of resorting to less rigorous approaches. In the face of increasingly numerous research combinations, in which the LL approach remains central, we believe that at the theoretical level the theme of languages, of relations between languages, of linguistic analysis, should remain a priority. This is both for the implications for the individual territories that these analyses may provide, and so as not to reduce the languages to one of the many objects of the Landscape, while the focus of the discipline is Linguistic Landscape.

With this contribution, an attempt has been made to show how the study of LL can be useful for a better understanding of multilingualism in urban settings and to grasp many aspects of the contact between languages, cultures, identities and society. Cities today, as a place of meeting/clash of individuals, groups and communities, are the ideal place to observe the effects of language contact, which manifests itself in the LL in all its forms, with examples of borrowings, code-switching and transfers on all levels of the language (lexical, but also morphological, syntactic, pragmatic and even writing systems). LL, as a privileged space which also allows for forms of creativity, is full of signs characterized by the contact between different cultures and languages. Analyzing a LL through the tools and theoretical frameworks belonging to language contact can provide relevant historical, social, cultural, and of course linguistic information. Think of Blommaert's work (2013), which was explained in Section 4: through a diachronic analysis of the LL, it is possible to observe the signs of contact between immigrant and national languages, transfer phenomena between them and their evolution due to prolonged contact. Likewise, consider the potential of the taxonomy proposed by Reh (2004), which is useful for investigating effects on languages structural properties. In addition, it allows shifting the focus towards a more functional perspective: see in this sense the works of Huebner (2006) as well as all those oriented to the study of language policies, ideologies, and power relations, which have been taken into account during this work.

To conclude, we would like to mention the study conducted by Woldemariam and Lanza (2014) in virtue of its representativeness of LL and language contact conjunction. In fact, through the analysis of the LL of two regional capitals in Ethiopia and, in particular, of the use of Amharic (national language) as opposed to that of Tigrinya and Oromo (regional languages, only recently become literacy languages), the scholars firstly observe structural contact phenomena and then interpret them as linked to questions of agency and power in the domain of multilingual practices.

We have introduced this work by recalling how multilingualism and superdiversity are phenomena now common to all large (but often also small) cities in the world and we have adopted a vision of language as a social activity, for which multilingual speakers have 'a complex repertoire of linguistic structures at their disposal' (Matras, 2009, p. 4), which allows them to use

their languages in everyday life, including ‘city life.’ This perspective empowers speakers, as it makes them active agents in adapting their competence to new linguistic practices: LL, as a symbolic and ideologically charged space, appears as the ideal place where we can observe all that, power issues, relations and hierarchies, language attitudes, and ideologies.

7. Further readings

Martin-Jones, M., Blackledge, A. and Creese, A., eds. (2012). *The Routledge handbook of multilingualism*. New York: Routledge.

This handbook deals with the issue of multilingualism across political and historical contexts, and in relation to education and other institutional sites, social and cultural change. Particular emphasis is placed on the use of ethnographic techniques for the study of urban and social multilingualism.

García, O., Flores, N. and Spotti, M., eds. (2017). *The Oxford handbook of language and society*. New York: Oxford University Press.

The 27 chapters of this book, which is strongly interdisciplinary and with global contributions, challenge the basic concepts of sociolinguistics, proposing a critical poststructuralist perspective that examines the socio-historical context and develops new theoretical and methodological tools that challenge the dominant concepts.

Pütz, M. and Mundt, N. (2018). *Expanding the linguistic landscape: Linguistic diversity, multimodality and the use of space as a semiotic resource*. Bristol: Multilingual Matters.

In this book, LL is viewed as a metaphor and expanded to include a wide variety of discursive modalities. Contributions, from different world contexts, show new ways of renegotiating concepts such as language and landscape, in the light of multimodality, language power, conflict, discourse, and translation.

Laihonen, P. and Szabó T., eds. (2018). Studying the visual and material dimensions of education and learning (special issue). *Linguistics and Education*, 44.

In this special issue, established and emerging scholars deal with the visual and material dimensions of education, exploring the linguistic schoolscape of different contexts from a conceptual, theoretical, and practical point of view and investigating how people create, interpret, and negotiate meanings in the multilingual environments of learning.

8. Related topics

Urban youth speech styles in multilingual settings, pragmatic factors

Abbreviations

| | |
|-----|------------------------------|
| FL | Foreign Language |
| L2 | Second Language |
| LL | Linguistic Landscape |
| LPP | Language Policy and Planning |
| LUL | Lesser-used language |

Authors' contribution

C. Bagna takes responsibility for Sections 2 and 5; M. Barni for Section 1; M. Bellinzona for Section 3 and 4.

References

- Agnihotri, R.K. and McCormick, K. (2010). Language in the material world: Multilinguality in signage. *International Multilingual Research Journal*, 4(1), pp. 55–81.
- Backhaus, P. (2006). Multilingualism in Tokyo: A look into the linguistic landscape. *International Journal of Multilingualism*, 3(1), pp. 52–66.
- Backhaus, P. (2007). *Linguistic landscape: A comparative study of urban multilingualism in Tokyo*. Clevedon: Multilingual Matters.
- Bagna, C. and Barni, M. (2006). Per una mappatura dei repertori linguistici urbani: Nuovi strumenti e metodologie. In: N. De Blasi and C. Marcato, eds., *La città e le sue lingue. Repertori linguistici urbani*, 1st ed. Napoli: Liguori, pp. 1–43.
- Bagna, C. and Barni, M. (2007). La lingua italiana nella comunicazione pubblica/sociale planetaria. *Studi Italiani di Linguistica Teorica e Applicata*, XXXVI(3), pp. 529–553.
- Bagna, C., Gallina, F. and Machetti, S. (2018). L'approccio del Linguistic Landscape applicato alla didattica dell'italiano L2 per studenti internazionali. In: C.M. Coonan, A. Bier and E. Ballarin, eds., *La Didattica delle lingue nel nuovo millennio: Le sfide dell'internazionalizzazione*, 1st ed. Venezia: Edizioni Ca' Foscari, pp. 219–231.
- Barni, M. and Bagna, C. (2009). A mapping technique and the linguistic landscape. In: E. Shohamy and D. Gorter, eds., *Linguistic landscape: Expanding the scenery*, 1st ed. New York: Routledge, pp. 126–140.
- Barni, M. and Bagna, C. (2010). Linguistic landscape and language vitality. In: E. Shohamy, E. Ben-Rafael, and M. Barni, eds., *Linguistic landscape and the city*, 1st ed. Bristol: Multilingual Matters, pp. 3–18.
- Barni, M. and Bagna, C. (2015). The critical turn in LL: New methodologies and new items in LL. *Linguistic Landscape*, 1(1), pp. 6–18.
- Bellinzona, M. (2018). Linguistic landscape e contesti educativi. Uno studio all'interno di alcune scuole italiane. *Lingue e Linguaggi*, 25, pp. 297–321.
- Ben-Rafael, E., Shohamy, E., Hasan Amara, M. and Trumper-Hecht, N. (2006). Linguistic landscape as symbolic construction of the public space: The case of Israel. *International Journal of Multilingualism*, 3(1), pp. 7–30.
- Berruto, G. (2009). Confini tra sistemi, fenomenologia del contatto linguistico e modelli del code switching. In: G. Iannaccaro and V. Matera, eds., *La lingua come cultura*, 1st ed. Torino: UTET, pp. 3–34.
- Bilkić, M. (2018). Emplacing hate: Turbulent graffscapes and linguistic violence in postwar Bosnia-Herzegovina. *Linguistic Landscape*, 4(1), pp. 1–28.
- Blommaert, J. (2013). *Ethnography, superdiversity and linguistic landscapes: Chronicles of complexity*. Bristol: Multilingual Matters.
- Boudon, R. (1990). *La place du désordre. Critique des théories du changement social*. Paris: Presses universitaires de France.
- Bourdieu, P. (1991). *Language and symbolic power*. Cambridge: Harvard University Press.
- Brown, K. (2012). The linguistic landscape of educational spaces: Language revitalization and schools in southeastern Estonia. In: D. Gorter, H. Marten and L. Van Mensel, eds., *Minority languages in the linguistic landscape*, 1st ed. New York: Palgrave Macmillan, pp. 281–298.
- Bruyèl-Olmedo, A. and Juan-Garau, M. (2009). English as a lingua franca in the linguistic landscape of the multilingual resort of S'Arenal in Mallorca. *International Journal of Multilingualism*, 6(4), pp. 386–411.
- Canagarajah, S. (2011). Translanguaging in the classroom: Emerging issues for research and pedagogy. *Applied Linguistics Review*, 2, pp. 1–28.
- Cenoz, J. and Gorter, D. (2008). The linguistic landscape as an additional source of input in second language acquisition. *IRAL-International Review of Applied Linguistics in Language Teaching*, 46(3), pp. 267–287.
- Chern, C. and Dooley, K. (2014). Learning English by walking down the street. *ELT Journal*, 68(2), pp. 113–123.

- Chesnut, M., Lee, V. and Schulte, J. (2013). The language lessons around us: Undergraduate English pedagogy and linguistic landscape research. *English Teaching: Practice and Critique*, 12(2), pp. 102–120.
- Clemente, M., Andrade, A. and Martins, F. (2012). Learning to read the world, learning to look at the linguistic landscape: A primary school study. In: C. Hélot, M. Barni, R. Janssens, and C. Bagna, eds., *Linguistic landscapes, multilingualism and social change*, 1st ed. Frankfurt: Peter Lang, pp. 267–285.
- Collins, J. and Slembrouck, S. (2007). Reading shop windows in globalized neighborhoods: Multilingual literacy practices and indexicality. *Journal of Literacy Research*, 39(3), pp. 335–356.
- Coupland, N. and Garrett, P. (2010). Linguistic landscapes, discursive frames and metacultural performance: The case of Welsh Patagonia. *International Journal of the Sociology of Language*, 205, pp. 7–36.
- Creese, A. and Blackledge, A. (2010). Translanguaging in the bilingual classroom: A pedagogy for learning and teaching? *The Modern Language Journal*, 94(1), pp. 103–115.
- Dagenais, D., Moore, D., Sabatier, C., Lamarre, P. and Armand, F. (2009). Linguistic landscape and language awareness. In: E. Shohamy and D. Gorter, eds., *Linguistic landscape: Expanding the scenery*, 1st ed. New York: Routledge, pp. 253–269.
- De Mauro, T. (1980). *Guida all'uso delle Parole*. Roma: Editori riuniti.
- Duarte, J. and Gogolin, I. (2013). *Linguistic superdiversity in urban areas*. Hamburg: John Benjamins.
- Extra, G. and Yağmur, K. (2004). *Urban multilingualism in Europe: Immigrant minority languages at home and school*. London: Multilingual Matters.
- Fabian, J. (1983). *Time and the other: How anthropology makes its object*. New York: Columbia University Press.
- García, O. (2009). *Bilingual education in the 21st century: A global perspective*. West Sussex: Wiley- Blackwell.
- Goffman, E. (1963). *Behavior in public places*. New York: Free Press.
- Goffman, E. (1981). *Forms of talk*. Philadelphia: University of Pennsylvania Press.
- Gorter, D., ed. (2006). *Linguistic landscape: A new approach to multilingualism*. Clevedon, Buffalo and Toronto: Multilingual Matters.
- Gorter, D. and Cenoz, J. (2014). The linguistic landscapes inside multilingual schools. In: B. Spolsky, O. Inbar-Lourie and M. Tannenbaum, eds., *Challenges for language education and policy: Making space for people*, 1st ed. New York: Routledge, pp. 151–169.
- Gorter, D. and Cenoz, J. (2017). Linguistic landscape and multilingualism. In: J. Cenoz, D. Gorter and S. May, eds., *Language awareness and multilingualism*, 1st ed. Cham: Springer International Publishing, pp. 233–245.
- Gorter, D., Marten, H. and Van Mensel, L., eds. (2012). *Minority languages in the linguistic landscape*. New York: Palgrave Macmillan.
- Griffin, J. (2004). The presence of written English on the streets of Rome. *English Today*, 20(2), pp. 3–8.
- Hancock, A. (2012). Capturing the linguistic landscape of Edinburgh: A pedagogical tool to investigate student teachers' understanding of cultural and linguistic diversity. In: C. Hélot, M. Barni, R. Janssens, and C. Bagna, eds., *Linguistic landscapes, multilingualism and social change*, 1st ed. Frankfurt: Peter Lang, pp. 249–266.
- Hélot, C., Barni, M., Janssens, R. and Bagna, C., eds. (2012). *Linguistic landscapes, multilingualism and social change*. Frankfurt: Peter Lang.
- Huebner, T. (2006). Bangkok's linguistic landscapes: Environmental print, codemixing and language change. *International Journal of Multilingualism*, 3(1), pp. 31–51.
- Itagi, N. and Singh, S., eds. (2002). *Linguistic landscaping in India: With particular reference to the new states: Proceedings of a seminar*. Central Institute of Indian Languages and Mahatma Gandhi International Hindi University.
- Jaworski, A. and Thurlow, C., eds. (2010). *Semiotic landscapes: Language, image, space*. London: Bloomsbury Publishing.
- Kallen, J. (2009). Tourism and representation in the Irish linguistic landscape. In: E. Shohamy and D. Gorter, eds., *Linguistic landscape: Expanding the scenery*, 1st ed. New York: Routledge, pp. 310–324.

- Landry, R. and Bourhis, R. (1997). Linguistic landscape and ethnolinguistic vitality: An empirical study. *Journal of Language and Social Psychology*, 16(1), pp. 23–49.
- Leeman, J. and Modan, G. (2009). Commodified language in Chinatown: A contextualized approach to linguistic landscape. *Journal of Sociolinguistics*, 13(3), pp. 332–362.
- Lingscape. [online] Available at: <https://lingscape.uni.lu/>.
- LinguaSnapp. [online] Available at: www.linguasnapp.manchester.ac.uk/.
- Lou, J.J. (2016). *The linguistic landscape of Chinatown: A sociolinguistic ethnography*. Bristol: Multilingual Matters.
- Malinowski, D. (2009). Authorship in the linguistic landscape: A multimodal-performative view. In: E. Shohamy and D. Gorter, eds., *Linguistic landscape: Expanding the scenery*, 1st ed. New York: Routledge, pp. 107–125.
- Malinowski, D. (2015). Opening spaces of learning in the linguistic landscape. *Linguistic Landscape*, 1(1), pp. 95–113.
- Matras, Y. (2009). *Language contact*. Cambridge: Cambridge University Press.
- Matras, Y., Gaiser, L. and Reershemius, G. (2018). Multilingual repertoire management and illocutionary functions in Yiddish signage in Manchester. *Journal of Pragmatics*, 135, pp. 53–70.
- Monnier, D. (1989). *Langue d'accueil et langue de service dans les commerces à Montréal*. Quebec: Conseil de langue française.
- Papen, U. (2012). Commercial discourses, gentrification and citizen's protest: The linguistic landscape of Prenzlauer Berg, Berlin. *Journal of Sociolinguistics*, 16(1), pp. 56–80.
- Pappenhagen, R., Scavaglieri, C. and Redder, A. (2016). Expanding the linguistic landscape scenery? Action theory and 'Linguistic Soundscaping.' In: R. Blackwood, E. Lanza and H. Woldemariam, eds., *Negotiating and contesting identities in linguistic landscapes*, 1st ed. London: Bloomsbury Publishing, pp. 147–162.
- Pavlenko, A. (2010). Linguistic landscape of Kyiv, Ukraine: A diachronic study. In: E. Shohamy, E. Ben Rafael and M. Barni, eds., *Linguistic landscape in the city*, 1st ed. Bristol: Multilingual Matters, pp. 133–150.
- Peck, A. and Stroud, C. (2015). Skinscapes. *Linguistic Landscape*, 1(1), pp. 133–151.
- Pennycook, A. (2009). Linguistic landscape and the transgressive semiotics of graffiti. In: E. Shohamy and D. Gorter, eds., *Linguistic landscape: Expanding the scenery*, 1st ed. New York: Routledge, pp. 302–312.
- Piller, I. (2003). Advertising as a site of language contact. *Annual Review of Applied Linguistics*, 23, pp. 170–183.
- Puzey, G. (2012). Two-way traffic: How linguistic landscape reflect and influence the politics of language. In: D. Gorter, H. Marten and L. Van Mensel, eds., *Minority languages in the linguistic landscape*, 1st ed. New York: Palgrave Macmillan, pp. 127–147.
- Reh, M. (2004). Multilingual writing: A reader-oriented typology-with examples from Lira Municipality (Uganda). *International Journal of the Sociology of Language*, 170, pp. 1–41.
- Rosenbaum, Y., Nadel, E., Cooper, R. and Fishman, J. (1977). English on Keren Kayemet street. In: J. Fishman, R. Cooper and A. Conrad, eds., *The spread of English: The sociology of English as an additional language*, 1st ed. Rowley: Newbury House, pp. 179–196.
- Rowland, L. (2013). The pedagogical benefits of a linguistic landscape project in Japan. *International Journal of Bilingual Education and Bilingualism*, 16(4), pp. 494–505.
- Sayer, P. (2009). Using the linguistic landscape as a pedagogical resource. *ELT Journal*, 64(2), pp. 143–154.
- Scollon, R. and Scollon, S.W. (2003). *Discourses in place: Language in the material world*. London: Routledge.
- Scollon, R. and Scollon, S.W. (2004). *Nexus analysis: Discourse and the emerging internet*. London: Routledge.
- Semi, G. (2015). *Gentrification. Tutte le città come Disneyland?* Bologna: Il mulino.
- Shohamy, E., Ben Rafael, E. and Barni, M., eds. (2010). *Linguistic landscape in the city*. Bristol: Multilingual Matters.

- Shohamy, E. and Gorter, D., eds. (2009). *Linguistic landscape: Expanding the scenery*. New York: Routledge.
- Shohamy, E. and Waksman, S. (2012). Talking back to the Tel Aviv centennial: LL responses to top-down agendas. In: C. Hélot, M. Barni, R. Janssens and C. Bagna, eds., *Linguistic landscapes, multilingualism and social change*, 1st ed. Frankfurt: Peter Lang, pp. 109–125.
- Spolsky, B. (2009). Prolegomena to a sociolinguistic theory of public signage. In: E. Shohamy and D. Gorter, eds., *Linguistic landscape: Expanding the scenery*. New York: Routledge, pp. 25–39.
- Spolsky, B. and Cooper, R. (1991). *The languages of Jerusalem*. Oxford: Clarendon Press.
- Szabó, T. (2015). The management of diversity in schools: An analysis of Hungarian practices. *Apples: Journal of Applied Language Studies*, 9(1), pp. 23–51.
- Thomason, S. (2001). *Language contact*. Edinburgh: Edinburgh University Press.
- Trinch, S. and Snajdr, E. (2016). What the signs say: Gentrification and the disappearance of capitalism without distinction in Brooklyn. *Journal of Sociolinguistics*, 21(1), pp. 64–89.
- Tulp, S. (1978). Reklame en tweetaligheid: Een onderzoek naar de geografische verspreiding van franstalige en nederlandstalige affiches in Brussel. *Taal en Sociale Integratie*, 1, pp. 261–288.
- Vertovec, S. (2007). *New complexities of cohesion in Britain: Super-Diversity, transnationalism and civil-integration*. London: Commission on Integration and Cohesion.
- Weinreich, U. (1953). *Languages in contact*. The Hague: Mouton.
- Winford, D. (2003). *An introduction to contact linguistics*. Oxford: Blackwell.
- Woldemariam, H. and Lanza, E. (2014). Language contact, agency and power in the linguistic landscape of two regional capitals of Ethiopia. *International Journal of the Sociology of Language*, 228, pp. 79–103.
- Zotero. (2013). *Linguistic landscape bibliography*. [online] Available at: www.zotero.org/groups/216092/linguistic_landscape_bibliography

Urban youth speech styles in multilingual settings

Margreet Dorleijn, Maarten Kossmann, and Jacomine Nortier

1. Introduction and definitions

1.1 *General introduction and outline*

The phenomenon of urban youth language has been enjoying growing academic interest since the late 1990s. In this chapter, we will address issues of language contact related to urban youth languages, such as translanguaging and stylization. The way young people use highly creative ways to express their identities and demarcate group boundaries in urban societies in a globalizing world has attracted scholars from both linguistic, anthropological and sociological disciplines. Although migration, globalization and language contact have always existed, the extent to which these processes have developed over the past decades undoubtedly is the basis of this relatively recent academic interest.

As it stands today, around 50% of the world's population live in cities and by 2050, the number is expected to increase to 70% (United Nations, 2018). These numbers include relatively high proportions of children, teenagers and adolescents. Immense changes have taken place over the past few decades and in their search for a place in society, today's young people are confronted with ethnic, linguistic, and rapidly spreading cultural varieties and mixes.

'Urban youth language' is a rather elusive phenomenon, which can manifest itself in different, and often very volatile, ways. This undoubtedly has contributed to the liveliness of the debates about preferred methodology. In this introduction, we will attempt to define and demarcate the phenomenon addressed in this chapter.

Before we do so, we first want to address the question why the multilingual language use of young city dwellers is particularly interesting.

There is no professional or lay dispute about the fact that young people, more than older people, are involved in a dynamic process of discovering themselves and the world. They experiment with different identities, and language is an important tool in their identity work. It is, at the same time, the means by which that identity is formed and negotiated in interaction with others (Eckert, 2008; Eckert and Wenger, 2005; Gal and Irvine, 1995; Irvine, 2001; Verschueren, 2004). As Bucholtz and Hall (2005) point out, this process takes place in

monolingual environments as well, but they argue that the dynamics can be observed more clearly in cases where multilingual resources are being used.

Over the past 50 years, urban environments have proven to be particularly favourable for the development of youth languages that employ multilingual repertoires. The reasons why non-Western European people moved to this part of the world are complex. Among them, decolonization and the need for cheap labour force played an important role, eventually leading to a high degree of urban multilingualism. In other parts of the world, rapid urbanization, bringing together people from many different linguistic backgrounds, has also provided fertile ground for the development of multilingual styles specific to young people.

1.2 Definitions

The title of this chapter is Urban Youth Speech Styles (henceforth UYSS). In Dorleijn, Mous and Nortier (2015), this term was proposed because it seems to cover the phenomenon that most researchers in the field are interested in: styles of speech used by young people in urban areas. While youth styles are also found in monolingual communities, we will restrict our discussion in this *Handbook* to UYSS using multilingual inputs.

We will use the term UYSS in the following way: *linguistic practices involving the use of linguistic material from different languages by young people in a multilingual urban environment, with a performative character, and which the speaker can control.*

The definition that UYSS are *linguistically distinctive speech styles mostly associated with urban youth* entails three dimensions. First, only speech styles (or registers, Mesthrie et al., 2009) are concerned. This means that only ways of speaking are included that, for the majority of their users, are (expressive) stylistic choices that are *deliberately chosen* in order to convey a certain image of the speaker and/or of the way her/his message is meant to be understood socially. Second, we delimit the term to such speech styles that are largely associated with younger members of society living in cities. Finally, we delimit our term to styles that show linguistic deviations from the common norm, parts of which are conventionalized and enregistered (Agha, 2005), that is, they are relatively stable features users can choose to employ. This delimitation thus distinguishes UYSS from the broader category of urban vernaculars, which are not specifically youthful, and do not always have the expressive associations of UYSS. It also distinguishes it from the many expressive registers that are not necessarily associated with youth, such as English *slang* or French *argot*. The linguistic part of the delimitation leaves aside highly individual stylistic choices or play with different (non-UYSS) registers that do not seem to have any conventionalization outside of some very small groups of peers.

Furthermore, by using the preceding definition, we exclude full-fledged languages that were acquired as L1s. The UYSS are used ‘on top of’ one or more functionally unrestricted languages as an extra mode or style (Nortier, 2018b).

The delimitation as given here does by no means imply that UYSS constitute a homogeneous group that can easily be contrasted with other ways of speaking. As will be argued later, UYSS show a great diversity, both as regards their linguistic strategies and their social meaning. Our use of the term UYSS here should not be interpreted as establishing a distinct category, but in the sense of a pre-established basis for comparison, allowing one to answer questions like: what are the differences and similarities between ways of speaking occurring in contexts X (urban youth) and having the basic characteristics Y (partly conventionalized performative behaviour)? Being not only an analytical or descriptive category, UYSS is used here as a ‘comparative concept’ in the sense of Haspelmath (2010).

Also, when considered as a comparative concept, our definition is not without problems. For example, one may wonder where ‘being young’ starts and ends. Does language use of pre-school children belong to the category of UYSS? Are people in their thirties still young? Is the language use of adults who still use UYSS (maybe as a kind of youth reminiscence) included in the definition? (cf. Dorleijn, Mous and Nortier, 2015 for Sheng; Rampton, 2015). Is it viable to exclude language use that is beyond productive control of the speaker?

There are many terms in use that try to capture UYSS (see for example Cornips and Nortier, 2008). Widely used terms include *multi-ethnolect* (Clyne, 2000, among others), *youth language* (Ziegler, 2018), *urban dialect* (Wiese, 2012), *youth interaction* (for example Djenar, Ewing and Manns, 2018), and *contemporary urban vernacular* (Rampton, 2015).

The choice of these labels often reflects differences in approach and methodology, and researchers tend to have strong opinions about which labels not to use. Thus, *multi-ethnolect* is rejected by researchers who argue that a ‘lect’ term implies a (more or less focused) set of linguistic features that separates it from a standard language or other ‘lects’ (cf. Aarsæther et al., 2015). The general term *youth languages* is very broad, and could in principle also include monolingual performative and expressive language play as for example exercised in student associations (e.g., Kiesling, 2001). Equally, it might include the multilingual language use of young people for whom this multilingual use is a way to communicate effortlessly with peers with the same linguistic background, and not at all as a stylistic device to be used for group demarcation or identity construction (Anchimbe, 2015, among others). The term *youth interaction* is even broader and refers to much more than just linguistic aspects. Terms like *urban dialects*, *ethnolects*, *multi-ethnolects* and *contemporary urban vernacular* suggest fixed, conventionalized varieties, something that does not do justice to the dynamics of these ways of speaking. According to Eckert (2008, p. 26), ‘The term ethnolect (like sociolect and the more generic dialect) reflects a view of language as a fixed rather than fluid entity, and of identity as compartmentalized, allowing one to think of an ethnolect as a discrete system indexical of ethnicity alone.’

Specific UYSS come under a broad range of labels, names and instances of which the following are examples from within Europe: *Perkerdansk* (Denmark), *Kebabnorsk* (Norway), *Rinkebysvenska* (Sweden), *Kiezdeutsch* or *Kanakspråk* (Germany), *Straattaal* (Netherlands), *Citétaal* (Belgium), *Verlan* (France, as one of the forms of ‘langue des jeunes’) (see also Nortier and Dorleijn, 2013). Examples from outside Europe are *Iscamtho* (South Africa), *Sheng* and *Engsh* (Kenya), *Camfranglais* (Cameroon), *Engligbo* (Nigeria), *Hong Kong Slanguage*, and *Walikan*, *Prokem* and *Gaul* in Indonesia. Sometimes the labels have pejorative associations, as in the case of German *Kanakspråk*: ‘The word Kanake is a highly derogatory term that has been used since roughly the 1970s to refer to visibly non-German foreigners or presumed foreigners, especially Turks’ (Loentz, 2006, p. 33). Sometimes the labels are coined by outsiders and sometimes by speakers themselves. An example of the latter is Dutch *Straattaal* (‘Street Language’), which was adopted by René Appel (1999) from users of the style. Moreover, whatever the exact history of such names, African UYSS are often referred to by their users under such terms, such as *Nouchi* in Abidjan and *Sheng* in Nairobi (Kouadio, 2006; Newell, 2009; Mazrui, 1995; Abdulaziz and Osinde, 1997).

The comparative concept UYSS implies one important essentialization: specific ways of speaking are presented as comparable, that is, as a coherent set of features, which can thus be contrasted with other ways of speaking within the community. Some scholars (e.g., Jaspers, 2008; Cornips, Jaspers and de Rooij, 2015) strongly reject this idea. This line of thought questions whether it makes sense to consider a specific UYSS an entity at all. Shouldn’t we consider it just a set of more or less independent variables, each variable being put into action for

different reasons, producing different kinds of indexicalities? Is there any reason to consider these variables as more than an arbitrary set put together by researchers? Aren't we creating categories that have no root in reality by naming such scientific constructs?

A full deconstruction of the notion of UYSS along these lines seems to miss the point, however. The feeling that there is something 'new' and 'other' going on in this type of language use is shared both by outside observers and by its users (see, e.g., Wiese, 2011b); moreover, as Madsen remarks for Copenhagen UYSS, 'certain linguistic features [. . .] are perceived as belonging together and represent a certain way of speaking signalling certain social meaning' (Madsen, 2011, p. 275). This is also clear from the remarkable success of naming strategies, not only with scientists and outsiders, but also by users themselves.

2. Historical overview of the study of UYSS

In the study of UYSS a dichotomy can be observed between studies that focus on the norms and conventions that exist within a certain style/variety on the one hand, and studies that concentrate on the way in which social actions lead to – or are accompanied by – specific language practices on the other.

The first type of studies has an overlap with 'classic' ethnolect studies such as Wölck (2002). Examples are Wiese's studies (2011a, 2011b) which describe the German UYSS *Kiezdeutsch* as a German dialect in emergence.

For the second type of study, with a focus on UYSS as expressive, performative speaking styles in social action, LePage and Tabouret-Keller (1985) and Rampton's pioneering study *Crossing* (1995) count as important milestones. LePage and Tabouret-Keller (1985) introduced the notion of 'act of identity,' which turned out to be a powerful tool to describe the relation between language and identity. They claim that the language spoken by somebody and his or her identity as a speaker of this language are inseparable. Language acts are acts of identity. Rampton's seminal work (1995) used an anthropological, linguistic and ethnographic approach, strongly grounded in Conversation Analysis, which inspired many later studies. As is often the case when a new research field opens up, there are, of course, precursors. An example of this is Hewitt (1982, 1986). Hewitt examined the sociolinguistic impact of the 'London Jamaican' creole used by young black Londoners on the language and culture of young Whites. His study was carried out among working-class adolescents in two areas in South London where young white residents appropriated characteristics of young Jamaicans. He placed the use of black language forms in the speech of Whites firmly in its social and political setting.

In Scandinavia, Kotsinas was among the first linguists to note a changing way of speaking among young people in Rinkeby, a multi-ethnic suburb near Stockholm in Sweden (Kotsinas, 1988, 1992). The variety was also observed in other suburban areas and its users were young people with a migrant background as well as native Swedes.

Before the anthropological approach gained currency among sociolinguists studying UYSS, there were important variationist studies on language use by specific groups with a migrant background that were then mainly referred to as 'ethnolects.' Well-known examples are Horvath (1985) and Wölck (2002), who mainly focus on the correlation of certain phonological features with groups with a minority background. In accordance with the variationist framework, they do not elaborate explicitly on aspects such as agentivity, performativity and expressiveness.

Some recent studies also choose an explicitly variationist approach where respondents are placed in an experimental situation. The objective of studies of this type is to describe

linguistic features which may be inherited during the acquisition process or are at least conventionalized. Such studies are important contributions when it comes to the – in our view central – question of what is contributed to UYSS by acquisition and socialization processes and what by agentive action (van Meel, 2016; Hinskens, 2011).

Other studies combine the quantitative description of linguistic aspects with a qualitative description of social action, e.g., Archakis and Papazachariou (2009).

3. Critical issues and topics

Two main strategies can be distinguished which set the UYSS apart from the languages that form their basis. The first will be called here ‘L2 stylization’ (cf. Muysken, 2013), the second will be called ‘paralexical insertion.’

3.1 *Linguistic strategies: L2 stylization*

L2 stylization refers to the emulation of the (perceived) way non-native speakers speak the dominant language. This deliberate use of L2 learner varieties can be considered stylizations (in the sense of Coupland, 2007), meaning that UYSS speakers are very well able to use more standard ways of speaking the dominant language (as opposed to real L2 interference), and that they only choose a restricted set of the many features associated to L2 learner varieties.

Three types of L2 stylization are common in UYSS:

- 1 Stylizations of non-native accents;
- 2 Stylizations of non-native syntax;
- 3 Introduction of a restricted group of function words common in code-switching, such as utterance modifiers (Matras, 1998) and expressive interjections.

Stylizations of non-native accents come in two types. The first type is the substitution of certain phonetic features of the dominant language that prove to be difficult for most L2 speakers. One may argue that the substitution of German [ç] by [ʃ] or [e] in UYSS variants of German belongs to this type (Auer, 2013; Wiese, 2012). Such substitutions may be considered to be largely independent from the type of L2 that is imitated – similar problems may be expected for L2 speakers with different L-1s, and thus one may argue that there is no specific L2 accent that is targeted here. Simplification of this type is far from general in UYSS, and one should beware of presenting UYSS phonetics as simplified versions of the non-UYSS variants. Thus, *Moroccan Flavoured Dutch* (cf. Nortier and Dorleijn, 2008), while presenting quite different phonetics, maintains the full set of standard Dutch vowel distinctions (Mourigh, forthcoming); this is very different from Moroccan L2 accents in Dutch.

The second type of L2 stylization is inspired by the L2 accents of specific L-1 groups. The afore-mentioned *Moroccan Flavoured Dutch* is a case in point. This variety, used both by young people with a Moroccan background and by people with different linguistic histories as well, features Moroccan-like intonation and syllabification, as well as some other features that are typical for older Moroccan L2 speakers of Dutch. For example, it presents highly complex consonant clusters (at least on the phonetic level), such as [ksɔxtɕfɛm] instead of autochthonous Dutch [ksɛxtɛxəfɛm] *ik zeg tegen hem* ‘I say to him.’ This can easily be understood from Moroccan Arabic and Berber, which both allow for this kind of clusters phonetically, but is very different from, for example, Turkish, another important migrant language in the Netherlands.

L2 accent stylization does not just mean talking with a particular (stereotyped) L2 accent. *Moroccan Flavoured Dutch* presents us with a clear example of how such stylizations can be inspired by, and still be different from, L2 pronunciations and stereotypes. In many older Moroccan's L2 Dutch, /s/ is pronounced [ʃ] in all positions of the word, a pronunciation which also belongs to the stereotype of the accent. In *Moroccan Flavoured Dutch*, this pronunciation is found as a stylistic variable, but here, it only takes place under the specific phonetic condition of clusters with sibilants followed by a continuant. Thus, /slapə/ (*slapen*) 'to sleep' can be pronounced [ʃla:pə], but /ste:ts/ (*steeds*) 'continually' is consistently pronounced [ste:ts] (Mourigh, 2017a). This shows that the stylization is not just an on-the-spot imitation of a stereotype, but that it is a convention that is not necessarily identical to the stereotype it is based on.

In addition to phonetics, L2 stylization also concerns morphosyntax and syntax. In UYSS based on continental Germanic languages, two features of stylization are widespread. The first feature is the neutralization of grammatical gender. Continental Germanic languages distinguish two or three genders with nouns, whose semantic distribution is highly erratic. Gender is mainly expressed on nominal modifiers. Confusion between grammatical genders is common among L2 speakers, and many of them choose to neutralize genders in favour of the most frequent form. This is something also commonly found in UYSS in this part of the world (Cheshire, Nortier and Adger, 2015).

In the same languages, L2 speakers often use word order deviant from the norm. In continental Germanic, there exists a rule that, in main clauses, the verb occupies the second position in the clause, irrespective of the nature of the first element in the clause. As this first element is often the subject of the clause, it is not unexpected that L2 speakers simplify the system by using strict SV(O) word order. This means that while the dominant variety would have variation between S – V and Adverb – V – S, L2 speakers use Adverb – S – V. Among UYSS, such constructions are well attested in Scandinavia and Germany (Freywald et al., 2015). The following is an example from Berlin:

1. *ab* *jetzt* *ich* *krieg* *immer* *zwanzig* *euro* (Freywald et al., 2015, p. 89)
 from now I get always twenty euros
 instead of standard German:
2. *ab* *jetzt* *krieg* *ich* *immer* *zwanzig* *euro*
 from now get I always twenty euros
 'From now on, I get always twenty euros.'

In addition to phonetic and (morpho)-syntactic features, UYSS that rely strongly on L2 stylizations also feature the use of a small set of borrowed function words. Most of these are utterance modifiers or interjections, such as *wəllah/vallah* 'I swear!; assuredly!' in a large number of European UYSS and *lan* 'post-sentence tag, man' in UYSS with an important Turkish component. However, in some UYSS of this type insertion of function words that are internal to the clause also occurs (Kossmann, 2017a), cf. the following example from *Moroccan Flavoured Dutch* in Gouda, where the Tarifyt Berber indefinite article *ižžən* is used in an otherwise Dutch sentence:

3. *ja*, *precies*, *als* *je* *bij* *ižžən* *weg* *komt*
 yes precisely if you at a road come
 'Yes, exactly, if you arrive there at a road.' (Mourigh, forthcoming)

It is not always easy to decide whether, in a certain situation, an L2 feature as used by a UYSS user is really a conscious choice (as implied by stylization). Many young people with a migratory background grow up with non-native L2 learner variants of the dominant language in their households, and some L2 features surfacing in their speech may not be conscious choices, but rather the effect of linguistic uncertainty about specific features of the dominant language. Thus, on the one hand some users of *Moroccan Flavoured Dutch* report gender neutralization as a conscious stylistic choice, using, for instance, the article *de* with nouns that have *het* in the dominant language (Nortier and Dorleijn, 2008). On the other hand, in written internet discourse aiming at the dominant variety of Dutch, some Moroccan heritage young people show large-scale hypercorrection, using, for instance, the non-UYSS article *het* with nouns that require *de* (e.g., *het jongen* instead of *de jongen* ‘the boy’) (cf. also Cornips, 2008).

In a number of UYSS the target of emulation does not seem to be an L2 stereotype, but rather a regional variant of the dominant language. This is the case in *Multicultural London English*, the phonetics of which are inspired by the Jamaican pronunciation of English (Cheshire et al., 2011). Here the dynamics seem to be the same as for continental European UYSS with the difference that the main target of emulation was a group that already spoke English before they migrated to Europe.

3.2 Linguistic strategies: paralexical insertion

The second main strategy in the linguistic constituency of UYSS will be called *paralexical insertion*. The term *paralexicon* has been borrowed from *argot* studies (Dubois et al., 2012; cf. also Mous, 2003). It refers to a large but restricted set of style-specific lexical items that may be used as substitutes for normal lexical items in order to attain a certain stylistic goal. This process mainly targets the lexicon. Such insertions take place into a matrix of more neutral speech. The matrix provides the grammatical grid for the utterance and most of the lexicon, while the style-specific lexical elements are inserted into lexical positions within this grid (Mesthrie and Hurst, 2013). In general, these insertions express concepts that could also have been expressed (with different expressivity) in the matrix language. The matrix of a UYSS can itself be a way of speaking involving a large amount of code-switching between several languages (Mesthrie and Hurst, 2013).

While many concepts have two instantiations – both one or several matrix forms and the paralexical form – many other concepts can only be expressed by means of the matrix language(s). The paralexical set may consist of words from foreign languages, phonologically manipulated elements from the matrix language, and matrix or foreign words with semantic changes. A European example of such a paralexical set in UYSS is Dutch *Straattaal*, which is characterized by a large set of words mainly taken over from Sranan Tongo, the creole language of Suriname, but expanded by words from other migratory languages and by manipulated Dutch and English items. The *Straattaal* paralexical set includes a wide variety of concepts somehow related to street style or otherwise important to adolescent urban life, e.g., *tata* ‘autochthonous Dutch,’ *sma* ‘girl,’ *scoro* ‘school,’ *oso* ‘house,’ *pata* ‘shoe,’ *wagi* ‘car.’

Indonesian provides another example. It has an open pronominal system, which has multiple forms to refer to speaker and addressee. First and second singular pronouns from other languages are easily inserted (Djenar, Ewing and Manns, 2018). Thus, for example, in youth interaction, an Arabic first singular *ana* ‘I’ can be inserted to indicate (or emphasize) the religious (muslim) identity of the speaker. Besides, elements from other Indonesian languages are inserted through ‘crossing’ to play around with (ethnic or other) identities.

Paralexical sets do not only consist of loanwords. They also include – and in some cases predominantly consist of – words in the matrix language that have been manipulated formally (e.g., by truncation or by suffixation, Kießling and Mous, 2004) or semantically. In addition to this, one often finds words whose etymology is fully opaque. A number of examples from Juba UYSS (which has a Juba Arabic matrix) are given in Table 19.1 (Nakao, 2013).

The paralexicon is not always bound to one single matrix language. In South Africa, the UYSS that are called *Tsotsitaals* in the scientific literature consist of a single (though highly variable) set of potential insertions, which can be combined with different matrix languages (Hurst, 2015). Paralexical sets are thus potentially independent from the matrix language.

Mesthrie and Hurst (esp. 2013), discussing the South-African situation, make a strict separation between, on the one hand, urban varieties of the matrix language (including ways of speaking involving a lot of code-switching), and on the other hand the paralexicon. This means that the matrix can be quite different from standard varieties but should still not be considered as specific to the UYSS. It also means, according to Mesthrie and Hurst, that the specificity of paralexical UYSS is purely lexical in nature; all unexpected syntactic features would be part of the matrix. While this seems to make sense in the case of most African UYSS, it is problematic for Dutch *Straattaal*, where a paralexical set is combined with a small number of morphosyntactic features, esp. gender neutralization, which do not seem to be common outside UYSS contexts.

It should also be stressed that there is no obligation to use the full set of words belonging to the paralexicon when choosing to use an UYSS. That is, just a few insertions may suffice to reach the intended style. There are no doubt differences between speech communities and speech situations concerning how many (and which) insertions are needed in order for speech to be recognized as UYSS (Dorleijn, Mous and Nortier, 2015).

UYSS with paralexicon are quite similar to the traditional European anti-languages called *argot* – in fact, the same definitions given for *argot* in the *Dictionnaire de linguistique et des sciences du langage* (Dubois et al., 2012, pp. 48–49) would apply to South African *Tsotsitaals*. The distinction between UYSS and *argot* may be fluid or irrelevant in this kind of speech styles.

The preceding rough division into the strategies L2 stylization and paralexical insertion should not be taken as a strict categorization of different UYSS. Many UYSS combine features from one and the other process; thus, Dutch *Straattaal* combines a paralexicon with L2 influences like gender neutralization. It is, however, interesting to compare to what extent one or the other process characterize the UYSS more or less. L2 stylizations seem to be extremely common in European UYSS, while only few of them use paralexical insertion as a major strategy. In fact, the Dutch situation is interesting, as two variants coexist, *Straattaal*, which is largely characterized by paralexification, and *Moroccan Flavoured Dutch*, which is a typical instance of L2 stylization. While the two can be used independently, they can also be combined. This is, in fact, quite easy, as *Straattaal* mainly comprises content words, while *Moroccan Flavoured Dutch* is rather about phonetics and about functional elements; put otherwise, *Moroccan Flavoured Dutch* can be used as the matrix to the insertion of *Straattaal* paralexicon

Table 19.1 Examples from Yuba UYSS

| | | |
|-----------------------|--------------|---|
| Loans: | <i>bong</i> | 'clothes' (< Bari) |
| Formal manipulations: | <i>bête</i> | 'house' (< Juba Arabic <i>bêt</i> + UYSS suffix <i>-e</i>) |
| | <i>bonto</i> | 'trousers' (< Juba Arabic <i>bontolôn</i>) |

(Kossmann, 2017b). Elsewhere in the world, L2 stylization seems to be less common than in Europe, and most UYSS are mainly (or only) characterized by paralexical insertion.

3.3 *The social meaning of UYSS*

Earlier, UYSS were delimited in terms of practices as being typical for urban youth. This, of course, leaves an enormous potential for associations and indexicalities attached to UYSS. To some extent, this is amplified by specific associations to ‘urbanity.’ In Europe and in North America, continents with long histories of urbanization, the adjective ‘urban’ is often associated with lower class urban culture, and, especially, with urban countercultures. Thus ‘urban dictionaries’ as can be found on the internet never concern typical city life registers such as bureaucratic language or languaging in university fraternities. On the other hand, in African and South Asian cities, which have very different population dynamics from Europe and North America, ‘urbanity’ has a strong and (for many people) positive association with modernity and upward social mobility (Nortier, 2017).

Most European UYSS are socially grounded in second or later generation migrants who have been using the dominant language in general society from a tender age. While many of these UYSS transcend ethnic boundaries (and have therefore been called multi-ethnolects, cf. Quist, 2008), users often have clear ideas about which ethnicities are the core users of the style. Thus, for Dutch *Straattaal* Surinamese heritage youth are considered the archetypical users, while uses by other migrant youth may be accepted or not, depending on whom one asks. However, the (quite frequent) use of *Straattaal* by people without a migratory background is considered to be unfitting by other *Straattaal* users (Cornips and de Rooij, 2013; Kossmann, 2017b; Mourigh, 2017b). Thus, while young people with all possible different heritage backgrounds practice *Straattaal*, it still has a clear grounding in certain groups as opposed to others. In other European UYSS, however, ethnicity may be (or have become) less important as an indexicality of the UYSS (Madsen, 2011).

Outside Europe, and particularly in sub-Saharan Africa, it seems that many UYSS have their roots in jargons related to marginal groups, criminals, prison populations, street children, etc. As a result, some of them still have a strong gang related flavour, as is the case of the different *Tsotsitaals* in South Africa (cf. Mesthrie and Hurst, 2013). In other cases, the UYSS has taken a different direction. For example, *Nouchi* in Abidjan (Ivory Coast), which is said to have developed among street children, is now commonly used among young people of all kinds of social backgrounds. According to Kube’s study of language use among Abidjan *lycéens* (high school students), over 80% of the interviewees said they knew *Nouchi*; her studies include the most prestigious elite school of the city. About one-third of the interviewees claimed they used *Nouchi* as their main or even exclusive choice when speaking with friends. To these users the indexicalities of *Nouchi* are therefore very different from the anti-elite counter-cultural associations of some other UYSS. In fact, Kube’s research suggests that to most of her interviewees *Nouchi* is just associated with age (Kube, 2005).

The situation of the UYSS in Nairobi is still different. *Sheng*, which has Swahili as its matrix, is associated with youth in general, although some versions of it still imply an element of being streetwise. *Engsh*, on the other hand, which developed as an anti-language to *Sheng*, nowadays ‘belongs’ to educated youth (Dorleijn, Mous and Nortier, 2015). African UYSS do not seem to index ethnic divisions (whether specific groups or more generally ‘migrants’), and there is little reason to assume they have ever done so in the past. European UYSS, on the other hand, have their roots in migrant communities. However, recently, a number of European UYSS have been said to develop into de-ethnicized youth varieties (Auer, 2013; Rampton,

2015), which would, at least at this point, make them more similar to African UYSS. This was also attested by Freywald et al. (2011) who emphasized this point and proved it through an attitude survey.

Even UYSS that have a counter-culture type of association are not always exclusively used to index a streetwise identity. Thus, Kossmann (2017a) argues that *Moroccan Flavoured Dutch* – often portrayed as a streetwise, ‘aggressive’ UYSS (Nortier and Dorleijn, 2008) – is also used by religious, law-abiding girls in order to mark a relaxed, not-so-serious stance towards their discourse. No doubt, similar complicated networks of indexicalities are to be found in other UYSS too.

UYSS are often described as being gendered, and more commonly used among males than among females (e.g., Mesthrie and Hurst, 2013; Nortier, 2017). This seems to be correlated with indexicalities of rebelliousness that are often associated with male behaviour. In UYSS that do not have such indexicalities there is little reason to expect strongly gendered patterning, and in fact, if such patterns are found, they are not very strong. Nortier (2017) reports how male and female young Moroccan-Dutch people comment negatively on Moroccan girls using *Straattaal*. The same participants, however, use *Moroccan Flavoured Dutch* without overt attention being paid to it. Examples from Algeria, Indonesia, and Hong Kong show that both women and men use youth varieties without the overt negative connotations known from Western Europe. Nortier (2017) suggests that this can be explained in the following way: There seems to be a gender restriction for varieties used as anti-languages (Halliday, 1976), while youth varieties that mark socially upward mobility can be used by both males and females.

4. Methodological matters

In order to envisage the multi-layered character of UYSS and the problems attached to its analysis, it is useful to take recourse to Silverstein’s notion of the *Total Linguistic Fact*. This refers to the dynamic ‘mutual interaction of meaningful sign forms contextualized to situations of interested human use, mediated by the fact of cultural ideology’ (Silverstein, 1985, p. 220). For a full understanding of UYSS it is necessary to analyze them not only at the level of linguistic forms, but also on the levels of linguistic practice and ideology and their interaction. According to Silverstein (1985, p. 221) every linguistic category ‘related to our ability to refer and predicate, which, carefully examined, [is] situated at such [a] triple intersection.’ Although early work on UYSS did not ignore practices and ideologies, most attention was paid to the structural linguistic dimension. Quist (2008) explicitly distinguishes a linguistic variety perspective and a level of practices and ideology. These approaches differ methodologically; the variety approach draws on dialectology and Labovian sociolinguistics and explores linguistic features characteristic of a ‘variety’ or ‘lect,’ whilst the practice and ideologies approaches are grounded in the social constructivist paradigm within sociolinguistics and explore the ways social meaning is constructed in situated discourse (Svendsen, 2015; Svendsen and Quist, 2010).

The different layers of the *Total Linguistic Fact* demand different methods. Thus, the level of linguistic form can be studied on the basis of recording of conversations, participant observations and even elicitation (cf. Nassenstein, 2018). Usually, experimental settings are avoided. However, see Freywald et al. (2011), where an attitudes survey in controlled group samples serves as de facto experiment.

The most common methods for studying the level of linguistic practices are in the realm of linguistic ethnography, which typically couples conversation analysis with interviews and

participant observation. In such studies, one remarks a strong reliance on the in-depth analysis of very small groups of users.

The study of UYSS language ideology concentrates on the analysis of metacommentary. Thus, in 2017, a special issue of *Applied Linguistics Review* was dedicated to metalinguistic comments and communities of practice (Dorleijn and Nortier, 2017).

The reliance on spoken conversation as found in the large majority of UYSS studies may be slightly outdated. Nowadays, Computer Mediated Communication (CMC), especially social media, is a major factor in the development of new patterns of communication. Undoubtedly, in order to understand UYSS, written data from the Internet plays a central role. Oral and CMC modalities have their own characteristics and challenges (Dorleijn and Nortier, 2012).

The usefulness of CMC data for an ethnographic approach is undisputed (see e.g., Androutsopoulos, 2008, 2013) who has done pioneering work in this approach). Where the aim of linguistic ethnography is to establish how group boundaries are set through language, which (speech) events are ordinary and which are exceptional, CMC provides an excellent resource (see e.g., Kossmann, 2017a; Nortier, 2017). More debate concerns the extent to which CMC data can be put on a par with (spontaneous) oral data (see Androutsopoulos, 2006; Dorleijn and Verschik, 2016). Clearly, written CMC data are different, in terms of the higher level of awareness implied in writing and should not be taken as faithful renderings of oral speech, among others because disambiguation strategies used in oral interaction (such as intonation, loudness, gaze, facial expressions, gestures) are lacking in (written) CMC. In oral communication, there is less time to reflect or to monitor and correct than in written communication, and it is more spontaneous. CMC users can correct themselves more easily than in oral encounters. On the other hand, in CMC, authors have to rely completely on (written) language for their disambiguation strategies, which arguably may lead to the more explicit use of certain linguistic features to convey a certain identity (whereas in oral interaction cues would be given in a different way, e.g., through physical appearance etc.).

In CMC the participants (authors) can play more easily with identities: for example, a middle-aged Dutch woman can pretend to be a 15-year-old Moroccan-Dutch boy if she adjusts her nickname and chooses the right words, unhindered by her physical properties.

5. Future directions

Most quantitative and qualitative studies of UYSS focus on particular instances that are either geographically, linguistically, or ethnically restricted – or all three at the same time. The next step would be to bring these case studies together in a comparative study. Earlier, we have identified two linguistic strategies (L2 stylization and paralexical insertion) that seem to apply to most if not all UYSS hitherto described. We could call them universal strategies. Can we, if we compare all individual cases of UYSS, find more universal tendencies? Is it possible to design a model which takes into account social and linguistic factors which can explain commonalities and differences? Or are UYSS indeed a loose collection of phenomena that have little in common, or are not different enough from other linguistic practices to warrant a separate category? Questions like the preceding can only be answered by a large-scale meta-study.

Work on UYSS generally takes a sociolinguistic, anthropological linguistic or discourse and interaction-oriented approach (often a combination of these). However, UYSS may also provide important insights to functionalist and cognitive linguists. From these perspectives the question of ‘attractiveness’ of the stylistic features that speakers select can be addressed. Why are particular features selected over others? To what extent do universal mechanisms such

as iconicity and economy play a role? Do these features possess a certain intrinsic/universal semiotic or emotional saliency?

Furthermore, studies that focus on the perception of UYSS through experiments such as Implicit Association Tests (Campbell-Kibler, 2012) or Matched Guise could yield fruitful results, especially when it comes to distinguishing UYSS from related varieties.

Another direction that has not been explored to the full (if at all) is an alignment with a language acquisition approach. While the register/style approach is dominant in the analysis of UYSS (and part of our preceding definition), one should not rule out that for some users the UYSS has become a neutral way of speaking. Thus, in the case of urban UYSS based on colonial languages, the ‘new’ variant may easily develop a wide range of usages. In such communities, children may become confronted mainly with the new variant, and acquire it as their first language. This seems to be the case at least for some speakers of *Nouchi* in Abidjan (Kouadio, 2006; Newell, 2009) and may be much more widespread. For much of the data discussed previously (e.g., about *Sheng* and *Moroccan Flavoured Dutch*), it is unclear to what extent they are the result of conscious manipulation by speakers, or the unconscious effect of language acquisition processes. Is this determined by individual circumstances of speakers (such as aptness, amount of input, etc.) or can generalizations be made?

This brings us to one more direction which could yield fruitful insights: the historical perspective. How do UYSS change? Do dialects of UYSS develop? In all UYSS, there is much variation and a much higher degree of innovation than one usually encounters in ‘mainstream’ languages. It is often stated, both by speakers and by researchers, that UYSS lexicons are subject to change constantly and that discontinuity of expression is the norm. This seems to be exaggerated, though, and in fact a large number of features prove to be persistent over a relatively long time-span (cf. also Hurst, 2015). As an example, we may adduce the use of Sranan (Surinamese) lexicon in Dutch *Straattaal*, which was already observed by René Appel in 1999. Even though many *Straattaal* users back then and nowadays have no knowledge of Sranan itself, the large majority of the Sranan-based lexicon listed in Appel (1999) is still in use 20 years later. Only few real-time analyses of linguistic change within UYSS have been undertaken so far, but one may speculate that innovations tend to focus on certain expressive and highly frequent parts of the lexicon (like evaluative adjectives and adverbs, or expressions for potential sexual partners and the ways to interact with them), and, moreover, that it is especially the most recent introductions that are substituted by innovative forms. This may lead to the observation that things constantly change, while on the other hand a considerable number of not only lexical, but also grammatical and phonological features remain the same. The implication of this continuity is that UYSS are not only defined by their dynamics but also present linguistic substance that can be studied as (linguistically, though not necessarily socially) stable features.

6. Further reading

Androutsopoulos, J. (2015). Networked multilingualism: Some language practices on Facebook and their implications. *International Journal of Bilingualism*, 19(2), pp. 185–205.

This paper proposes the term ‘networked multilingualism’ and presents findings from a case study to explore its implications for the theorizing of multilingualism.

Djenar, D., Ewing, M. and Manns, H. (2018). *Style and intersubjectivity in youth interaction*. Boston and Berlin: De Gruyter.

Djenar et al. (2018) sketch a South-East Asian perspective on (multilingual) linguistic practices of young people.

Nortier, J. and Svendsen, B.A., eds. (2015). *Language youth and identity in the 21st century. Linguistic practices across urban spaces*. Cambridge: Cambridge University Press.

In this book, linguistic practices from some north-West European countries are compared to each other and practices elsewhere.

Mesthrie, R. and Hurst, E. (2013). Slang registers, code-switching and restructured urban varieties in South Africa. An analytic overview of tsotsitaals with special reference to the Cape Town variety. *Journal of Pidgin and Creole Languages*, 28(1), pp. 103–130.

This article provides a rich overview of the features and analytical challenges presented by UYSS in South Africa.

7. Related topics

Processing multilingual data, social factors, borrowing, code-switching, linguistic landscape, and urban multilingualism

Abbreviations

| | |
|------|---------------------------------|
| CMC | Computer Mediated Communication |
| L1 | first language |
| L2 | second language |
| O | object |
| S | subject |
| UYSS | Urban Youth Speech Styles |
| V | verb |

References

- Aarsæther, F., Marzo, S., Nistov, I. and Ceuleers, E. (2015). Indexing locality: Contemporary urban vernaculars in Belgium and Norway. In: J. Nortier and B.A. Svendsen, eds., *Language, youth and identity in the 21st century. Linguistic practices across urban spaces*, 1st ed. Cambridge: Cambridge University Press, pp. 249–270.
- Abdulaziz, M. and Osinde, K. (1997). Sheng and Engsh: Development of mixed codes among the urban youth in Kenya. *International Journal of the Sociology of Language*, 125, pp. 43–63.
- Agha, A. (2005). Voice, footing, enregisterment. *Journal of Linguistic Anthropology*, 15(1), pp. 38–59.
- Anchimbe, E. (2015). Code-switching: Between identity and exclusion. In: G. Stell and K. Yakpo, eds., *Code-switching between structural and sociolinguistic perspectives*, 1st ed. Berlin: Walter de Gruyter, pp. 139–162.
- Androutsopoulos, J. (2006). Introduction: Sociolinguistics and computer-mediated communication. *Journal of Sociolinguistics*, 10(4), pp. 419–438.
- Androutsopoulos, J. (2008). Potentials and limitations of discourse-centred online ethnography. In: J. Androutsopoulos and M. Beißwenger, eds., *Data and methods in computer-mediated discourse analysis*. Special Issue, *Language@Internet*, 5, Article 9.
- Androutsopoulos, J. (2013). Code-switching in computer-mediated communication. In: S.C. Herring, S.D. Stein and T. Virtanen, eds., *Pragmatics of computer-mediated communication*, 1st ed. Berlin, Germany and New York: De Gruyter Mouton, pp. 659–686.
- Androutsopoulos, J. (2015). Networked multilingualism: Some language practices on Facebook and their implications. *International Journal of Bilingualism*, 19(2), pp. 185–205.
- Appel, R. (1999). Straattaal. De mengtaal van jongeren in Amsterdam [Street language. The mixed language of Amsterdam youth]. *Toegepaste Taalwetenschap in Artikelen*, 62, pp. 39–55.

- Archakis, A. and Papazachariou, D. (2009). On young women's prosodic construction of identity. Evidence from Greek conversational narratives. In: A. Stenström and A. Jørgensen, eds., *Youngspeak in a multilingual perspective*, 1st ed. Amsterdam and Philadelphia: John Benjamins, pp. 13–29.
- Auer, P. (2013). Ethnische Marker im Deutschen zwischen Varietät und Stil. In: A. Depperman, ed., *Das Deutsch der Migranten (Institut für Deutsche Sprache, Jahrbuch 2012)*, 1st ed. Berlin and Boston: De Gruyter, pp. 9–40.
- Bucholtz, M. and Hall, K. (2005). Identity and interaction: A sociocultural linguistic approach. *Discourse Studies*, 7(4–5), pp. 585–614.
- Campbell-Kibler, K. (2012). The implicit association test and sociolinguistic meaning. *Lingua*, 122(7), pp. 753–763.
- Cheshire, J., Kerswill, P., Fox, S. and Torgersen, E. (2011). Contact, the feature pool and the speech community: The emergence of multicultural London English. *Journal of Sociolinguistics*, 15(2), pp. 151–196.
- Cheshire, J., Nortier, J. and Adger, D. (2015). Emerging multiethnolects in Europe. *Queen Mary's Occasional Papers Advancing Linguistics*, 33(27).
- Clyne, M. (2000). Lingua franca and ethnolects in Europe and beyond. *Sociolinguistica*, 14, pp. 83–89.
- Cornips, L. (2008). Loosing grammatical gender in Dutch: The result of bilingual acquisition and/or an act of identity? *International Journal of Bilingualism*, 12(1–2), pp. 105–124.
- Cornips, L., Jaspers, J. and de Rooij, V. (2015). The politics of labelling youth vernaculars in the Netherlands and Belgium. In: J. Nortier and B.A. Svendsen, eds., *Language, youth and identity in the 21st century. Linguistic practices across urban spaces*, 1st ed. Cambridge: Cambridge University Press, pp. 45–69.
- Cornips, L. and Nortier, J., eds. (2008). Ethnolects? The emergence of new varieties among adolescents. *Special Issue of International Journal of Bilingualism*, 12(1–2).
- Cornips, L. and de Rooij, V. (2013). Selfing and othering through categories of race, place, and language among minority youths in Rotterdam, The Netherlands. In: P. Siemund, I. Gogolin, M.E. Schulz and J. Davydova, eds., *Multilingualism and language diversity in urban areas. Acquisition, identities, space, education*, 1st ed. Amsterdam and Philadelphia: John Benjamins, pp. 129–164.
- Coupland, N. (2007). *Style: Language variation and identity*. Cambridge: Cambridge University Press.
- Djenar, D., Ewing, M. and Manns, H. (2018). *Style and intersubjectivity in youth interaction*. Boston and Berlin: De Gruyter.
- Dorleijn, M., Mous, M. and Nortier, J. (2015). Urban youth speech styles in Kenya and the Netherlands. In: J. Nortier and B.A. Svendsen, eds., *Language, youth and identity in the 21st century. Linguistic practices across urban spaces*, 1st ed. Cambridge: Cambridge University Press, pp. 271–289.
- Dorleijn, M. and Nortier, J. (2012). Bilingualism and youth languages. In: C. Chapelle, ed., *The encyclopedia of applied linguistics*, 1st ed. Oxford: Wiley-Blackwell, pp. 480–487.
- Dorleijn, M. and Nortier, J., eds. (2017). Metalinguistic discourse on multilingual urban and youth speech styles and linguistic practices. *Special Issue of Applied Linguistics Review*, 10(3).
- Dorleijn, M. and Verschik, A., eds. (2016). Using multilingual written internet data in code-switching and language contact research. *Special Issue of Journal of Language Contact*, 9(1).
- Dubois, J., Giacomo, M., Guespin, L. and Marcellesi, C. (2012). *Le dictionnaire de linguistique et des sciences du langage*. Paris: Larousse.
- Eckert, P. (2008). Variation and the indexical field. *Journal of Sociolinguistics*, 12(4), pp. 453–476.
- Eckert, P. and Wenger, É. (2005). Communities of practice in sociolinguistics. *Journal of Sociolinguistics*, 9, pp. 582–589.
- Freywald, U., Cornips, L., Ganuza, N., Nistov, I. and Opsahl, T. (2015). Beyond verb second – A Matter of novel information-structure effects? Evidence from Norwegian, Swedish, German and Dutch. In: J. Nortier and B.A. Svendsen, eds., *Language, youth and identity in the 21st century. Linguistic practices across urban spaces*, 1st ed. Cambridge: Cambridge University Press, pp. 73–92.
- Freywald, U., Mayr, K., Özçelik, T. and Wiese, H. (2011). Kiezdeutsch as a multiethnolect. In: F. Kern and M. Selting, eds., *Ethnic styles of speaking in European metropolitan areas*, 1st ed. Amsterdam and Philadelphia: John Benjamins, pp. 43–75.

- Gal, S. and Irvine, J. (1995). The boundaries of languages and disciplines: How ideologies construct difference. *Social Research*, 62(4), pp. 967–1002.
- Halliday, M. (1976). Anti-Languages. *American Anthropologist*, 78(3), pp. 570–584.
- Haspelmath, M. (2010). Comparative concepts and descriptive categories in crosslinguistic studies. *Language*, 86(3), pp. 663–687.
- Hewitt, R. (1982). White adolescent creole users and the politics of friendship. *Journal of Multilingual and Multicultural Development*, 3(3), pp. 217–232.
- Hewitt, R. (1986). *White talk Black talk: Inter-racial friendship and communication amongst adolescents*. Cambridge: Cambridge University Press.
- Hinskens, F. (2011). Emerging Moroccan and Turkish varieties of Dutch: Ethnolects or ethnic styles? In: F. Kern and M. Selting, eds., *Ethnic styles of speaking in European metropolitan areas*. (Studies in Language Variation 8), 1st ed. Amsterdam and Philadelphia: John Benjamins, pp. 101–129.
- Horvath, B. (1985). *Variation in Australian English: The sociolects of Sydney*. Cambridge: Cambridge University Press.
- Hurst, E. (2015). Overview of the tsotsitaals of South Africa; their different base languages and common lexical core items. In: N. Nassenstein and A. Hollington, eds., *Youth language practices in Africa and beyond*, 1st ed. Berlin: De Gruyter Mouton, pp. 169–184.
- Irvine, J. (2001). ‘Style’ as distinctiveness: The culture and ideology of linguistic differentiation. In: P. Eckert and J. Rickford, eds., *Style and sociolinguistic variation*, 1st ed. Cambridge: Cambridge University Press, pp. 21–43.
- Jaspers, J. (2008). Problematising ethnolects: Naming linguistic practices in an Antwerp secondary school. *International Journal of Bilingualism*, 12(1–2), pp. 85–103.
- Kiesling, S. (2001). Stances of whiteness and hegemony in fraternity men’s discourse. *Linguistic Anthropology*, 11(1), pp. 101–115.
- Kießling, R. and Mous, M. (2004). Urban youth languages in Africa. *Anthropological Linguistics*, 46(3), pp. 303–341.
- Kossmann, M. (2017a). Key and the use of Moroccan function words in Dutch internet discourse. *Nederlandse Taalkunde*, 22(2), pp. 223–248.
- Kossmann, M. (2017b). Is Dutch Straattaal a mixed multiethnolect? A Moroccan perspective. *Applied Linguistics Review*, 10(3).
- Kotsinas, U.-B. (1988). Rinkebysvenska – en dialekt. In: P. Linell, V. Adelswärd, T. Nilsson and P. Petersson, eds., *Svenskans beskrivning 16*, 1st ed. Linköping: Linköping University Press, pp. 264–278.
- Kotsinas, U.-B. (1992). Immigrant adolescents’ Swedish in multicultural areas. In: C. Palmgren, K. Lövgren and G. Bolin, eds., *Ethnicity in youth culture*, 1st ed. Stockholm: Stockholm University, pp. 43–62.
- Kouadio, J.N. (2006). Le nouchi et les rapports dioula-français. *Le français en Afrique*, 21, pp. 177–191.
- Kube, S. (2005). *Gelebte Frankophonie in der Côte d’Ivoire. Dimensionen des Sprachphänomens Nouchi und die ivorische Sprachsituation aus der Sicht Abidjaner Schüler*. Münster: LIT Verlag.
- LePage, R. and Tabouret-Keller, A. (1985). *Acts of identity*. Cambridge: Cambridge University Press.
- Loentz, E. (2006). Yiddish, Kanak Sprak, Klezmer, and HipHop: Ethnolect, minority culture, multiculturalism, and stereotype in Germany. *Shofar An Interdisciplinary Journal of Jewish Studies*, 25(1), pp. 33–62.
- Madsen, L.M. (2011). Late modern urban youth style in interaction. In: M. Selting and F. Kern, eds., *Ethnic styles of speaking in European metropolitan areas*, 1st ed. Amsterdam and Philadelphia: John Benjamins, pp. 265–290.
- Matras, Y. (1998). Utterance modifiers and universals of grammatical borrowing. *Linguistics*, 36(2), pp. 281–331.
- Mazrui, A. (1995). Slang and code-switching: The case of Sheng in Kenya. *Afrikanistische Arbeitspapiere: Schriftenreihe des Kölner Instituts für Afrikanistik*, 42, pp. 168–179.
- Meel, L. van. (2016). *The roots of ethnolects. A sociophonological study in Amsterdam and Nijmegen*. PhD. Radboud Universiteit Nijmegen.

- Mesthrie, R. and Hurst, E. (2013). Slang registers, code-switching and restructured urban varieties in South Africa. An analytic overview of tsotsitaals with special reference to the Cape Town variety. *Journal of Pidgin and Creole Languages*, 28(1), pp. 103–130.
- Mesthrie, R., Swann, J., Deumert, A. and Leap, W. (2009). *Introducing sociolinguistics*, 2nd ed. Amsterdam and Philadelphia: John Benjamins.
- Mourigh, K. (2017a). Stance-taking through sibilant palatalisation in Gouda Moroccan Dutch. *Nederlandse Taalkunde*, 22(3), pp. 421–446.
- Mourigh, K. (2017b). A Dutch multiethnolect? Metalinguistic commentary from Gouda. *Applied Linguistics Review*, 10(3).
- Mourigh, K. (unpublished manuscript). *Moroccan Dutch in Gouda*.
- Mous, M. (2003). The linguistic properties of lexical manipulation and its relevance for Ma'a and for mixed languages in general. In: P. Bakker and Y. Matras, eds., *The mixed language debate: Theoretical and empirical advances*, 1st ed. Berlin: De Gruyter Mouton, pp. 209–235.
- Muysken, P. (2013). Ethnolects of Dutch. In: F. Hinskens and J. Taeldean, eds., *Language and space: An international handbook of linguistic variation. Volume 3, Dutch*, 1st ed. Berlin: De Gruyter Mouton, pp. 739–760.
- Nakao, S. (2013). The prosody of Juba Arabic: Split prosody, morphophonology, and slang. In: M. Laffkioui, ed., *African Arabic: Approaches to dialectology*, 1st ed. Berlin and Boston: De Gruyter Mouton, pp. 95–120.
- Nassenstein, N. (2018). 'I swear they said this': Kritische Gedanken zu afrikanischen Jugendsprachen und ihren Beschreibungspraktiken. In: A. Hollington, N. Nassenstein and A. Storch, eds., *Critical youth language studies – Rethinking concepts*. Special Issue of *The Mouth*, 3, pp. 29–63.
- Newell, S. (2009). Enregistering modernity, bluffing criminality: How Nouchi speech reinvented (and fractured) the nation. *Journal of Linguistic Anthropology*, 19(2), pp. 157–184.
- Nortier, J. (2017). Online metalinguistic comments and gender roles in multilingual youth speech styles & practices among Moroccan girls and women in the Netherlands. *Applied Linguistics Review*, 10(3).
- Nortier, J. (2018b). Youth languages. In: A. Ziegler, ed., *Jugendsprachen/Youth Languages. Aktuelle Perspektiven internationaler Forschung/Current Perspectives of International Research*, 1st ed. Berlin and Boston: De Gruyter, pp. 3–24.
- Nortier, J. and Dorleijn, M. (2008). A Moroccan accent in Dutch: A socio-cultural style restricted to the Moroccan community? *International Journal of Bilingualism*, 12(1–2), pp. 125–143.
- Nortier, J. and Dorleijn, M. (2013). Multi-ethnolects: Kebabnorsk, Perkerdansk, Verlan, Kanakensprache, Straattaal, etc. In: P. Bakker and Y. Matras, eds., *Contact languages: A comprehensive guide*, 1st ed. Boston: De Gruyter Mouton, pp. 229–271.
- Nortier, J. and Svendsen, B.A., eds. (2015). *Language youth and identity in the 21st century. Linguistic practices across urban spaces*. Cambridge: Cambridge University Press.
- Quist, P. (2008). Sociolinguistic approaches to multiethnolect: Language variety and stylistic practice. *International Journal of Bilingualism*, 12(1–2), pp. 43–61.
- Rampton, B. (1995). *Crossing. Language & ethnicity among adolescents*. Manchester, UK: St. Jerome Publishing.
- Rampton, B. (2015). Contemporary urban vernaculars. In: J. Nortier and B.A. Svendsen, eds., *Language, youth and identity in the 21st century. Linguistic practices across urban spaces*, 1st ed. Cambridge: Cambridge University Press, pp. 24–44.
- Silverstein, M. (1985). Language and the culture of gender. In: E. Mertz and R. Parmentier, eds., *Semiotic mediation*, 1st ed. New York: Academic Press, pp. 219–259.
- Svendsen, B.A. (2015). Introduction. In: J. Nortier and B.A. Svendsen, eds., *Language, youth and identity in the 21st century. Linguistic practices across urban spaces*, 1st ed. Cambridge: Cambridge University Press, pp. 3–23.
- Svendsen, B.A. and Quist, P. (2010). Introduction. In: P. Quist and B.A. Svendsen, eds., *Multilingual urban Scandinavia: New linguistic practices*, 1st ed. Bristol: Multilingual Matters, pp. xiii–xxiii.
- United Nations. (2018). *Department of economic and social affairs: Population* [online] Available at: www.un.org/en/development/desa/population/.

- Verschueren, J. (2004). Notes on the role of metapragmatic awareness in language use. In: A. Jaworski, N. Coupland and D. Galasinski, eds., *Metalanguage: Social and historical perspectives*. Berlin and New York: De Gruyter Mouton, pp. 53–74.
- Wiese, H. (2011a). Grammatical innovation in multiethnic urban Europe: New linguistic practices among adolescents. *Lingua*, 119, pp. 782–806.
- Wiese, H. (2011b). Ein neuer urbaner Dialekt im multiethnischen Raum: Kiezdeutsch. (A new urban dialect in the multiethnic public space: Kiezdeutsch). In: D. Läßle, M. Messling and J. Trabant, eds., *Stadt und Urbanität im 21. Jahrhundert*, 1st ed. Berlin: Kadmos, pp. 146–161.
- Wiese, H. (2012). *Kiezdeutsch: ein neuer Dialekt entsteht*. München: C. H. Beck.
- Wölck, W. (2002). Ethnolects – between bilingualism and urban dialect. In: Li Wei and J. Dewaele, eds., *Opportunities and challenges of bilingualism*, 1st ed. Berlin: De Gruyter Mouton, pp. 157–170.
- Ziegler, A., ed. (2018). *Jugendsprachen/Youth Languages. Aktuelle Perspektiven internationaler Forschung/Current perspectives of international research*. Berlin and Boston: De Gruyter.

Part 4

Linguistic areas



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

The Balkans

Victor A. Friedman

1. Introduction and definitions

The Balkans represent the first region to be identified as a sprachbund (linguistic league) or linguistic area, i.e., a locus of contact-induced change owing to multi-lateral, multi-directional, mutual multilingualism. Note that here, following Friedman and Joseph (2017, forthcoming) I treat the term *sprachbund* as a loanword that has been nativized in English. In this model, multilingualism is shared by speakers of the various languages, it is stable across generations, and it involves varied social groups. While no linguistic situation is unchanging, the combination of the factors mentioned here differentiate the sprachbund from other contact situations such as a diaspora, a colony, or that of endangered indigenous languages. Owing to the complexity of a sprachbund, the directionality of contact-induced change is not always discernible (Thomason and Kaufman, 1988, p. 93), nor is such directionality necessarily relevant, the point being the fact of convergence itself (Ilievski, 1973). This section defines the basic terms of relevance to the study of the Balkans as a linguistic area.

1.1 *The definition of the Balkans as an area*

The term *area* in the sub-heading title can refer to the physical peninsula, the geo-political construct, or the linguistic area, which was Emeneau's (1956) interpretation of Trubetzkoy's (1930) *Sprachbund* 'language union' itself Trubetzkoy's own translation of the Russian *jazykovej sojuz* 'linguistic union' (Trubetzkoy, 1923). The name *Balkan* itself should also be elucidated, as it was not used for the peninsula until the nineteenth century, first as a geographic name in 1808 (Sundhaussen, 1999, p. 31; Todorova, 1997, p. 25). *Balkan* is Turkish for 'forested mountain' and is also the name of the mountain range running east-west across the middle of what is today Bulgaria and known as *Stara Planina* 'old mountain' in Balkan Slavic, *[H]aimos* in Greek, and *Haemus* in Latin. This usage simply took a significant geographical feature and applied it *pars pro toto* to the entire peninsula, which is unproblematically defined on three sides by the Adriatic, Ionian, Mediterranean, Aegean, and Black Seas (including the Dardanelles, Sea of Marmara, and Bosphorus). At the time it was first used as a geographical designation, most of the geographic Balkan peninsula was – or had recently

been – under Ottoman rule and known as ‘Turkey in Europe.’ The first geopolitical use of ‘Balkan’ dates from Minchin’s (1886, p. 92, 400 et passim) reference to ‘Balkan states,’ when Bulgaria, Greece, Montenegro, Romania, and Serbia were all independent or autonomous states (albeit with boundaries differing from today’s), and European Turkey had shrunk to an area that would, after its final dismemberment in the wake of the Balkan Wars and World War I, become Albania and parts of Bulgaria, Greece, and the Kingdom of Serbs, Croats, and Slovenes (Yugoslavia after 1929) as well as a small part of the Republic of Turkey. For most of the twentieth century the geopolitical Balkans was understood to be comprised of Albania, Bulgaria, Greece, Romania, European Turkey, and what is now former Yugoslavia. In the twenty-first century, the term Western Balkans has been used for that part of southeastern Europe that is not in the EU.

As a linguistic area, however, the definition of ‘Balkan’ is more complex. It is basically the part of the Balkan Peninsula where languages from different Indo-European groups plus certain Turkic languages have converged to a significant degree. This entire linguistic area was at one time part of the Ottoman Empire, thus omitting Slovenia (parts of which, however, were called ‘Balkanic Italy’ in the 1920s) as well as the Kajkavian region of Croatia. Some former Ottoman European possessions, e.g., Hungary, are of marginal linguistic relevance, having contributed only the occasional lexical item (Schubert, 1999). Of the languages spoken in the region, about which more will be said in Section 1.2, Slavic, Albanian, Romanian, and Romani all contribute to a heuristic description of the northern boundary of the linguistic Balkans. A sprachbund differs from a nation-state and from the concept of the genealogical language family in that ‘boundary’ is not a legally or ideologically fixed (or reified) descriptor but rather, as Hamp (1989) pointed out, ‘a series of differential bindings.’ The Balkan sprachbund is thus not a case of ‘all or nothing’ (pace Stolz, 2006) but rather a descriptor of a group of convergent languages resulting from centuries of sustained, intense and intimate multi-lateral, multi-directional, mutual multilingualism, described by Friedman and Joseph (2017) as their ‘4-M’ model.

The Balkans as a language area (sprachbund) has a northern boundary that is more complex than that of any political formation. The northern border region extends along the Albanian-Slavic contact zone (pre-1878) to the Romance-South Slavic contact zone (where Romance gradually replaced Slavic as the state language). Certain specifically Balkan features of Romani follow roughly these same boundaries. It must be remembered that these mappings are fuzzy, not the sharp, unbroken lines on some political map, although politics has played its part in the complexity. The implications of these language-contact facts will be discussed in Section 3.1.2.

1.2 The Balkan languages

In this chapter the terms Albanian, Romani, Balkan Romance, Balkan Slavic, Balkan Turkic, and Greek are used as convenient labels to refer to the totality of the respective linguistic systems, each of which is clearly differentiated from the others but within each of which language/dialect distinctions have different complexities. Here only those matters of importance to this description of the Balkans as a language contact area will be briefly elucidated. It should also be noted that for all of these systems, there are linguistic enclaves both inside and outside the Balkans which are beyond the scope of this chapter (see Friedman and Joseph, forthcoming, for details).

Albanian is divided into two major dialects by a bundle of isoglosses running along or just south of the river Shkumbî in the middle of Albania: Geg in the north and Tosk in the

south. The dialects of Montenegro, Kosovo, and most of North Macedonia are Geg. Those of Greece and southwestern North Macedonia are Tosk. The Greek of mainland Greece is also divided into a northern and a southern area, the main isogloss runs along the Gulf of Corinth and north of Attica, but some of the dialects of southern Albania do not have the distinctively northern features (Joseph and Brown, 2015). Mention should be made here of Tsakonian, the sole descendent of Doric Greek, all other surviving Hellenic languages/dialects being descended from the Hellenistic koinē of late antiquity. While of great interest, it is outside the general scope of this chapter. All Romani dialects show some effects of Balkan language contact (cf. Matras, 1994, who refers to Romani as a ‘Balkanised Indo-Aryan language’). Those Romani dialects that remained in the Balkans, however, show more such effects than those outside the Balkans. The dialects of Romani of most relevance for the Balkans are the so-called Balkan group (subdivided into Balkan I and Balkan II) and South Vlax. (It should be noted that these dialects can all be co-territorial, although Balkan II dialects are limited to northern Bulgaria, North Macedonia, and Kosovo, while Balkan I dialects are concentrated mainly in the regions that remained under Ottoman control longer (see Matras, 2005; Boretzky and Iglar, 2004). For Balkan Romance, the dialects north of the Danube constitute Romanian, and those south of the Danube make up Aromanian – spoken in Albania, northern Greece, North Macedonia, and southwestern Bulgaria – and Meglenoromanian, spoken in two villages in southeastern North Macedonia and some adjacent villages across the border in northern Greece. Balkan Slavic is a term referring to Bulgarian, Macedonian, and – for many important features – the so-called Torlak dialects of the former Serbo-Croatian. Balkan Turkic is comprised of two languages: Turkish and Gagauz, both of which are in the Oghuz branch. Kipchak languages such as Kuman and Tatar were or are spoken in the Balkans, but Kuman had disappeared by the end of the middle ages and Tatar did not arrive until the period of the Russo-Turkish wars of the nineteenth century. The Turkish of the Balkans is also called Rumelian Turkish, with East Rumelian Turkish being more or less part of the Turkish standard (Istanbul Turkish is East Rumelian), while West Rumelian Turkish – spoken today mostly in Kosovo and North Macedonia – is heavily Balkanized (Ibrahimi, 1982; Matras, 2003/2004; Friedman, 2006; Matras and Tufan, 2007).

The classic group of Balkan languages (as opposed to languages of the Balkans, which could mean any language that happens to be spoken in the region) is Albanian, Balkan Romance, Balkan Slavic, and Greek (Sandfeld, 1926/1930). To this Romani (Friedman, 2000b; Matras, 1994) and – as noted previously – West Rumelian Turkish should be added. Moreover, Balkan Judezmo also shows features that developed in the Balkans (Friedman and Joseph, 2014) and that distinguish it from Judezmo spoken outside the Balkans. Finally, we can mention Balkan Armenian, i.e., the Armenian spoken in the Balkans for centuries prior to the flood of refugees in the Balkans escaping the genocide taking place in Anatolia. Although this dialect differs from that of the arrivals from Anatolia (Friedman field notes), it has yet to be studied. The information in Adamou (2008) is a tantalizing indicator that there is more to be done (see Section 3.1.1).

2. Historical overview

This section presents main points of the external histories of the Balkan languages. That section is followed by a history of the field of Balkan linguistics, which, in its early years, was the history of contact linguistics.

2.1 *The Balkan languages: external histories*

Although the Balkan Peninsula has been inhabited by humans for many tens of thousands of years, nothing identifiable survives of their languages – aside from the occasional toponym or other lexical item that defies any other explanation – prior to the Indo-European invasions. For Hellenic, at least, there were other Indo-European as well as non-Indo-European speakers when those speakers arrived in the Balkans. Based on the similarities between Greek and Armenian, it is hypothesized that Hellenic reached the Balkans either from north of the Black Sea, or from south of it. While the northerly route is the most popularly accepted, the possibility of a southern route cannot be ruled out at this point (Hamp, 2013, p. 10).

The speakers of the Indo-European dialect that became Albanian migrated from north of the Carpathians, where they had been in contact with the Indo-European dialects that would become Germanic and Slavic (Hamp, 1994, 2010). Without going into the details of the debate (see Fine, 1983, pp. 10–11; Woodard, 2008, pp. 7–9), Albanian is unquestionably descended from one of the ancient Indo-European Balkan languages mentioned in ancient sources. However, the attestations for these languages range from a few ill-understood inscriptions to a few glosses and anthroponyms to a single word or nothing at all. Hamp (2013, pp. 8–9) has, on the basis of shared innovations with Messapic, argued that Albanian is indeed related to Illyrian (as opposed to, e.g., Thracian or Daco-Moesian), but it descended from a sibling of Illyrian that was, at one time, closer to the Danube and in contact with Daco-Moesian (Hamp, 1982). Be that as it may, Albanian and Greek represent the two language systems spoken today that were spoken in the Balkans prior to the arrival of Latin speakers.

The migrations mentioned in the previous two paragraphs were completed before the arrival of the next linguistic contributor: Latin. The Romans began incursions into the Balkans beyond the towns on the Dalmatian Coast around the turn of the BCE/CE millennium. Latin reached its height in the Balkans during the second century CE. After that it was in retreat, and it disappears from inscriptional evidence in the Balkans in the sixth century CE. Romance speakers, however, clearly did not disappear. They are well attested in late medieval sources (Fine, 1987, pp. 11–14) but we do not have a single datable document until 1521, and manuscripts without dates probably are no earlier than 1500 (Mareş, 2000). Mention should be made of the Jireček line, named for the historian Konstatine Jireček, who first proposed a boundary between Greek and Latin as languages of power or prestige based on inscriptional evidence (Jireček, 1911, pp. 38–39). The line runs from the Albanian coast east and northeast across North Macedonia, then east across Bulgaria to the Black Sea. A number of alternatives have been proposed modifying Jireček's start point, end point and details of the route in between (Kaimio, 1979, pp. 86–89). However, regardless of the exact northern and southern limits, it is now generally recognized that there was a region of Greek-Latin bilingualism running across Albania, North Macedonia, and Bulgaria (including the adjacent corner of Serbia). North of this region, Latin dominated. South of the region, Greek generally predominated in the inscriptional evidence, but Romance (and Thracian) continued to be spoken.

Slavic speakers began arriving in the Balkans south of the Danube in significant numbers during the sixth and seventh centuries. The linguistic, toponymic, and documentary evidence shows that Slavic speakers were present everywhere in the Balkans, down the southernmost Peloponnese (Fine, 1983, p. 36, 1987, p. 166). It was during this period that speakers of all the languages that had been spoken on the peninsula other than Greek, Latin/Romance, or the ancestor of Albanian either shifted to one of those three or to Slavic. During the Late Middle Ages, Greek (represented by Byzantium) and South Slavic (represented by Bulgaria and Serbia) vied as the languages of power based on their respective imperial conquests, each always at the expense of

the other. This was the basic linguistic situation when Turkic speakers arrived in the Balkans via Anatolia. Turkic speakers had arrived from north of the Black Sea on more than one occasion, but, as medieval Slavonic sources make clear, their influence was limited to a few lexical items.

The first linguistically significant migration were speakers of what is now known as Gagauz, toward the end of the Byzantine period (Wittek, 1953). Once the Ottoman Turks entered, and eventually conquered, the Balkans during the fourteenth to fifteenth centuries, Turkish became a language of power and prestige, but in some regions it also became the language of the countryside. It is most likely that Romani, which arrived in Anatolia in the Middle Ages, was spoken in Byzantine Europe prior to the arrival of the Ottomans (Matras, 2002, pp. 17–18). The attested dispersal of Romani throughout Europe occurred during the same century as Ottoman expansion in the Balkans, as a result of which, Romani, like Romance and Slavic, has extensive branches both inside and outside the Balkans. As such, it contributes, as do Slavic and Romance, to the understanding of convergences versus inherited or parallel (typological) similarities in the Balkan languages.

Finally, the latest arrival to be considered here was Balkan Judezmo, which is descended from the medieval Judeo-Iberian (Spanish/Portuguese) of the Jews expelled from Spain in 1492 and Portugal in 1497 who found refuge in the Ottoman Empire. Balkan Judezmo deserves consideration owing to five centuries of Judezmo contact with other Balkan languages and the availability of non-Balkan dialects of Judezmo and historical sources relating to the relevant dialects and languages of and from Iberian Peninsula.

2.2 *Balkan linguistics as a field*

In the nineteenth century, much academic effort was being expended in the establishment of genealogical relationships among languages and elucidating resemblances between languages that were due to their descent from a common ancestor. While studying the histories and structures of the Romance and Slavic languages, scholars noticed that Balkan Romance and Balkan Slavic resembled one another in ways that they did not resemble other Romance and Slavic languages, respectively. Moreover, they noticed that some of these similarities were shared with Albanian, and in some instances also with Greek. The earliest known published observation for all four groups is Leake (1814, p. 380), who attributed the similarities to Slavic. However, Kopitar's (1829, p. 86) formulation is the one most frequently cited: he wrote that – in modern terms – Albanian, Balkan Romance, and Balkan Slavic give the impression that *'nur eine Sprachform herrscht, aber mit dreierley Sprachmaterie'* [italics in original] (only one grammar dominates, but with three lexicons). He attributed this to the influence the pre-Latin substratum (which he called 'Thracian'). During this period it was the postposed definite article in Albanian Balkan Romance, and Balkan Slavic as well as the replacement of infinitives by analytic subjunctives and the origin of future marking particles in lexical 'want' that attracted most attention, although the reduction of case marking in Balkan Slavic was also noted.

In the context of nineteenth and early twentieth century ideologies of racial purity and eugenics, languages were taken as defining human populations called 'races' (cf. Hammond, 1923, p. 9). Under this ideology, language change was viewed as degeneration from an earlier state of purity, and language contact was equated with miscegenation. It was thus that Schleicher (1850:143) described Albanian, Balkan Romance, and Balkan Slavic as a group that was united by the fact that each was *'die verdorbensten ihrer Familie'* (the most corrupt in their families).

After Kopitar (1829), the next seminal work was Miklosich (1861), which was concerned mainly with Slavic lexicon in Romanian. Miklosich (1861) also contained the first list of

Balkan morphosyntactic features, which, in addition to the three just mentioned, included genitive-dative merger, object reduplication, and the expression of ‘teens’ as ‘numeral on ten’ (this last not present in Greek). Miklosich also identified a number of phonological features, but that question will be addressed in Section 3.1.10.

The decade 1920–1930 saw the publication of five works – two of them being the same item but in different languages – that taken together established Balkan linguistics as a field of investigation: Trubetzkoy (1923, 1930), Seliščev (1925), and Sandfeld (1926, 1930). Trubetzkoy (1923), in Russian, contained a passage that became, in German, his Proposition 16 at the First International Congress of Linguists in the Hague in 1928 (Trubetzkoy, 1930). This proposition defined the *sprachbund* as opposed to the language family, i.e., it provided the basic conceptual framework for distinguishing similarities owing to contact-induced change versus similarities due to descent from a common ancestor. The basics of that definition are worth citing here and are presented in Table 20.1.

Seliščev (1925) introduced the term ‘Balkanism’ – that is, a convergent feature shared by at least some of the Balkan languages that arguably owes its existence to contact-induced change – and also provided an updated list of features that took into account the decades of research since Miklosich (1861). Seliščev (1925) would have been more widely recognized, had it not been eclipsed by Sandfeld (1930), the French translation of the original Danish (Sandfeld, 1926). This epoch-making compendium of Balkan linguistic features, with roughly half the pages devoted to lexicon or phraseology, with the explicit exclusion of Romani and Judezmo, and in ignorance of West Rumelian Turkish, nonetheless has stood the test of time as a valuable resource. There are two reasons for this. First, in one form or another, Sandfeld described an enormous number of Balkan linguistic features. In many cases, subsequent research has refined and expanded on his initial observation, but the seed of the idea can be identified in his work. Second, of necessity Sandfeld made use mostly of folkloric, i.e., dialect, texts. Sandfeld was writing at a time when folkloric texts were still the bedrock of nationalism and when Balkan national standard languages were at most young, and sometimes not yet existent. It can be noted that Sandfeld attributed the convergences to Greek and the East Roman (Byzantine) Empire. Attempts have also been made to credit the *sprachbund* to Latin, i.e., the Roman Empire (Haarmann, 1978). Nonetheless, while this or that feature may have an identifiable or possible source (cf., e.g., Hamp, 1982; Gołąb, 1997), such is not always the case, and it is actually the Ottoman Empire and the ‘4-M’ conditions that were obtained during it that produced the *sprachbund* as we know it (see, e.g., Section 3.1.1).

Table 20.1 The differences between a language family and a *sprachbund* (Trubetzkoy, 1923, 1930; translation Friedman and Joseph, forthcoming)

| <i>Language family</i> | <i>Sprachbund/Language area</i> |
|--|--|
| regular sound correspondences | other similarities in the structure of the sound system, but no regular sound correspondences |
| correspondences in the phonological expression of morphological categories | similarity in the principles of morphological structure, but no agreement in the phonological form of morphological elements |
| common basic vocabulary | a large number of common culture words, but no common basic vocabulary |
| – | a great similarity with respect to syntax |

Although research on the Balkans continued through the Great Depression, and even, to some extent, during World War II, the next major compendium was Georgiev, Gălăbov and Zaimov (1968), which was Volume VI of the seven volumes of proceedings of the First International Congress of the Association for Southeast European Studies held in Sofia in 1966. This volume, dedicated entirely to Balkan linguistics, had contributions from every academic Balkan linguist in the world at that time, despite the problems of the Cold War, and ran to 900 pages. Schaller (1975) was an attempt to update Sandfeld (1930), but see Joseph (1986) on its problems. As of 2019, more than a dozen handbooks and compendia giving a broad overview of Balkan linguistics and/or languages have appeared, none of them in English (see Friedman and Joseph (forthcoming) for lists and discussion; Tomić (2006) is rich in data, but, unfortunately, contains many errors, see Sims, 2008; Joseph, 2011). Of these, Asenova (2002) is the most significant advance beyond Sandfeld (1930) for the four classic Balkan language groups, despite its Bulgarian bias, and it has a significant resume in English. A feature that is both a strength and a weakness of this account is the fact that examples are almost all from standard languages. On the one hand, such an approach demonstrates the extent to which Balkan features survived the puristic tendencies of standardization. On the other hand, it is in the colloquial rather than the standard language that the speaker-to-speaker contact that produces a sprachbund can best be seen.

3. Critical issues and topics

In the simplest terms, the critical issues and topics can be stated as a compound question: what features of a Balkan language are Balkanisms (in Seliščev's (1925) sense), and where do they come from? In the case of elements such as loanwords and the occasional morpheme or phonological element, this answer is usually – albeit not always – fairly straightforward. In the case of morphosyntax and calquing, there are, on occasion, two challenges: directionality (i.e., which language was the source of the feature) and typology (this latter in Hamp's, 1977 sense of achronic resemblance; cf. also Matras and Sakel, 2007 on *matter vs pattern*). An additional challenge is sometimes motivated by the kind of nineteenth-century ideology discussed in Section 2.2, i.e., the desire to explain an item or feature as an internal development rather than the result of contact-induced change (cf. the discussion in Kazazis, 1972, 1977). While such an ideology has radically affected the vocabulary of most Balkan languages, the colloquial has, in this respect, been more conservative than the written standards (see Friedman, 2005 on Romani and Aromanian, and likewise Kosovo Albanian). Section 3.1 treats the features most commonly cited in the early specialist and all the non-specialist literature. Section 3.2 examines the lexicon, and Section 3.3 discusses phonology.

3.1 Features and their origins: morphosyntax

This subsection is of necessity limited to the most salient commonly cited features. See Section 5 for a discussion of the uses and misuses of these items.

3.1.1 A particle derived from etymological 'want' to mark future

Although typologically unremarkable, when placed in its proper historical context, the use of a particle derived from 'want' is an areal feature on two counts: it developed as such during the Ottoman period (Asenova, 2002, p. 214), and it was selected over the use of 'have' for the rest of Romance, perfective 'be' for the rest of Slavic, and other constructions for the rest

of Romani (see Mufwene, this volume on feature selection, cf. also Adamou, 2010 on language ecology). It is also more widespread in Albanian than superficial accounts realize (all of southern Geg, parts of north-western and east central Geg). At the same time, negative future marking using ‘have’ or a possessive construction is typical of Balkan Slavic, West Rumelian Turkish, and Aromanian and Romani in Macedonia.

3.1.2 Replacement of infinitive by analytic subjunctive

As Joseph (1983) observed in his classic work, not every Balkan language or dialect replaced infinitives with subjunctives to the same extent. Moreover, infinitive replacement occurs outside of the Balkans. Still, as shown in Friedman and Joseph (2017), the more deeply embedded a language or dialect is in the Balkans (Tosk Albanian, Aromanian, West Macedonian, Greek), the weaker the category of infinitive. In the case of the former Serbo-Croatian, only the Torlak dialects show complete replacement, while for ideological reasons such replacement is discouraged in Standard Croatian. Replacement of infinitive by optative is also a feature of West Rumelian Turkish (Friedman, 1982; Ibrahim, 1982; Matras, 2003/2004), and Judezmo uses subordinate clauses in *ke* ‘that’ where Spanish would have an infinitive (Friedman and Joseph, 2014). Thessaloniki Armenian is characterized by less infinitive usage than dialects in Armenia (Adamou, 2008).

3.1.3 Postposed definite article

The fact that postposed definiteness marking occurs elsewhere is irrelevant to the historical timing of the rise of postposed definiteness marking in the Balkans, where the syntactic conditions differ from those of, e.g., Scandinavian, Armenian, or North Russian (which last is more emphatic than referential). The development of a postposed demonstrative into a postposed definite article in Romance and Slavic occurred while the speakers of these languages were in the Balkans, and in contact with the ancestor of Albanian, which may have already had the construction (Hamp, 1982; Gołąb, 1997).

3.1.4 Replacement of synthetic comparatives by analytic

The replacement of synthetic comparatives with analytic ones is most complete in the core Balkan area. As with many other features, it is timing and distribution that argue for language contact as playing a role, especially for Balkan Slavic and Romani, which latter always uses borrowed material from local contact languages.

3.1.5 Replacement of conditional by anterior future

Another typologically well attested feature whose timing and distribution in the Balkans point to contact-induced change. For both Romanian and Balkan Slavic alternative inherited conditionals remain in competition (Gołąb, 1964; Belyavski-Frank, 2003), but not in Aromanian, where the Balkan conditional has triumphed. The temporal trajectory is clear enough.

3.1.6 Object reduplication (resumptive clitic pronouns)

As Kallulli (2008) and Friedman (2008) have shown, the exact rules for resumptive clitic pronouns differ among the various Balkan languages in their details, but not in their broad

outlines. Especially telling in this respect is the fact that Aromanian follows the reduplication rules of Greek, Albanian, or Macedonian, depending on which is the current nation-state language of contact (cf. Beis, 2000; Markovikj, 2007), and Balkan Judezmo employs the construction where no other Ibero-Romance language does. As with other features, while traces of the phenomenon can be found prior to the Ottoman period, the current systems took their shape at that time.

3.1.7 *Simplification of the declensional system*

The paucity of case marking in Balkan Slavic attracted the attention of Slavic Balkanists in the nineteenth and early twentieth centuries because non-Balkan Slavic has been relatively conservative in preserving inherited case marking, albeit in restructured systems in every instance. Wahlström (2015), however, is the first detailed analysis of the processes of Balkan Slavic case loss. Romani, too, has been quite conservative: its restructured case system happens to recapitulate that of Sanskrit, albeit with new material (Friedman, 1991; Matras, 2002, pp. 78–94). A careful examination of the loss of synthetic substantival case marking in Balkan Slavic, however, reveals that precisely at all the margins, i.e., zones of most contact, accusative, and sometimes dative, marking is preserved on some substantives or definite forms thereof, much as in the respective non-Balkan contact languages. It is thus the case that both case marking – as well as its replacement, i.e., prepositional usages (Asenova, 2002, pp. 97–103; Markovikj, 2015) – show patterns that point to contact, but not always a straightforward simplification.

3.1.8 *Clitic order*

The alignment of Balkan Slavic clitic order with non-Slavic contact languages is never cited in superficial accounts, but the patterning is obviously Balkan. All the Indo-European languages started out with some sort of Wackernagelian clitic ordering, with clitics coming after the first stressed element in a clause, and the former Serbo-Croatian still has such a rule. In Bulgarian (and eastern Macedonian), clitics are tied to the verb phrase, coming before the finite verb unless it is clause initial, in which case they must follow. In Macedonian, however, as in Albanian, Balkan Romance, and Greek, clitics come before the finite verb even if this makes them clause initial.

3.1.9 *Evidentials*

Turkish entered the Balkans with an evidential system at a time when none of the other Balkan languages had one. This can even be said for Albanian, since the admirative was in the process of shifting from a modal to an evidential at the time of the earliest dated, extensive documents (sixteenth century). Thus, the evidential strategies (Aikhenvald, 2003) of Albanian and Balkan Slavic, both of which began by utilizing native material, was subsequently expanded in different ways in Albanian, Macedonian, and Bulgarian. From Albanian, a morpheme was reinterpreted as an evidential particle and borrowed into the Frasherote dialect of Aromanian in Gorna Belica (Bela di supră). Meglenoromanian was influenced by Macedonian in the same way as Albanian by Turkish, i.e., an inverted perfect (participle+auxiliary) was re-purposed to render what Friedman (2012) calls the admirative complex. Individual dialects of Balkan Romani borrow markers or particles from Turkish and Slavic to render various evidential effects (cf. also Matras, 1995 for Vlax Romani). See Friedman (2018a) for additional details and references.

3.1.10 ‘Teen’ as ‘on ten’

Given that this manner of forming teens is found throughout Slavic but only in Balkan Romance for Romance and also in Albanian, it looks like Slavic influence. However, Hamp (1992), on the basis of the fact that the numeral ‘ten’ is neuter in Latin but feminine in Albanian and Romanian but masculine in Slavic suggests that this could be an old development from the period when the ancestor of Albanian was in contact with the ancestor of Slavic north of the Carpathians (for which other evidence exists, Hamp, 1994, 2010). This would thus be a pre-Balkan areal feature that was copied into Balkan Latin under the influence of pre-Albanian.

3.1.11 Analytic perfect using ‘have’

The use of a possessive to express past resultativity is another typologically widespread feature (Benveniste, 1966). In Balkan Slavic, however, it was limited to southwestern Macedonian and the Macedonian and Bulgarian dialects spoken along the former Via Egnatia (Koneski, Vidoeski and Jašar-Nasteva, 1968, p. 534; Bojadžiev, 1991; Adamou, 2012; since World War Two, the ‘have’ perfect has spread throughout the Republic of North Macedonia as a result of its being part of the Macedonian standard language). Gołąb (1984, p. 135) provides convincing evidence that its presence in southwest Macedonian dialects is due to a calque on Balkan Romance, where the construction is attested already during the Latin period. For Greek, the ‘have’ perfect has a later, and different origin, having developed from a reinterpreted modal pluperfect during the early modern period (Friedman and Joseph, forthcoming and references therein). The Albanian perfect is already in place at the time of the earliest texts. In the extreme southwestern Macedonian dialects (Kostur-Korča (Greek Kastoria-Albanian Korça) region), the ‘have’ perfect has completely replaced the old ‘be’ perfect (except some admirative-complex usages of the latter), which is arguably connected to the complex multilingualism of the region (see Makartsev (2013) for a detailed study of the situation in Albania). Adamou (2012) suggests the possibility of multiple factors in the origins of the ‘have’ perfect, and this, too, is consistent with the facts. Thus, the ‘have’ perfect in the Balkans serves as an example of how superficial typology misses the historical facts essential for an account of contact-induced change. See also RMS on ‘have’ perfects in Romani (Matras and Elšik, 2001–2005, 2008).

3.2 Features and their origins: lexicon

The lexical aspect of the Balkan languages that first drew the attention of researchers was the presence of a large number of Turkish loanwords in all of them (e.g., Miklosich, 1884–90). In fact, all of the Balkan languages have contributed to one another’s vocabulary, albeit to varying extents and in varying lexical areas, with differences also depending on dialect and region. Most recently, Friedman and Joseph (2014, 2017, forthcoming) have identified a class of shared Balkan loanwords that they label Essentially Rooted in Conversation (ERIC). ERIC loans are typically closed class and borrowing resistant items such as pronouns, numerals, kinship terms, and bound morphology as well as conversational items that are not borrowing resistant but are likewise not ‘culture words’ (in Trubetzkoy’s, 1930 sense) such as discourse particles, interjections and connectives and also idioms, taboo expressions, and other phraseologies (cf. the Leipzig-Jakarta list in Tadmor, Haspelmath and Taylor, 2010, pp. 238–241). It is the fact that such lexicon passes from one language to another via face-to-face interaction among speakers that makes it characteristic, or even diagnostic, of a sprachbund (Friedman and Joseph, 2017, forthcoming). Mention should also be made here of the *Maly dialektologičeskij atlas*

balkanskih jazykov (MDABJa): *Serija leksičeskaja* of which the seventh volume (Domosilet-skaya and Sobolev, 2018) is the most recent. Each volume is dedicated to a specific cultural-lexical set of fields such as beekeeping, farming, gardening, meteorology, animal husbandry, family, spiritual culture, etc. (Sobolev et al., 2018). The MDABJa, whose overseeing editor is A. N. Sobolev, is based on a detailed morphosyntactic and lexico-ethnographic questionnaire for 12 locations: one each for south Aromanian, central Geg, northern Tosk, northern Greek, and southern Greek and seven Slavic points: one each in southwestern North Macedonia, southwestern Bulgaria, the Bulgarian Rhodopes, northeastern Bulgaria, southeastern Serbia, southern Montenegro, and central Dalmatia (Croatia). The complete questionnaire has been published for a number of these points, and there is also a grammatical series, of which the first volume has appeared (Sobolev, 2005).

3.3 Features and their origins: phonology

Unlike morphosyntax and lexicon, where features are shared over broad areas in the region, Balkan phonological developments are highly localized, i.e., there is no such thing as Balkan phonology but rather there are Balkan phonologies. Of the various phonological Balkanisms that have been proposed at least since Miklosich (1861), the most frequently cited is the presence of stressed schwa and/or a higher back unrounded vowel. Such a vowel, however has different origins and explanations in the dialects of a single language, much less the Balkan languages in general. Thus, for example, in northern Macedonian (and Torlak Serbian) stressed schwa comes from Common Slavic **ǔ* and **ĩ*, while in eastern Macedonian it comes from vocalic **ɪ*, but in adjacent Bulgarian dialects (and most but not all other Bulgarian dialects, Ivić, 1968) it comes from **ǔ* and back nasal **ǔ̃*, while in southwestern Macedonian it comes only from back nasal **ǔ̃*. Schwa also accompanies the reflexes of vocalic /*r*/ and /*l*/ in these dialects, as is also the case in most of Bulgarian (cf. also Trummer, 1983 on Slavic-Albanian denasalization patterns). Meanwhile, in the west-central Macedonian dialects, stressed schwa is completely absent – as is the case in Greek – except sometimes in Turkisms, which is also the case in Geg Albanian. In Tosk Albanian, stressed schwa comes from nasal *â* and *ê*, which however, are retained as such in Geg. In Romanian, stressed schwa comes from a variety of sources. Rosetti (1958) points to similarities with Bulgarian, which may in fact be relevant given the importance of Bulgarian and Church Slavonic in Romania into the nineteenth century. Schwa is also generally lacking in Romani, except in loanwords.

What we do see in the Balkans are localized shared phonological developments. Thus, for example, in the Slavic, Albanian, Turkish, and Romani dialects of Kosovo mellow palatals whose voiceless realization is represented by orthographic <ć> in Serbian, <q> in Albanian, /*k*+front vowel (including orthographic <â> and <û>) in Turkish, and /*k*+/j/ or front vowel in Romani (although the grapheme <q> is sometimes also used for this sound in Kosovo-based written Romani), merge with the strident palatal – Serbian, Romani <č>, Turkish, Albanian <ç> – which replaces the mellow palatal in pronunciation (cf. Ivić, 1968). Similarly, in northern Greece, local Aromanian and Macedonian dialects borrow interdental fricatives in Greek words, and these then make their way into native words, while in northern Greek dialects, palatal fricatives and affricates occur (Margariti-Ronga and Papadamou, 2019). The Geg dialects with fewest – or even no – nasals are those in western North Macedonia, where Macedonian-Albanian bilingualism has a tradition going back many generations. In fact, teasing out the details of contact-related phonological change is one of the current themes of research (see Section 4).

4. Current contributions and research

Current research on Balkan linguistics includes all those areas that have been traditional until now as well as newer directions. Two major traditional areas can be distinguished: historical studies and fieldwork studies. Both types of research are also in conversation with newer trends in linguistics. Although a number of single-language historical and synchronic studies, e.g., Markopoulos (2009) and Pană Dindelegan (2013), respectively, also pay some attention to Balkan linguistics, their monoglot focus is not about language contact *per se*. In the realm of historical linguistics, recent advances have been in understanding the Balkan phonologies alluded to in Section 3.3, but also in morphosyntax. Thus, for example, Dombrowski (2013, 2019) examines a number of phonological changes in particular dialects of Albanian, Balkan Slavic, and West Rumelian Turkish that resulted from Balkan language contact. Friedman (2018b), independently from, but parallel to, Lindstedt (2016) and Sawicka (2014), provides evidence that the preservation of nasality (by means of homorganic nasals before stops) in the reflexes of Common Slavic nasal vowels in peripheral southwestern Macedonian dialects – spoken in what is today northwestern Greece and southeastern Albania – is not an archaism but rather an archaic innovation due to language contact (with Albanian and Aromanian in the west and Greek in the south). Sawicka also looks at larger areal tendencies. Morozova and Rusakov (2018) is exemplary of the kind of qualitative ethnolinguistic fieldwork that is highly relevant both to analyzing the Balkans today and reconstructing the past. By examining not only dialectal features but also marriage patterns and who speaks in what language to whom as well as generational histories, they not only model how language contact is driving language change in the present, but also provide data that support speculations of how change was effected in the past. Adamou (2016) and Dombrowski (2019) are also exemplary of how modern approaches to Balkan linguistic data can inform linguistic theory in general. Dombrowski examines Albanian-Slavic contact-induced phonological changes as well as the effects of Balkan language contact on West Rumelian Turkish. Adamou uses Balkan Slavic, Romani, Turkish, and Greek data from northern Greece. Curtis (2012) examines Albanian-Slavic contact in general, and Prendergast (2017) examines locative determiner omission as a contact-induced change in Albanian, Aromanian, and Macedonian. Sobolev et al. (2018) is a collection of studies of a number of Balkan contact zones, including the core zone where today Albania, Greece, and North Macedonia share borders, i.e., the Ohrid-Prespa region, including Bitola, Korça, and Kastoria, as well as evidence of medieval Tsakonian-Slavic contact in the south of Greece, Slavic-Albanian contact in the highlands of north Albania and south Montenegro as well as eastern Albania and western North Macedonia.

Adamou and Shen (2019) is a particularly significant example of how Balkan linguistic research can contribute to linguistic theory writ large. Already in Weinreich (1953, pp. 116–122) the issue of ideological biases against bi/multi-lingualism were addressed, but Adamou and Shen's (2019) research demonstrates the importance of Balkan linguistics for larger theoretical and also practical concerns for a larger neurolinguistic field that, otherwise, has restricted itself to relatively simple situations available in the anglophone world (cf. Friedman, 2018c and sources cited therein). Elšík and Matras (2006) is especially relevant to theories of markedness. Also worth noting is the field of contact ethnolinguistics as exemplified by Makartsev and Wahlströhm (2016).

5. Main research methods

The bedrock of Balkan linguistics remains fieldwork studies and historical-textual analyses as exemplified in the works cited in Section 4. Such work includes newer methodologies such

as those enabled by corpora (Adamou, 2016) and novel considerations of dialectological data (Dombrowski, 2013, 2019) as well as more recent fields such as studies of linguistic landscapes (Kramer, Friedman and Ivković, 2014).

There are two problematic methodologies that should be mentioned here owing to their relative visibility. One is a numerological approach that picks a relatively small number of features (usually from those in Section 3.1) and attempts to quantify the relative ‘Balkanness’ of a language by their presence or absence. Such flawed methodology results in inaccurate pictures of the Balkans as a linguistic area. Thus, in one version (Campbell, Kaufman and Smith-Stark, 1986) Romanian is declared the ‘most Balkan,’ while in another (van der Auwera, 1998) it is Bulgarian, and in yet another the Balkans is a fragmented periphery to a chimerical ‘Charlemagne sprachbund’ (Haspelmath, 1998). As argued in Friedman and Joseph (2017), by ignoring the more nuanced picture that can only be obtained from both historical knowledge and a full consideration of dialectological data, the actual differential bindings (Hamp, 1989) that together make up the sprachbund are lost, and inaccurate characterizations are made. Thus, to cite an illustrative example, constructions of the type dative subject + intransitive verb to mean ‘feel like’ as in Macedonian *mi se jade burek* ‘I feel like eating burek (me.DAT ITR eat. PRS.3sg burek),’ if examined only in terms of standard languages, is found in Balkan Slavic, South Balkan Romance (Aromanian and Meglenoromanian), and Albanian, as well as Balkan dialects of Romani, but not in Romanian or Greek. However, the construction is found precisely in the Greek spoken around Kastoria, which was in contact with Macedonian in the past. Given that the construction is well attested throughout Slavic, it is arguably a Slavonicism in the other Balkan languages. A simple numerological approach, however, misses this and many other facts, and it ends up with misleading generalizations (cf. Papadamou and Papanastasiou, 2013; Friedman and Joseph, 2017, 2018; Guentchéva, 2010).

Formalist syntax of Balkan languages, as pointed out by Joseph (2001) is not concerned with the historical circumstances that produced the Balkans as a linguistic area. Rather, in its search for putative universals, it reveals interesting typological tendencies that are independent of the socio-historical realities of contact linguistics. Joseph (2001) therefore distinguishes between ‘comparative Balkan syntax’ and ‘comparative syntax of the Balkans.’ By comparative Balkan syntax, Joseph understands those features of the syntax of the various Balkan languages that can be argued to reflect contact-induced or convergent change over time, which change is grounded in the social conditions of what Friedman and Joseph call their ‘4-M’ model (see Section 1.1). Joseph’s understanding of comparative syntax of the Balkans is that the languages of the Balkans are, like any other human language, available for speculation on typological tendencies. But such speculations are not tied to any area. Once space is invoked, then time, too, should be taken into consideration.

6. Future directions

Perhaps the most urgent direction for Balkan linguistics, one that is being pursued on numerous fronts but could be expanded, is the documentation of remaining contact zones in the Balkans. The Austrian Academy of Sciences’ Vanishing Languages and Cultural Heritage web site is one of the few such resources focusing on the Balkans, as opposed to language-specific or more general web sites for endangered languages. Such documentation is especially important for Aromanian, the most endangered Balkan language, and for the minority languages of Greece and European Turkey, which both have many rich areas of multilingualism in need of documentation. At the same time, North Macedonia, where urban Balkan multilingualism remains significant and where the largest percentage of the population (roughly a third)

declares a language other than the eponymous nation-state one (Macedonian) as mother tongue, provides an ideal setting for the construction of a multilingual corpus. Moreover, all the other Balkan nation-states all have regions of Balkan multilingualism that are in need of documentation, some of it currently ongoing. The concept of a multilingual Balkan linguistic atlas, one that could take into account the various dialects of the various languages and oriented toward mapping the actual instantiations and occurrences of various Balkanisms, is, close to a century after it was first mooted (cf. Friedman, 2000a), a desideratum. Finally, it is worth noting the increased engagement of Balkan linguistics with various trends in general linguistics as indicated earlier. These are obviously engagements worth continuing.

7. Further reading

All of the works just cited are rich in possibilities for further reading. A few items not cited so far are added here.

Friedman, V.A. (2013). The languages of the Balkans. In: M. Aronoff, ed., *Oxford bibliographies online: Linguistics*, 1st ed. Oxford: Oxford University. [online] Available at: www.oxfordbibliographies.com/view/document/obo-9780199772810/obo-9780199772810-0108.xml [Accessed 30 May 2019].

This selective, annotated bibliography, organized by topics and including brief introductions and descriptions for each topic, makes an excellent starting point for an investigation of the Balkans as a linguistic area.

Kahl, T., Metzelin, M. and Schaller, H., eds. (2012). *Balkanismen heute – Balkanisms Today – Balkanizmy segodnja*. Vienna: Lit Verlag.

The articles in this collection by an international group of scholars address a broad range of Balkan linguistic topics.

Krapova, I. and Joseph, B.D., eds. (2019). *Balkan syntax and (Universal) principles of grammar*. Berlin: De Gruyter Mouton.

This collection of articles looks at the ways comparative Balkan syntax can inform general linguistic theory.

Sedakova, I., Makartsev, M. and Civ'jan, T., eds. (2019). *Balkanskij tezaurus: Kommunikacija v složno-kul'turnyh obščestvah na Balkanah* (Balkanskie čtenija 15). Moscow: Russian Academy of Sciences (RAN).

This collection of lengthy abstracts from an international Balkanological conference gives a good sense of various directions of research in Balkan linguistics as a broad field.

Tomić, O.M., ed. (2004). *Balkan syntax and semantics*. Amsterdam: John Benjamins.

This collection of articles, which served as the basis for Tomić (2006), has a good selection of topics and is not marred by the numerous errors in the latter work.

8. Related topics

Social factors, pragmatic factors, typological factors, borrowing, convergence, Anatolia

Abbreviations

| | |
|------|------------------------------------|
| 3sg | third person singular |
| DAT | dative |
| ERIC | Essentially Rooted in Conversation |
| ITR | intransitive |
| PRS | present |

References

- Adamou, E. (2008). Arménien. In: E. Adamou, ed., *Le patrimoine plurilingue de la Grèce – Le nom des langues II*, 1st ed. Louvain-la-Neuve: Peeters Publishers, pp. 71–76.
- Adamou, E. (2010). Bilingual speech and language ecology in Greek Thrace: Romani and Pomak in contact with Turkish. *Language in Society*, 39(2), pp. 147–171.
- Adamou, E. (2012). Le parfait avec “avoir” dans l’aire balkanique: Approche multifactorielle et diversifiée d’un balkanisme. In: C. Chamoreau and L. Goury, eds., *Changement linguistique et langues en contact*, 1st ed. Paris: CNRS Éditions, pp. 101–109.
- Adamou, E. (2016). *A corpus driven approach to language contact*. Boston and Berlin: De Gruyter Mouton.
- Adamou, E. and Shen, X.R. (2019). There are no language switching costs when codeswitching is frequent. *International Journal of Bilingualism*, 23, pp. 53–70.
- Aikhenvald, A.Y. (2003). Evidentiality in typological perspective. In: A.Y. Aikhenvald and R.M.W. Dixon, eds., *Studies in evidentiality*, 1st ed. Amsterdam: John Benjamins, pp. 1–31.
- Asenova, P. (2002). *Balkansko ezikoznanie*, 2nd ed. Veliko Tărnovo: Faber.
- Beis, S. (2000). *Le parler aroumain de Metsovo: Description d’une langue en voie de disparition*. PhD. Université Paris 5, René Descartes.
- Belyavski-Frank, M. (2003). *The Balkan conditional in South Slavic*. Munich: Otto Sagner.
- Benveniste, É. (1966). La construction passive du parfait transitif. In: É. Benveniste, ed., *Problèmes de Linguistique Générale*, 1st ed. Paris: Gallimard, pp. 176–186.
- Bojadžiev, T. (1991). *Bălgarskite govori v Zapadna (Belomorska) i Iztočna (Odrinska) Trakija*. Sofia: Universitetsko izdatelstvo “Sv. Kliment Ohridski”.
- Boretzky, N. and Iglă, B. (2004). *Komentierter dialektatlas des Romani*, 2 vols. Wiesbaden: Harrassowitz.
- Campbell, L., Kaufman, T. and Smith-Stark, T. (1986). Meso-America as a linguistic area. *Language*, 62, pp. 530–570.
- Curtis, M.C. (2012). *Slavic-Albanian language contact, convergence, and coexistence*. PhD. The Ohio State University.
- Dombrowski, A. (2013). *Phonological aspects of language contact along the Slavic periphery: an ecological approach*. PhD. University of Chicago.
- Dombrowski, A. (2019). Laterals and fricatives in the Albanian-Slavic contact zone: Micro-scale phenomena with macro-scale ramifications. In: J. Pennington, V. Friedman and L. Grenoble, eds., *And thus you are everywhere honored: Papers to honor Brian D. Joseph*, 1st ed. Bloomington, IN: Slavica, pp. 53–64.
- Domosiletskaya, M.V. and Sobolev, A.N. (2018). *Malyi dialektologičeskij atlas balkanskih jazykov. Serija leksičeskaja. Vol. VII. Pčelovodstvo*. Saint Petersburg: Nauka.
- Elšik, V. and Matras, Y. (2006). *Markedness and language change: The Romani sample*. Berlin: De Gruyter Mouton.
- Emeneau, M. (1956). India as a linguistic area. *Language*, 32, pp. 3–16.
- Fine, J.V.A. (1983). *The early medieval Balkans: A critical survey from the sixth to the late twelfth century*. Ann Arbor: University of Michigan.
- Fine, J.V.A. (1987). *The late medieval Balkans: A critical survey from the late twelfth century to the Ottoman conquest*. Ann Arbor: University of Michigan.
- Friedman, V.A. (1982). Balkanology and Turcology: West Rumelian Turkish in Yugoslavia as reflected in prescriptive grammar. In: A.A. Barentsen, R. Sprengers and M.G.M. Tielemans, eds., *South Slavic and Balkan linguistics*, 1st ed. Amsterdam: Rodopi, pp. 1–77.
- Friedman, V.A. (1991). Case in Romani: Old grammar in new affixes. *Journal of the Gypsy Lore Society (Fifth Series)*, 1(2), pp. 85–102.
- Friedman, V.A. (2000a). After 170 years of Balkan linguistics: Whither the millennium? *Bulletin de l’Association Internationale d’Études du Sud-Est Européen*, 30, pp. 203–216.
- Friedman, V.A. (2000b). Romani in the Balkan linguistic league. In: K. Tzitzilis and K. Symeonidēs, eds., *Valkaniké Glossologia: Sygkhronia kai Diakhronia*, 1st ed. Thessaloniki: University of Thessaloniki, pp. 95–105.

- Friedman, V.A. (2005). From orientalism to democracy and back again. In: R. Detrez and P. Plas, eds., *Developing cultural identity in the Balkans: Convergence vs. divergence*, 1st ed. Berlin: Peter Lang, pp. 25–43.
- Friedman, V.A. (2006). West Rumelian Turkish in Macedonia and adjacent areas. In: H. Boeschoten and L. Johanson, eds., *Turkic languages in contact*, 1st ed. Wiesbaden: Harrassowitz, pp. 27–45.
- Friedman, V.A. (2008). Balkan object reduplication in areal and dialectological perspective. In: D. Kallulli and L. Tasmowski, eds., *Clitic doubling in the Balkan languages*, 1st ed. Amsterdam: John Benjamins, pp. 25–63.
- Friedman, V.A. (2012). Perhaps mirativity is phlogiston, but admirativity is perfect. *Linguistic Typology*, 16(2), pp. 505–527.
- Friedman, V.A. (2018a). Where do evidentials come from? In: A.Y. Aikhenvald, ed., *The Oxford handbook of evidentiality*, 1st ed. Oxford: Oxford University, pp. 124–147.
- Friedman, V.A. (2018b). Reflexes of Common Slavic nasal vowels in Southwest Macedonian dialects revisited: An areal and Balkanological account. In: C. Bethin, ed., *American contributions to the XVI international congress of Slavists: Linguistics*, 1st ed. Bloomington, IN: Slavica, pp. 85–98.
- Friedman, V.A. (2018c). Healthy and wealthy and wise: Evidence of the benefits of multilingualism from neuropsychology and culture. In: V. Latifi, V.A. Friedman, and M. Markovikj, eds., *Multiculturalism and language contact*, 1st ed. Tetovo: Southeast European University, pp. 13–24. [e-book] Available at: https://drive.google.com/file/d/1q4W_CpBnuJoNRbsDI4M_NrdjOJ-4yxoW/view [Accessed 30 May 2019].
- Friedman, V.A. and Joseph, B.D. (2014). Lessons from Judezmo about the Balkan sprachbund and contact linguistics. *International Journal of the Sociology of Language*, 226, pp. 3–23.
- Friedman, V.A. and Joseph, B.D. (2017). Reassessing sprachbunds: A view from the Balkans. In: R. Hickey, ed., *Cambridge handbook of areal linguistics*. 1st ed. Cambridge: Cambridge University, pp. 55–87.
- Friedman, V.A. and Joseph, B.D. (2018). Non-nominative and depersonalized subjects in the Balkans: Arealty vs. genealogy. In: J. Barðdal, N. Pat-El and S.M. Carey, eds., *Non-canonically case-marked subjects. The Reykjavík-Eyjafjallajökull papers*, 1st ed. Amsterdam: John Benjamins, pp. 23–53.
- Friedman, V.A. and Joseph, B.D. (forthcoming). *The Balkan languages*. Cambridge: Cambridge University.
- Georgiev, V., Gălăbov, I. and Zaimov, J. (1968). *Actes du Premier Congrès International des Etudes Balkaniques et Sud-est Européennes, Vol. VI (Linguistique)*. Sofia: Bălgarska Akademija na Naukite.
- Gołąb, Z. (1964). *Conditionalis typu balskańskiego w językach południowosłowiańskich*. Cracow: Polska Akademia Nauk.
- Gołąb, Z. (1984). *The Arumanian Dialect of Kruševo in SR Macedonia SFR Yugoslavia*. Skopje: Makedonska Akademija na Naukite i Umetnostite.
- Gołąb, Z. (1997). The ethnic background and internal linguistic mechanism of the so-called Balkanization of Macedonian. *Balkanistica*, 10, pp. 13–19.
- Guentchéva, Z. (2010). L'impersonnel dans les langages balkaniques: Convergences et divergences. In: P. Assenova, A. Petrova and Ts. Ivanova, eds., *Glagolnata sistema na balkanskite ezici – nasledstvo i neologija/The Verbal System of the Balkan Languages – Heritage and Neology*, 1st ed. Veliko Tărnovo: Faber, pp. 24–50.
- Haarmann, H. (1978). *Balkanlinguistik*. Tübingen: Tübinger Beiträge zur Linguistik.
- Hammond, C.S. (1923). *1923 Atlas of the world and gazetteer*. New York and London: Funk and Wagnalls.
- Hamp, E.P. (1977). On some questions of areal linguistics. In: K. Whistler, R.D. van Valen, Jr., C. Chiarello, J.J. Jaeger, M. Petrucek, H. Thompson, R. Javkir and A. Woodbury, eds., *Proceedings of the 3rd annual meeting of the Berkeley linguistics society*. Berkeley: Berkeley Linguistics Society, pp. 279–282.
- Hamp, E.P. (1982). The oldest Albanian syntagma. *Balkansko ezikoznanie*, 25, pp. 77–79.
- Hamp, E.P. (1989). Yugoslavia – A crossroads of Sprachbünde. *Zeitschrift für Balkanologie*, 25(1), pp. 44–47.

- Hamp, E.P. (1992). Albanian. In: J. Gvozdanović, ed., *Indo-European numerals*, 1st ed. Berlin: De Gruyter Mouton, pp. 835–921.
- Hamp, E.P. (1994). Albanian. In: R.E. Asher, eds., *The encyclopedia of language and linguistics*, 1st ed. Oxford: Pergamon, pp. 65–67.
- Hamp, E.P. (2010). On the etymology of Crimean Gothic *Apel*. *North-Western European Language Evolution (NOWELE)*, 58/59, pp. 453–480.
- Hamp, E.P. (2013). (with annotations and comments by D.Q. Adams). The expansion of the Indo-European languages: An Indo-Europeanist's evolving view. *Sino-Platonic Papers*, 239, pp. 1–10.
- Haspelmath, M. (1998). How young is standard average European? *Language Sciences*, 20, pp. 271–287.
- Ibrahimi, S. (1982). Ndikimi i gjuhës shqipe në rrafshin fonologjik të turqishtes së folur në Maqedoni dhe Kosovë. *Studime fililogjike*, 36(19.2), pp. 51–61.
- Ilievski, P. (1973). Kon interpretacijata na modelot na udvoeniot objekt vo makedonskiot jazik. *Godišen zbornik, Filozofski fakultet na Univerzitetot Skopje*, pp. 205–220.
- Ivić, P. (1968). Liens phonologiques entre les langues balkaniques. In: V. Georgiev, I. Gălăbov and J. Zaimov, eds., *Premier Congrès International des Etudes Balkaniques et Sud-est Européennes, Vol. VI (Linguistique)*. Sofia: Bălgarska Akademija na Naukite, pp. 133–144.
- Jireček, C. (1911). *Geschichte der Serben*. Gotha: Friedrich Andreas Perthes.
- Joseph, B.D. (1983). *The synchrony and diachrony of the Balkan infinitive*. Cambridge: Cambridge University.
- Joseph, B.D. (1986). A fresh look at the Balkan sprachbund: Some observations on H. W. Schaller's *Die Balkansprachen*. *Mediterranean Language Review*, 3, pp. 105–114.
- Joseph, B.D. (2001). Is Balkan comparative syntax possible? In: M.L. Rivero and A. Ralli, eds., *Comparative syntax of Balkan languages*, 1st ed. Oxford: Oxford University, pp. 17–43.
- Joseph, B.D. (2011). Review of Tomić 2006. *Slavica Iaponica*, 29, pp. 123–131.
- Joseph, B.D. and Brown, C.G. (2015). On hybrid forms in language contact: Some evidence from the Greek of southern Albania. *Albano-Hellenica*, 6, pp. 1–8.
- Kahl, T., Metzelin, M. and Schaller, H., eds. (2012). *Balkanismen heute – Balkanisms Today – Balkanizmy segodnja*. Vienna: Lit Verlag.
- Kaimio, J. (1979). *The Romans and the Greek language (Comentationes Humanarum Litterarum 64)*. Helsinki: Societa Scientarium Fennica.
- Kallulli, D. (2008). Clitic doubling, agreement and information structure: The case of Albanian. In: D. Kallulli and L. Tasmowski, eds., *Clitic doubling in the Balkan languages*, 1st ed. Amsterdam: John Benjamins, pp. 227–236.
- Kazazis, K. (1972). The status of Turkisms in the present-day Balkan languages. In: H. Birnbaum and S. Vryonis, Jr., eds., *Aspects of the Balkans: Continuity and change. contributions to the international Balkan conference held at UCLA, October 23–28, 1969*. The Hague: Mouton, pp. 87–116.
- Kazazis, K. (1977). Review of Nà ksetourképsoume të glōssa mas, Giákōbos S. Dizikirikēs [Let Us Deturkicize Our Language!, Giakovos S. Dizikirikis]. *Folia Slavica*, 1(2), pp. 301–306.
- Koneski, B., Vidoeski, B. and Jašar-Nasteva, O. (1968). Distribution des balkanismes en macédonien. In: V. Georgiev, I. Gălăbov and J. Zaimov, eds., *Premier Congrès International des Etudes Balkaniques et Sud-est Européennes, Vol. VI (Linguistique)*. Sofia: Bălgarska Akademija na Naukite, pp. 517–546.
- Kopitar, J. (1829). Albanische, walachische und bulgarische Sprache. *Jahrbücher der Literatur*, 46, pp. 59–106.
- Kramer, C.E., Friedman, V.A. and Ivković, D. (2014). Seeing double: Latin and Cyrillic in the linguistic landscapes of Macedonia. In: L. Mitkovska, ed., *Proceedings of the international applied linguistics conference practice in language, language in practice*. Skopje: FON University, pp. 14–20.
- Leake, W.M. (1814). *Researches in Greece*. London: John Booth.
- Lindstedt, J. (2016). South Macedonian decomposed nasal vowels are not an archaism but an early Balkanism. In: *Paper presented at the eleventh meeting of the Slavic linguistic society, 24–26 September*. Toronto, CA.
- Makartsev, M.M. (2013). Innovacii v glagol'noj sisteme slavjanskogo govora Bobošticy (obl. Korči, Albanija). In: A.M. Moldovan and S.M. Tolstaja, eds., *Slavjanskoe jazykoznanie: XV Meždunarodnyj s'ezd slavistov*, 1st ed. Moscow: Indrik, pp. 512–534.

- Makartsev, M.M. and Wahlströhm, M., eds. (2016). *In search of the center and periphery: Linguistic attitudes, minorities, and landscapes in the central Balkans*. (Slavic Helingensia 49). Helsinki: University of Helsinki.
- Mareş, A. (2000). Considerații pe marginea datării Psaltiri Humuzaki. *Limba română*, 49, pp. 675–683.
- Margariti-Ronga, M. and Papadamou, E. (2019). Macedonian-Epirote isoglosses. In: A. Ralli, P. Barka, B.D. Joseph, and S. Bompalas, eds., *MGDLT8, Proceedings of the 8th international conference of modern Greek dialects and linguistic theory (Gjirokastër, 4–6 October 2018)*. Patras: University of Patras, pp. 33–143.
- Markopoulos, T. (2009). *The future in Greek from ancient to medieval*. Oxford: Oxford University.
- Markovikj, M. (2007). *Aromanskiot i makedonskiot govor od ohridsko-struškiot region (vo balkanski kontekst)*. Skopje: Makedonska Akademija na Naukite i Umetnostite.
- Markovikj, M. (2015). The dominant exponents of dynamic spatial relations in Macedonian (in a Balkan context). In: D.L. Dyer, B.D. Joseph and C.E. Kramer, eds., *Ođ Čukazo u nazad: Papers to Honor Victor A. Friedman on the Occasion of his Retirement, Balkanistica*, 28, pp. 295–310.
- Matras, Y. (1994). Structural Balkanisms in Romani. In: N. Reiter, ed. *Sprachlicher Standard und Substandard in Südosteuropa und Osteuropa*, 1st ed. Wiesbaden: Harrassowitz, pp. 195–210.
- Matras, Y. (1995). Verb evidentials and their discourse function in Vlach Romani narratives. In: Y. Matras, ed., *Romani in contact: The history, structure, and sociology of a language*, 1st ed. Amsterdam: John Benjamins, pp. 95–123.
- Matras, Y. (2002). *Romani: A linguistic introduction*. Cambridge: Cambridge University.
- Matras, Y. (2003/2004). Layers of convergent syntax in Macedonian Turkish. *Mediterranean Language Review*, 15, pp. 63–86.
- Matras, Y. (2005). The classification of Romani dialects: A geographic-historical perspective. In: B. Schrammel, D. Halwachs and G. Ambrosch, eds., *General and applied Romani linguistics: Proceedings of the sixth international conference on Romani linguistics*. Munich: Lincom Europa, pp. 7–22.
- Matras, Y. and Elšik, V. (2001-2005, 2008). *Romani morpho-syntactic database*. University of Manchester. [online] Available at: <http://romani.humanities.manchester.ac.uk/rms/> [Accessed 30 May 2019].
- Matras, Y. and Sakel, J. (2007). Investigating the mechanisms of pattern replication in language convergence. *Studies in Language*, 31, pp. 829–865.
- Matras, Y. and Tufan, Ş. (2007). Grammatical borrowing in Macedonian Turkish. In: Y. Matras and J. Sakel, eds., *Grammatical borrowing in cross-linguistic perspective*. 1st ed. Berlin: De Gruyter Mouton, pp. 215–227.
- Miklosich, F. (1861). Die slavischen Elemente im Rumunischen. *Denkschriften der Kaiserlichen Akademie der Wissenschaften, Philosophisch-historische Klasse*, 12, pp. 1–70.
- Miklosich, F. (1884–1890). Die türkischen Elemente in den südost- und osteuropäischen Sprachen. *Denkschriften der kaiserlichen Akademie der Wissenschaften in Wien, philosophisch-historische Klasse*, 34(35), pp. 37, 38.
- Minchin, J.G.C. (1886). *The growth of freedom in the Balkan Peninsula*. London: John Murray.
- Morozova, M.S. and Rusakov, A. Yu. (2018). Slavic-Albanian interaction in Velja Gorana: Past and present of a balanced language contact situation. In: V. Latifi, V.A. Friedman and M. Markovikj, eds., *Multiculturalism and language contact*. [ebook]. Tetovo: Southeast European University, pp. 38–59. Available at: https://drive.google.com/file/d/1q4W_CpBnuJoNRbsDI4M_NrdjOJ-4yxoW/view [Accessed 30 May 2019].
- Pană Dindelegan, G., ed. (2013). *The grammar of Romanian*. Oxford: Oxford University Press.
- Papadamou, E. and Papanastassiou, G. (2013). The position of the northern dialects of the prefecture of Kastoria. In: M. Janse, B.D. Joseph, A. Ralli and M. Bagriacik, eds., *Proceeding of the fifth international conference on modern Greek dialects and linguistic theory (MGDLT 5)*. Patras: University of Patras, pp. 389–400.
- Prendergast, E.H. (2017). *The origin and spread of locative determiner omission in the Balkan linguistic area*. PhD. University of California, Berkeley, CA.
- Rosetti, A. (1958). Slavo-Romanica: sur la constitution du système vocalique du roumain. *Romano-slavica*, 1, pp. 27–30.

- Sandfeld, K. (1926). *Balkanfilologien*. Copenhagen: Bianco Lunos Bogtrykkeri.
- Sandfeld, K. (1930). *Linguistique balkanique*. Paris: Klincksieck.
- Sawicka, I. (2014). *Fonetskite promeni vo arelna lingvistika (vrz osnova na jazičnata situacija vo jugoistočna Evropa)*. Skopje: Makedonska Akademija na Nauite i Umetnostite.
- Schaller, H.W. (1975). *Die Balkansprachen. Eine Einführung in die Balkanphilologie*. Heidelberg: Carl Winter.
- Schleicher, A. (1850). *Die Sprachen Europas in systematischer Übersicht*. Bonn: König.
- Schubert, G. (1999). Der Einfluss des Ungarischen in Südosteuropa. In: U. Hinrichs, ed., *Handbuch der südosteuropa-Linguistik*, 1st ed. Wiesbaden: Harrassowitz, pp. 677–700.
- Sedakova, I., Makartsev, M. and Civ'jan, T., eds. (2019). *Balkanskij tezaurus: Kommunikacija v složnokulturnyh obščestvah na Balkanah* (Balkanske čtenija 15). Moscow: Rossijskaja Akademija Nauk.
- Seliščev, A. (1925). Des traits linguistiques communs aux langues balkaniques: un balkanisme ancien en bulgare. *Révue des Études Slaves*, 5, pp. 38–57.
- Sims, A. (2008). Review of Tomić 2006. *Journal of Slavic Linguistics*, 16, pp. 331–347.
- Sobolev, A.N., ed. (2005). *Malyi dialektologičeskij atlas balkanskih jazykov. Serija grammatičeskaja. Vol. I. Kategorii imeni suščestvitel'nogo*. München: Verlag Otto Sagner.
- Sobolev, A.N., Kisiler, M.L., Kozak, V.V., Koner, D.V., Makarova, A.L., Morozova, A.S. and Rusakov, A. Y. (2018). Južnoslavjanske dialektu v simbiotičeskih soobščestvah Balkan. *Acta linguistica petropolitana*, 14, pp. 1–10.
- Stolz, T. (2006). All or nothing. In: Y. Matras, A. McMahon and N. Vincent, eds., *Linguistic areas: Convergence in historical and typological perspective*, 1st ed. Houndmills, Basingstoke and Hampshire: Palgrave Macmillan, pp. 32–50.
- Sundhaussen, H. (1999). Südosteuropa und Balkan: Begriffe, Grenzen, Merkmale. In: U. Hinrichs, ed., *Handbuch der südosteuropa-Linguistik*, 1st ed. Wiesbaden: Harrassowitz, pp. 7–42.
- Tadmor, U., Haspelmath, M. and Taylor, B. (2010). Borrowability and the notion of basic vocabulary. *Diachronica*, 27(2), pp. 226–246.
- The Austrian Academy of Sciences' Vanishing Languages and Cultural Heritage. [online] Available at: www.oeaw.ac.at/vlach.
- Thomason, S.G. and Kaufman, T. (1988). *Language contact, creolization, and genetic linguistics*. Berkeley: University of California.
- Todorova, M. (1997). *Imagining the Balkans*. Oxford: Oxford University.
- Tomić, O.M., ed. (2004). *Balkan syntax and semantics*. Amsterdam: John Benjamins.
- Tomić, O.M. (2006). *Balkan morpho-syntactic features*. Dordrecht: Springer.
- Trubetzkoy, N.S. (1923). Vavilonskaja bašnja i smešenie jazykov. *Evrazijskij vremennik*, 3, pp. 107–124.
- Trubetzkoy, N.S. (1930). Proposition 16. In: C. de Boer, J. van Ginneken and A.G. van Hamel, eds., *Actes du Premier Congrès International des Linguistes à La Haye, du 10–15 Avril 1928*. Leiden: A.W. Sijthoff, pp. 17–18.
- Trummer, M. (1983). Südslavische Sprachgeographie und Balkanlinuistik. In: N. Reiter, ed., *Ziele und Wege der Balkanlinguistik*. Berlin: Osteuropa-Institut an der Freien Universität Berlin, pp. 245–254.
- van der Auwera, J. (1998). Revisiting the Balkan and Meso-American linguistic areas. *Language Sciences*, 20(3), pp. 259–270.
- Wahlström, M. (2015). *Loss of case inflection in Bulgarian and Macedonian*. Helsinki: University of Helsinki.
- Weinreich, U. (1953). *Languages in contact*. The Hague: Mouton.
- Wittek, P. (1953). Les gagaouzes—gens de Kaykâüs. *Rocznik orientalistyczny*, 17, pp. 12–24.
- Woodard, R.D. (2008). Introduction. In: R.D. Woodard, ed., *The ancient languages of Europe*. Cambridge: Cambridge University, pp. 1–18.

Anatolia

Anaïd Donabedian and Ioanna Sitaridou¹

1. Introduction

The geographical extent of Anatolia (Turkish Anadolu, also called Asia Minor) roughly corresponds to the Asian part of contemporary Turkey and is bordered by Mesopotamia, Iran and the Caucasus. Anatolia hosted people and languages from various phyla, some of which have since become extinct. Alongside Hattic and Hurro-Iranian, the entire Anatolian branch of the Indo-European languages comprising Hittite, Luwian, Palaic, Lycian, Carian, Lydian, Pisidian, Sidetic, and later, Phrygian also perished ca. 200 BC (see Watkins, 2001, p. 62). These languages, together with Armenian, Greek and Aramaic, are considered to have constituted an ‘Old Anatolian’ or Graeco-Anatolian Sprachbund since the second millennium BC (Luraghi, 2010; Watkins, 2001; Romagno, 2015).

Since late antiquity and during the medieval period, the main language families represented in the area are Indo-European (Greek, Armenian, Iranian, Kurdish), Semitic (Aramaic, Arabic), Turkic (Saljuqian/Seljukian or Western Oghuz, Kipchak), Kartvelian (Laz), and to some extent, North-West Caucasian (Adyghe). It is precisely on this more recent Anatolian Sprachbund (in the sense of Trubetzkoy, 1928, p. 18, but see also Tzitzilis, to appear, on a different definition) that we focus in this chapter.

1.1 Historical background

Language contact in Anatolia arose out of complex geopolitical conditions such as moving boundaries and wars, implantation of tribes or clans, migrations and deportations; cultural practices such as endogamy-exogamy, religious affiliation and influence, and prestige; lifestyle such as territory-sharing practices (nomadism vs. sedentarism); never-ending identity-making processes such as tribal/ethnic affiliation and identity (ex)appropriation; social practices such as community organization and chiefdoms; education policies, if any, the influence of a standard/‘high’ language on the spoken language, degrees of diglossia, bidialectalism and bilingualism; and socio-economic factors such as land and resources management, ethnic repartition of professions, fiscality, etc.

Since the end of the Persian Achaemenid Empire and the beginning of the Hellenistic period (323 BC), Western and Central Anatolia was under Greek rule. Iranian domination remained

in South-East Anatolia, which was ruled by the Arsacide dynasty, followed by the Sassanides until the Arab invasions in AD 650. Importantly, the cultural influence of Greek became increasingly significant in the whole Anatolia with the Christianization of Greeks, Armenians and Assyrians and other autochthonous phyla since the inception of Christianity and definitely by late second century. Greek progressively supplanted Aramaic as the language of administration even in the fringes of Anatolia, for instance, in the Armenian plateau. Since the late eleventh century, the incipient Turkification of Anatolia progressively culminated to the formation of the Ottoman Empire in the fourteenth century, which finds Anatolia under a single political rule with increasing domination of Turkish (Oghuz) in an otherwise multilingual context. The system of millet devolving cultural and educational autonomy to most of the confessional minorities of the empire, allowed the existence of cultural and educational institutions, some of which were even allowed to implement their own language and educational policy in the nineteenth century. However, the system was sectarian with no language or educational rights or recognition for non-Turkic-speaking Muslim minorities (such as native Kurds, Zazakis, Arabs, Laz, or the significant community of Circassians settled in Anatolia since 1855). Moreover, the inegalitarian socio-economic and judiciary system of the Ottoman Empire, despite often privileging the non-Muslim populations (namely, Christians and Jews) from an economic point of view, also led to the development of local strategies for constant ‘identity negotiation’ often resulting in language and/or religious shift.

Religious belonging is an important factor for language-community building in Anatolia. It is therefore unsurprising that when a community experiences partial or collective conversion, linguistic diversification ensues soon after, often leading to language loss or shift. As shown by Khan (2007b), Jewish and Christian populations settled in the same village show important dialectal variation in Neo-Aramaic; likewise by Napiórkowska (2013) for Urmia (spoken in the borders of Anatolia with Iran); Akkus (2018, pp. 460, 469), following Blanc (1964), notices the same phenomenon for speakers of Arabic: ‘Muslim, Jews and Christian spoke radically different dialects despite living in the same town.’ Moreover, Sitaridou (2014b, p. 29) argued that religious separation of the converted Greek-speaking communities in Pontus led to further isolation and thus survival of archaic features in the Pontic Greek varieties spoken by the Islamized populations. Similarly, Hemshinli Armenian following Islamization in the eighteenth century: while the converted group stayed in situ, the non-converted one migrated in Abkhazia. The latter, because of contact with Standard Eastern Armenian through education during the Soviet period, developed differently.

Having said this, religion was also the catalyst in establishing intra-community solidarities, hence intensified language contact, even in trans-religious contexts. It is well-known that Neo-Aramaic speakers in Anatolia used to be trilingual in North-Eastern Neo-Aramaic (NENA), Turkish and Arabic (Chyet, 1995); in Urfa, the Arabic-speaking women’s extra-community contact was with Armenian speakers, rather than Turkish-speaking people, which has resulted nowadays in a community of Assyrians who are native speakers of the Armenian dialect of Urfa despite having been born in Aleppo. To the ever-complex Anatolian context, we should also mention cases where language shift took place without being preceded by religious conversion. For instance, Greek orthodox communities shifted to Turkish in Cappadocia (Karamanlides) and Pontus (Urum) (before the latter moving to Georgia and south-eastern Ukraine, see Skopeteas, 2014).

Religion aside, the relationship of a spoken variety to its closest standard cognate – the latter generally emerged or became more prominent during the nineteenth century as a result of the rise of national awareness in Europe, the Balkans and the collapsing Ottoman Empire – also played an important role in shaping the evolution of the spoken variety. In fact, it could be argued

that under the influence of the standard (soon-to-become national) language, contact between genetically distinct languages was progressively inhibited if not stigmatized. It is this period after all during which ‘purity,’ linguistic or otherwise, became the main desideratum (see for instance Fallmerayer, 1845). During this period, some speech communities, such as Armenian and Greek, developed an expansive network of schools, which promoted the ‘high’ register and thus greatly affected the spoken (and low-register) varieties. However, other languages in the region were not concerned with such competition with a more prestigious variety, either because they were not in large-scale contact with the cognate ‘high’/standardized language, as was the case for Arabic of Anatolia, or because there was no cognate standard, as was the case for Neo-Aramaic and Kurdish.

1.2 Profiling the languages of Anatolia ca. 1900

Here we sketch the main languages of Anatolia ca. 1900 (that is, shortly before the massive population shifts in the area), bringing into consideration some (by no means exhaustive) elements of ethno-religious affiliation in order to convey the complexity of the interplay between language, religion and ethnicity – albeit, undoubtedly the latter cannot be considered/conceptualized by the standards of today. Some of these speech communities have phenomenally survived in remote pockets (e.g., Romeyka) of Turkey today, or not so remote (e.g., Zazaki). We indicate whether they are still to be encountered in Turkey today alongside some indication about linguistic vitality to the extent that such information is available. It goes without saying that it is of paramount importance to document the language(s) of any surviving speech communities; however, as noted by Chyet (1995, p. 223), these speech communities are often vulnerable, thus rendering contemporary field working on the contact conditions almost impossible.

- **Antioch Arabic:** (a) Hatay (still spoken); main confession Sunni, also Alawi and Greek Orthodox, (Arnold, 1998, 2006); (b) Çukurova (moribund), mainly Alawis (Procházka, 1999).
- **Anatolian Arabic:** spoken in Mardin, Siirt, Diyarbakir, Kozluk-Sason-Muş, main confession Sunni, also Alawi, few Greek Orthodox and Jews (Jastrow, 2006, p. 87, 2015; Akkuş, 2018).
- **Armenian** (Western Armenian standard and dialects): spoken in Istanbul, Izmir, all over Anatolia, especially Eastern Anatolia, mainly Armenian Orthodox (also Catholic, Protestants) (Adjarian, 1909, among others); There are also Turkish-speaking Armenian Orthodox communities in Çukurova, Kutahya, and several others (Balta and Ölmez, 2011).
- **Hemshinli Armenian** (Homshetsma Armenian dialect): low vitality variety spoken in the Black Sea and hinterland, mainly Muslims (Vaux, LaPorta and Tucker, 1996).
- **North Western Neo-Aramaic:** Tur Abdin (Turoyo), Diyarbakir (Mlahso, extinct), Jacobites (Jastrow, 1985; Poizat, 2016).
- **North-Eastern Neo-Aramaic:** Hertevin (also spoken in Iraq and Iran), Assyrian Orthodox, Chaldean, Jewish (Chyet, 1995; Coghill, 2018); Assyrians (Jastrow, 1988; Khan, 2007b; Napiórkowska, 2013).
- **Circassian** (Adyghe, also said Tcherkesse): spoken in Kaiseri, Kahramanmaraş Tokat, and urban centres, majorly Sunni (Dumézil, 1965)
- **Cappadocian Greek:** originally spoken in the area between the Ottoman cities of Nevşehir (Greek Νεάπολη), Kaiseri (Greek Καισάρεια) and Niğde (Greek Νίγηρη) (see Janse, 2004 et seq.; Dawkins, 1916; Karatsareas, 2011); Greek Orthodox (Catholic communities

also attested); in the same area, there were also Turkish-speaking Greek Orthodox communities (Karamanlidika) (see Balta and Ölmez, 2011). Currently, spoken in pockets in Northern Greece (moribund) by refugees following the Exchange of Populations between Greece and Turkey in 1923.

- Turkish-speaking **Greek** (Urum): was spoken in Kars, Erzurum, Trabzon, and Gümüşhane; Greek Orthodox (Catholic communities also attested); in the nineteenth century they moved to Georgia and since 1990s to Greece (but also Cyprus and Russia) (see Xanthopoulou-Kyriakou, 1991; Eloeva, 1998; Skopeteas, 2014 Höfler, 2016 and references therein).
- Romeica (Greek ρωμαίικα later called Istanbulite **Greek**): spoken in Istanbul (now has assimilated to Standard Modern Greek), Greek Orthodox (see Komondouros and McEntee-Atalians, 2007).
- Romeica (same autoglossonym as Istanbulite Greek which later came to be known as Pontic **Greek**): spoken in the Black Sea, Trabzon, Greek Orthodox (see Drettas, 1997, i.a.); spoken in Greece since the Exchange of Populations between Greece and Turkey in 1923 (low vitality).
- Romeyka (an archaic branch of Pontic **Greek**): still spoken in Çaykara, Sürmene and Tonya, but also by migrants in Istanbul and Germany, France and Belgium (see Mackridge, 1987; Sitaridou, 2013, 2014a, 2014b, 2016); Greek speakers were islamized in 16/17th C. and thus exempted from the Exchange of Populations between Greece and Turkey in 1923 (low vitality in Kara Deniz, severely endangered in Istanbul and elsewhere).
- Phrasiot **Greek**: was originally spoken in a region which comprises today the south, south-eastern part of the province of Kayseri, around the towns of Develi and Yahyalı, and the northern part of Adana, close to the borders with Kayseri. Spoken in West and Central Macedonia since the Exchange of Populations between Greece and Turkey in 1923 (low vitality) (Bağrıaçık, 2018).
- Judeo-**Spanish**: spoken in Istanbul, Izmir by Jews (low vitality) (Bornes-Varol, 2008).
- Northern **Kurdish** (Kurmanji, Zazaki): spoken in South-Eastern Anatolia, up to the Lake Van; several dialects, Muslim (Sunni and Alevi), also spoken by Yezidi communities (Haig, 2014b).
- Central **Kurdish** (Sorani): spoken in South-Eastern borders, Sunni Muslim (MacKenzie, 1961, 1962).
- **Laz**: still spoken in the Black Sea in Pazar, Arhavi by Muslim speakers (low vitality) (Öztürk and Pöchtrager, 2011; Lacroix, 2009).
- Domari (**Doms**), Lomavren (**Loms**) Romani (**Roms**): spoken in Southern Anatolia by Muslims and North-East Anatolia (especially Posha, with Armenian lexification) by Armenian Orthodox (Matras, 2002, 2012; Scala, 2014; Herin, 2016).
- **Turkish** (standard and dialects): spoken in the whole of Anatolia, by Sunni Muslims, also Alevi (Brendemoen, 2002, among others).

All languages of Anatolia, at least since the eleventh century AD, are in contact with Turkic, but have not always been in contact with each other. In Southern Anatolia, Kurdish and Neo-Aramaic overlap almost exactly (Chyet, 1995; Haig, 2001, 2017), while Arabic (Jastrow, 2006; Arnold, 1998; Akkuş, 2018) and Domari also overlap (Matras, 2002), though to a lesser extent. In Western Anatolia, the main locus of contact is the Black Sea Coast from Samsun to Artvin, which is also the extension zone of the Laz and which also hosts Lom. Neither Greek nor Laz seem to have been in contact with Kurdish dialects and/or Zazaki, which are contained south of the Kars-Erzurum-Erzincan line, or with neo-Aramaic, the latter to be found south of Van. Armenian is in contact with all the languages of Anatolia, except probably Circassian.

2. Critical issues and topics

The relevance of modern Anatolia for language contact and theoretical linguistics came to light since at least Asia Minor Greek’s description as a mixed language (Dawkins, 1916; Thomason and Kaufman, 1988, pp. 215–222; Poplack and Levey, 2010, p. 392). Two issues continue to dominate the discussion, namely: (1) what zones or subareas can be identified within Anatolia; and (2) whether modern Anatolia is a Sprachbund or not (Tzitzilis, 1989; to appear; Haig, 2014a) given its great typological heterogeneity and overall difficulty in identifying a common diasystem.

2.1 Candidate isoglosses or a common diasystem?

The conventional way to delimit Anatolia geographically is by reference to today’s Turkish borders. However, a wide range of languages spoken in Anatolia falls across a dialectal continuum which is of equal importance to the study of language contact in Anatolia (see Figure 21.1), and which, crucially, extends to countries outside Anatolia: Syria and Irak for Arabic (Anatolian Arabic is part of the Iraqi group and has cognates in Syria, see Akkus, 2018) and for Indo-Aryan languages (Domari and Romani); Syria, Irak, Iran, and Armenia for Neo-Aramaic and the Kurdish languages. Moreover, Armenian is vernacular in the Caucasus (Armenia, Georgia, Azerbaijandjan), Syria (Aleppo, Qamishli, Kessab, Damask), and Iran (Isfahan, Urmia, Tabriz). The boundary between Western and Eastern Armenian dialects roughly coincides with Turkey’s eastern border, which, importantly, has undergone several changes during modern history; crucially, several Western Armenian dialects of the Erzurum group are vernacular beyond this area (for instance, in Akhalkalak and Akhaltskha in Georgia; and in Gyumri in Armenia). Turkish is also spoken far beyond Anatolia’s borders (namely, in the Balkans, Caucasus, and Iran)—Eastern Anatolian Turkic dialects being very close to Azerbaijani. Greek is mainly spoken in Greece and Cyprus, with some islets in central Europe (e.g., Romania), but also in Turkey, in the Caucasus (Russia, Georgia) as well as in Ukraine, Armenia and Azerbaijandjan.

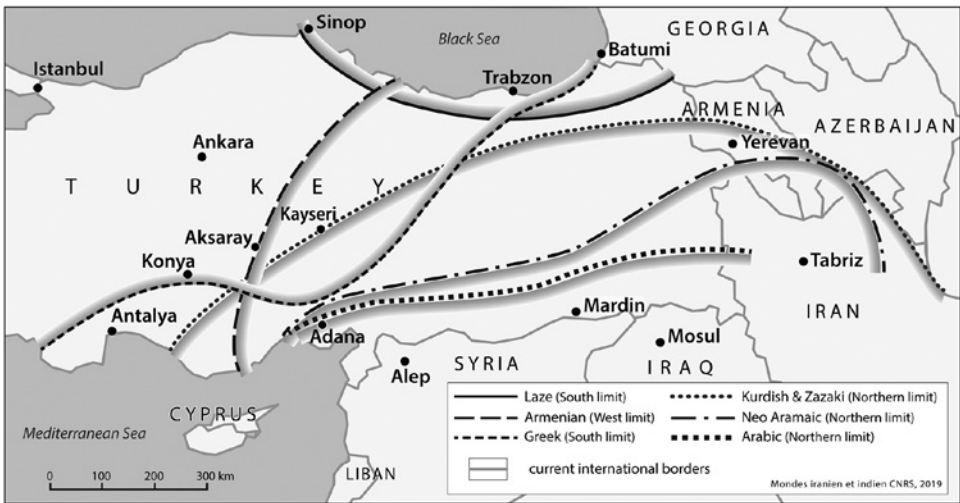


Figure 21.1 A schematic representation of the main dividing ethno-linguistic lines in Anatolia before the massive shifts of the early twentieth century

Overall, Anatolia's linguistic borders are far less obvious than the geographical ones seem to be. In fact, neighbouring linguistic areas (such as the Balkans, Caucasus, and Araxes area in Mesopotamia) all involve some of Anatolia's languages and display many similar features, a fact which has prompted several scholars to talk about the region in terms of transition zone(s) (see Haig, 2017 for making such a claim about Eastern Anatolia) or buffer zone (see Stilo, 2005, 2012 in reference to the Iranian languages), while Friedman (1996) points out common typological and areal features for the Balkans and the Caucasus. Considering the continuum mentioned earlier, another option is to consider Anatolia as a part of a larger Western Asian area which includes the Middle East and the Caucasus (Haig and Khan, 2018).

Thus, in the literature there is a seemingly contradictory double-pursuit: on the one hand, to discover the isoglosses that divide Anatolia into different zones, and the quest of pan-Anatolian features, on the other. It follows that the more difficult it is to find (coinciding) isoglosses, the more ground there is to claim that languages in Anatolia have evolved into a common diasystem.

2.2 *The quest for pan-Anatolian features*

Any proposal for the existence of an Anatolian Sprachbund is tantamount to a definition of a Sprachbund. Crucially, if we define the latter as (a) a contiguous linguistic area where 'three or more languages share structural features as a result of contact rather than as a result of accident or inheritance from a common ancestor' (Thomason, 2001, p. 99) it is not the same as (b) a convergent development from separate languages leading towards a common diasystem, which, ultimately, may lead to full isogrammatism (Tzitzilis, to appear). The latter, needless to say, is harder to prove, thus more controversial.

As a first step, commonly shared features have to be identified to be subsequently tested for pan-Anatolian status (which if we were to follow definition (a) they cannot be the result of inheritance). To this end, Matras (2009, p. 270; see also Haig, 2001; Matras, 2000; Chyet, 1995) suggests an inventory of 12 candidate features, as shown in (1):

1. Pan-Anatolian features by Matras (2000, 2009, p. 270, 2010):

- (i) uvular /q/
- (ii) /ɑ:/æ/ opposition
- (iii) pharyngeals in Arabic loans
- (iv) echo expressions
- (v) verbal morphology
- (vi) structural and phraseological calques such as finite subjunctives in modal complements
- (vii) enclitic copula
- (viii) postposed particle 'too'
- (ix) 'either . . . or' constructions
- (x) light verbs like 'to do' and copulas like 'to become'
- (xi) focus and modal particles
- (xii) *ki* subordinator

However, Haig (2014a, pp. 14 sq. and 28) disputes the relevance of shared verbal morphology as a key pan-Anatolian feature (as originally proposed by Matras, 2010, p. 75), tagging it rather as one of the distinctive features of the 'Mesopotamian sphere' of Anatolia (alongside

pronominal clitics cross-referencing of verbal arguments, and grammatical gender on nouns), as opposed to the ‘Caucasian-Caspian sphere’ (which is characterized by recipient arguments flagged through postpositions or case-suffixes and the experiencers of verbs of desire/volition expressed as canonical subjects).

Other features which have claimed to be pan-Anatolian are verb final order (Haig, 2001, p. 199) – which we discuss at length in Section 4 – mirativity/evidentiality (Johanson and Utas, 2000) and differential object marking (Janse, 2004; Key, 2012). The latter two however, are disputed because differential object marking (DOM) is widespread in various language families outside Anatolia (such as Romance, Slavic, Iranian, Caucasian) and mirativity is represented in a large belt ranging from Balkans to Himalaya. More recently, Haig (2017, p. 418) claimed that only ‘a few candidates remain that could be considered shared features of a pan-Anatolian region,’ as in (2):

2. Pan-Anatolian features (Haig (2001, 2014a, 2017, p. 418):

- (i) the enclitic recalled-topic marker
- (ii) an obligatory clause-final copula (similar to (1vii))
- (iii) ‘either . . . or’ constructions based on ‘ya(n) . . . ya(n)’ (see (1ix))
- (iv) a grammaticalized indefinite article, accompanying indefinite singular NPs
- (v) echoic reduplication (see (1iv) as in;)
- (vi) the use of a general complementizer *ki* (with variant vowel values) (see (1xii))

Nevertheless, empirical refutation of (2) is still easy to come across. For instance: (i) is not found in Cappadocian; (iv) and (vi) are not found in Romeyka. As for (ii), Haig (2017, p. 406) suggests that the medial copula pattern described for Siirt NENA (Jastrow, 1980, p. 148) could have spread to other varieties, especially among Semitic languages of Anatolia, which can be rather ascribed to a fadeout phenomenon. Overall, the features in (2) are more convincing as Eastern Anatolian features than as pan-Anatolian ones.

2.3 Zoning attempts

Controversies such as the ones discussed in Section 2.2 led researchers to reject the claim for a pan-Anatolian linguistic area and, make, instead, several attempts at zoning. Haig (2014a, p. 29) proposes a significant number of sub-areas or clusters within Anatolia, defined in terms of epicentres of contact and gradual fadeout areas. Stilo (2012, 2015) proposes areas of overlap exhibiting mixed features. Haig (2017), instead, puts forward the proposal that there are two areal epicentres, namely the south-eastern Kurdish/NENA/Arabic ‘Mesopotamian’ region and the northern Turkic/Kartvelian/Armenian ‘Caspian/Caucasian’ region, with Eastern Anatolia as an intermediate zone.

A survey of the literature can maximally yield four zones, as shown here and in Figure 21.2:

- (a) **Western Anatolia**, historically dominated by Asia Minor Greek, is a buffer zone between Southern and Eastern Anatolian regions, the Balkan Sprachbund (with Istanbul and Eastern Black Sea epicentres) and the Caucasian Sprachbund (see Tzitzilis, to appear; Sitaridou, 2016; Haig, 2017).
- (b) **Eastern Anatolia** is a buffer zone between the Caspian/Caucasian regions and Mesopotamia (Haig, 2017). Isoglosses proposed by Haig—related to a reduced sample of 10 languages/dialects represented in the area—are argued to show a clear division between Central and Northern Kurdish, Arabic, Nena, Domari on one hand; Turkish, Homshetsma Armenian, Laz, Eastern Armenian on the other, with Zazaki in the middle of the zone.

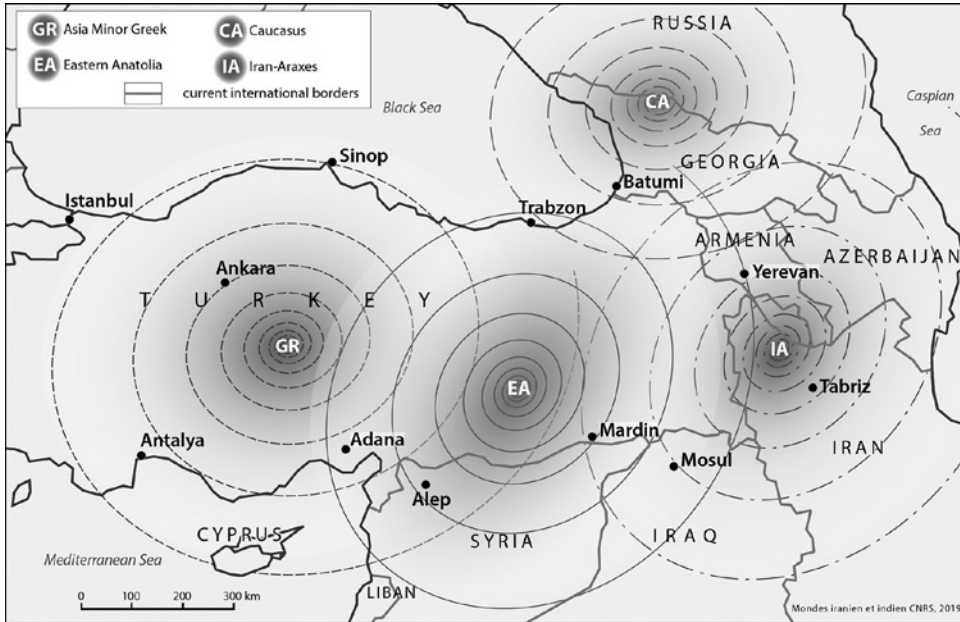


Figure 21.2 Four zones with epicentres and fade-outs

However, it is important to clarify that in Eastern Anatolia, in its current conceptualization by Haig, Pontic Greek and Western Armenian are totally absent; thus it is uncertain as to whether they belong to Eastern Anatolia or Western Anatolia on his account. Tzitzilis (to appear) assumes that Pontus is Eastern Anatolia and that Pontic Greek is in contact with Eastern Armenian only; however, even if there is fade-out contact with Eastern Armenian, the contact is mainly with Western Armenian.

- (c) **Southern Anatolia** is a buffer zone between the Levant (e.g., Cyprus and Syria/Lebanon), Western and Eastern Anatolia regions, and Mesopotamia. It is a zone of intense contact between the Semitic languages (NENA and Anatolian Arabic) and the Iranian languages (Kurdish, Zazaki) with variable impact of Turkish, and locally of Armenian as well (Haig and Khan, 2018, p. 40).
- (d) **Araxes-Iran** is a buffer zone between the Caucasian region, Eastern Anatolia, and Iran (Stilo, 2012). It includes the fringes of Eastern Anatolia and Southern Anatolia, outgrowing towards Iran and the Caucasus. The claimed focal point, namely Araxes is rather a line, namely the border river, which separates the former Soviet countries of the South Caucasus from Turkey and Iran.

3. Current contributions and research

In our pursuit, there are some important methodological issues to take into account: (1) the description and documentation of the varieties still spoken in Anatolia is rather limited; (2) the historical record (or the lack thereof) does not always allow essential comparisons; or, when it does exist it is often ignored (see Section 3.2); (3) existing attempts to account for separate linguistic areas in Anatolia, albeit notable, have been biased, most of the time, by an unbalanced account of the different languages present in the region. For instance, there is an important

critical mass of works on the interplay between Semitic, Iranian, and Turkic (see Haig and Khan, 2018; Johanson and Utas, 2000; Johanson and Bulut, 2006), with a focus on Eastern Anatolia and its south-eastern margins (Northern Iran, Mesopotamia and at a less extent, Caucasus), but the same cannot be said about the role of Armenian, despite being spoken in a very expansive Eastern terrain and has been under extensive contact with Zazaki, Kurmanji, Neo Aramaic, and Turkish dialects (but see Haig, 2017 for a mention of Hemshin).

Nevertheless, some languages in the area have indeed benefited from large-scale dialectological documentation. Even if data processing methods vary, and despite the lack of interoperability between existing databases (Manchester Database of Kurdish dialects; The North Eastern Neo-Aramaic Database Project in Cambridge; the Atlas of Armenian dialects currently developed in Armenia at NASRA; etc.), current scholarship allows to sketch several isoglosses in Anatolia. In what follows, we will concentrate on phonetics/phonology, morphology, and syntax, leaving aside the lexicon and phraseological calques, which are anyway expected to occur to a significant degree due to the fact that the area is an ethno-cultural melting pot.

3.1. *Phonetic and phonological isoglosses*

Language contact in Anatolia has affected the phonetics of most languages. Vocalism is less challenging, probably because the vowel repertoires are somewhat convergent in the Semitic and Iranian languages of the region (Haig, 2017, p. 401); hence, it is only certain vowels that deserve particular attention.

To start with, the extension of the back-central unrounded vowel [u] – and possibly the development of [æ] and [œ] too – under the influence of Turkish is reported for languages which do not have an inherited schwa such as Asia Minor Greek (in particular, as a result of vowel harmony, see Dawkins (1916, pp. 43, 67) and Oeconomides (1908, p. 2) although the topic remains understudied). The extension of front rounded vowels [y] and [ø] and unrounded [a], which sometimes affects, to varying degrees, both loanwords and inherited lexicon, is also attributed to the influence of Turkish, the only language of the area which has both front rounded vowels and unrounded [a] (Haig and Khan, 2018, p. 13).

As for consonants, alveolar, and post-alveolar stops [t], [ʒ] and affricates [tʃ], [dʒ], which are unknown in mainland varieties of Greek, are considered one of the main distinctive characteristics of Asia Minor Greek (Dawkins, 1916, pp. 74, 80, 86). In the case of affricates, in particular, their emergence could have been facilitated by the lack of phonemic opposition, in both Greek and Turkish, between alveolar and post-alveolar affricates, namely /ts/≠/tʃ/, /dz/≠/dʒ/ (feature *-/ts/≠/tʃ/*) (see Stilo, 2007; Haig, 2014a, 2017). Like most of the language families in the area, they display, instead, a single realization, which may be (with possible dialectal shifts) either post-alveolar (Asia Minor Greek and Arabic) or alveolar (Iranian, Turkic, and Aramaic). The opposite feature (*+/ts/≠/tʃ/*) concerns two groups, namely Armenian and Kartvelian languages, which are geographically contiguous, yet phylogenetically distinct, and have been in long-standing contact (earlier than their first attestations from the fifth century AD), but also some non-related languages spoken in overlapping areas, like Pontic Greek (Drettas, 1997). Given that this opposition, namely *+/ts/≠/tʃ/*, is restricted to Armenian and Kartvelian, it may be labelled as a feature characterizing the ‘Caucasian’ border of Anatolia (North-East), despite the fact that the extension of Armenian towards central Anatolia was much pervasive between the thirteenth and nineteenth centuries. Lastly, Zazaki (Paul, 1998) and Kurmanji (Haig and Khan, 2018, p. 113) display the opposition for affricates only.

The phonemic distinction between one voiced (i.e., /b/) and two unvoiced series (usually one aspirated i.e., /p^h/ and one more or less ejective or glottalized /pʼ/) – present in Kartvelian,

Eastern dialects of Armenian, Pontic Greek, Kurmanji and Zazaki – is generally considered a Caucasian feature in the context of the East Anatolia transition zone (Haig, 2017, p. 402), whereas Western languages display a two-way opposition, often with aspiration of unvoiced stops (e.g., Asia Minor Greek, Western Armenian, Turkish). The three-way distinction is, to some extent, in complementary distribution with the Mesopotamian pharyngealized stops [ʕ] since languages that display the latter (e.g., Sorani, Arabic, and Aramaic) tend to lack the former, and vice versa. However, the opposition is not absolute. On one hand, Zazaki and Kurmanji represent an overlapping zone where both features are displayed. Asia Minor Greek, Turkish and Western Armenian, on the other, lack both. The overall picture, which shows two pairs of isoglosses defining four sub-areas, is shown in Table 21.1.

Phonemic /f/ (/f/≠p/ and /f/≠v/) (see Stilo, 2015) is found in most Anatolian languages, including Asia Minor Greek, Turkic, and Iranian languages, thus forming a large belt from West to East. In the North-East, Georgian, which has one of the largest phonemic inventories of Anatolia, lacks /f/, while Laz uses it only in loanwords from Turkish (Lacroix, 2009). Armenian, which initially transcribed Greek /f/ as /p^h/, eventually acquired/nativized /f/ (and created the corresponding grapheme) by the twelfth century AD after intensive contact with Frankish Crusaders; this development makes rather difficult to consider Stilo's (2015, p. 351) classification of Armenian among languages with ' /f/ in loans only ' as valid given the different time-frames. Having said this, a few Armenian dialects still lack it: for example, in Karabagh dialect, /f/=p^h/ is still found in loanwords: Russ. *futljar* 'case' > [p^hutljar], in contrast to neighbouring Azerbaijani (which provided a shibboleth during the war in the late 1980s). As for the Southern belt, historically, Semitic displays an inherited /f/, but in a very specific system where some oppositions lack (/f/ = /v/, /p/ = /b/, /w/ = /v/), while there is a distinct /θ/ phoneme. The effect of contact is thus in the reorganization of the system of phonemic oppositions around /f/. In Anatolian Arabic (Akkus, 2018, p. 457), in some cases such as in Tillo/Siirt (Southern-Turkey), /f/ results from the shift of Old Arabic interdental /θ/, which either became sibilant /s/ (Sason), dental stop /t/ (Diyarbakir), or remained /θ/ (Mardin). In Western Neo-Aramaic, too, contact-induced reorganization of phonological systems is also observed, namely a parallel shift affecting p>f and b>p (Arnold, 2000, p. 352). It seems that the most fine-grained phonological distinctions are observed in the southernmost of the NENA dialects, in Iraq, i.e., Amedya (Greenblatt, 2011, p. 13) and Barwar (Khan, 2008, p. 29), which display the full range of distinctions, namely /p/≠p^h/≠/f/≠/v/≠/w/, although such claims have to be treated with caution given that several authors mention that some phonemes are used only for borrowings (i.e., see Kim, 2010, p. 230, with regard to Turoyo in Southern Turkey, who mentions that only /b/, /f/ and /w/ are inherited, while /p/ and /v/ are being used only for loanwords) and thus the data are ultimately not comparable.

The three-way isogloss resulting from this reorganization, +/f/, -/f/ and /f/ in loans only (Stilo, 2015, p. 351), gives rise to a +/f/ belt from West (Turkic, Greek) to East (Iranian), including Armenian, and a North-Eastern (Caucasian) and Southern (Semitic) fringe with a fade-out (which comprises Georgian and some varieties of Aramaic which are -/f/, Laz, Udi

Table 21.1 Pairs of phonemic isoglosses defining Anatolian sub-areas

| Areas | /p ^h /≠/p'/ | [ʕ] |
|--------------------------------------|------------------------|-----|
| Caucasian belt | + | - |
| Southern (Mesopotamian) belt | - | + |
| Eastern Anatolia (transitional area) | + | + |
| Western Anatolia | - | - |

and some varieties of Aramaic which are /ff/ in loans only). Nevertheless, the fact that some North Caucasian languages beyond Anatolia, for example Abkhaz, display +/ff/ (as well as /p/, /pʰ/, /v/, /w/) is challenging for this classification. Moreover, regarding the Southern belt, considering the complexity of interactions with Kurdish and Iranian languages on the one hand, and with different varieties of Arabic on the other, the three-way isogloss does not seem capable of reflecting the situation unless a phonemic account is given for the entire set of oppositions involving bilabial and labio-dental stops (/p/, /pʰ/, /pʷ/, /f/, /v/, /w/) and even interdental (/θ/).

Final stress (including the possibility of post-stress unaccented suffixes, e.g., the definite article in Armenian, some inflexional suffixes in Kurmanji) is widely attested across Anatolia, with, however, a certain degree of variation: for instance, final stress is the rule for Turkic-speaking and for overlapping Eastern and South-Eastern areas (Kartvelian, Armenian, Iranian including Kurdish dialects), but not for Western (i.e., Greek) and Southern Anatolia (Levant and Mesopotamia, in this case mainly Semitic languages). However, some contact-induced exceptions blur the phylogenetically driven isoglosses: for instance among Neo-Aramaic dialects, which are surrounded by stress-final languages (Kurdish, Turkish, and Modern Persian), the most innovative ones (mainly the Jewish dialects) shifted to final stress, while the more conservative Christian dialects did not (Napiórkowska, 2013, p. 93).

3.2 Morphological isoglosses

One morphological isogloss which could potentially constitute a pan-Anatolian feature stemming from Turkish is echoic *m*-reduplication (or ‘fixed-segment’ reduplication, see Matras, 2009; Haig, 2001), mostly used to refer to a whole conceptual domain or to recall any word from the context with scepticism or irony (‘X and so on,’ ‘allegedly X,’ ‘such things like X’), as in: Arm. *tun* ‘house’ *tun-mun* ‘house and related things,’ ‘house and other belongings,’ Tk. *beyaz* ‘white’ *beyaz-meyaz* ‘white and such things,’ ‘allegedly white,’ Kurm. *xwarin* ‘to eat’ *xwarin-marin* ‘eating and such things,’ Laz *dadzi* ‘thorn’ *dadzi-madzi* ‘thorns and so on’ (Lacroix, 2009, p. 123); also reported in Cappadocian as contact-induced from Armenian (Bağriaçık, 2016); and even outside Anatolia, in Persian and Abkhaz.

Echoic *m*-reduplication is slightly different from partial leftward (‘quasi-fixed segmentism’) emphatic reduplication of adjectives (Donabedian, 2018, p. 104; Demir, 2018, among others): Tk. *çiplak* ‘naked’ *çip-çiplak* ‘all naked,’ Capp. *mavro* ‘black’ *mas-mavro* ‘pitch black’ (Bağriaçık and Janse, 2016, p. 201), Arm. *sev* ‘black’ *sep-sev* ‘pitch black,’ etc. According to Bağriaçık and Janse (2016, p. 205), in Turkish, the choice of consonant in the reduplicated syllable could be determined by morphology-lexicon, while in Armenian and Cappadocian, it is determined by the phonetic value of the adjacent consonant, which suggests that this diffusion is a case of ‘selective copying’ (Johanson, 2002). Importantly, reduplication phenomena are extended over the whole of Anatolia, and thus may rather be pan-Anatolian features; therefore, dividing isoglosses seem not to reflect contemporary distribution but rather correspond to previous generalizations about grammatical borrowing (Matras, 2011; Johanson, 2002, among others).

An example of diffusion induced by convergent grammaticalization paths, which is also a candidate for pan-Anatolian feature status, is the formation of indicative verbs with a specific prefix for progressive-indicative, originating from a location or equative marker (Matras, 2009, p. 260, 2010, p. 75), as opposed to a subjunctive marked by an alternative prefix or the lack of a prefix. Matras shows that the present indicative forms in Turoyo Aramaic, Kurmanji Kurdish, Persian, Western Armenian and Levantine Arabic follow a common scheme: indicative prefix (respectively, *ko-*, *di-*, *mī-*, *gə-*, *ba-*) + root + person marking (clitic or inflectional).

However, Haig (2014a, pp. 20–22) argues that this feature while present in Persian and Levantine Arabic, is absent in Turkish and Zazaki and irregularly found in Neo-Aramaic dialects; furthermore, it is also absent in Asia Minor Greek as well hence it is not a plausible pan-Anatolian candidate. Representing Eastern Anatolia as a transitional zone binding a Mesopotamian and a Caucasian-Caspian sub-sphere, he shows that the prefixed indicative schema is characteristic of the former, but not displayed in the latter. This is consistent with one of the isoglosses ascribed to the Araxes-Iran Area by Stilo (#12), concerning periphrastic formation of indicative tenses. This situation is perfectly illustrated by Armenian dialects. On the one hand, as recalled by Matras, the Western Armenian indicative present and imperfect (*kə tesn-em* ‘I see, I am seeing’; *kə tesn-ei* ‘I was seeing’) are formed with a proclitic particle *kə* [gə], originating from the defective verb *kam* ‘there is,’ as opposed to the subjunctive forms without particle (*tesn-em*, *tesn-ei*); on the other, EArm displays a periphrastic present/imperfect with *-um* converb, consistent with Stilo’s isogloss and Haig’s assumptions concerning a Caucasian-Caspian trend (*tesn-um em* ‘I see, I am seeing,’ *tesn-um ei* ‘I was seeing,’ with subjunctives identical to WArm). However, there are a couple of challenging asymmetries: first, the *-um* converb displays the specifically EArm locative morpheme (*-um*), which overlaps partially with the proposed Eastern Anatolian isogloss; second, in WArm, *kə*, due to its veridical function, is not compatible with negation, which triggers negation to be rendered periphrastically with the verb ‘to be’: *č-em tesn-er NEG=AUX.ISG see-CVB*, a schema originating from Middle Armenian *č-em i tesn-el, NEG=AUX.ISG LOC see-INF*, which involves the old locative preposition. This case demonstrates how one feature (namely, preposed particle originating from locative-existential) proposed as pan-Anatolian and one feature considered a dividing isogloss (namely, periphrastic imperfect tenses) are interlinked, and can therefore be treated in different ways: either, as Haig (2017) does, by considering Eastern Anatolia a transitional zone, and in this case WArm data and EArm data act as a confirmation of this hypothesis, since the dividing line between WArm dialects and EArm dialects is situated in this area (with a fadeout belt defined by a Rize-Van western line and an Akhalkalaki-Guymri-Nakhitchevan eastern line); or, to achieve a more fine-grained resolution of contact, by disentangling the variable ‘prefix’ and the variable ‘locative-existential,’ and adding a third isogloss relating to the affirmative/negative asymmetry (see for example Jastrow, 1988, p. 55 in connection with a reverse asymmetry in Hertevin Neo-Aramaic, where only the negative form is prefixed) – the latter also seems to be relevant in Romeyka (Sitaridou, 2016).

The formation of the comparative/superlative, which could also potentially be included as a feature of an emergent pan-Anatolian diasystem, is essentially an analytic construction with ablative or a ‘source’ adposition or case marker, with comparative marking on the adjective as optional (lit. ‘he is big(-ger) from-me’); in particular, either an ablative case suffix as in Turkish (*-den/-dan/-ten/-tan*), Laz (*-šen*, Lacroix, 2009, p. 93) and Armenian (WArm *-e*, EArm *-ic*) ex. *injm-e mej e* 1SG-ABL big COP.3SG; or a preposition with a local meaning ‘from,’ e.g., *ji* in Kurmanji, *mən* (or variants thereof) in NENA, Arabic (Haig, 2017), and Pontic Greek (*atos as emen tranos en* he.nom from I.ACC big be.3SG) (Tzitzilis, to appear).

Regarding morphological typology, while only a few languages of the area (Turkish, Kartvelian) are genetically agglutinative, we see a change towards agglutination in genetically fusional languages. The typical example is Cappadocian (Janse, 2004), in which Greek imposes fusional morphology whereas Turkish imposes agglutinative morphology, resulting in (3) whereby Cappadocian uses agglutinative morphology parallel to the Turkish agglutination: the endings *-ju* and *-jv* function like the Turkish agglutinative suffixes *-in* and *-lær*, respectively, displaying a transparent relation between form and grammatical function which

dialects of the region rather suggests a fade-out phenomenon. Yet the landscape displays several challenging paradoxes. First, a gender system ‘could be more deeply entrenched’ in some North-Western dialects such as Maraş, Malatya, and Antep (Haig, 2004), than in some Kurmanji dialects spoken more in the South. The second paradox is the apparent mirror situation of Southern Kurdish, which lacks gender despite being in deeper contact with Arabic, at least in its Western part, and Northern Kurdish, which has gender, despite being closer to Turkish and Armenian. As for possible explanations, sociolinguistic parameters and the history of settlements in the region could help us to sketch some diachronic paths (as suggested in Haig and Khan, 2018, p. 108, referring to Dahl (2004) and Trudgill’s (2011) generalizations about ‘mature’ systems).

Let us now turn our attention to case marking and pronominal clitics. The latter is one of the important dividing lines of the region despite being found in areas beyond Anatolia and displaying geographical epicentres and gradual fade-out (Haig, 2014a, p. 27). The existence of clitic pronouns marking arguments of finite verbs is well known in Iranian and Semitic languages, but limited to the Southern and South-Eastern part of Anatolia (Neo Aramaic and Sorani), and does not concern Kurmanji, Zaza, Armenian, Turkish, or Greek. Laz also displays inherited argument marking clitics in its verbal morphology, thus not diverging from other Kartvelian languages.

Several correlations have been suggested in the literature: Jügel and Samvelian (2016) highlight a complimentary distribution between pronominal clitics (Sorani) and case marking (Kurmanji). In this case, as clitics are inherited in Sorani (Haig and Öpengin, 2018, p. 162), one can consider their loss in Kurmanji as the result of contact, probably due to convergence with Armenian (Haig and Öpengin, 2018, p. 163). In any case, it seems that the diffusion of pronominal clitics is clearly a one-way process, allowing maintenance or loss, but not spread towards languages which do not exhibit such a marking at all. However, possessive articles used as object clitics in some Armenian dialects spoken in Iran, such as Khoy, suggest that exceptions are possible (see Martirosyan in Haig and Khan, 2018, p. 85). At most, one could consider some phenomena as a ‘fadeout’ manifestation. One such fadeout could well be the enclitization of clitic pronouns in Pontic Greek (Chatzikyriakidis, 2012) (but see Sitaridou, 2014a for a different possibility, namely reanalysis of Hellenistic Greek postverbal strong pronouns). Note as well Stilo’s (2015) isogloss ‘Possessive pronoun is an oblique form encliticised to noun’ (Araxe#10) highlighting the fact that some sets of clitic pronouns attach both to verbs as object indices and to nouns as possessors.

Restricted or absent agreement inside the NP, which is typologically related to agglutination as discussed previously, is also connected to the status of bare nouns. Many languages of the area allow bare count nouns in argument position depending on information structure (more likely for bounded preverbal objects than for topical subjects, see Donabedian, 2010a, 2010b). One of the correlates of the availability of bare nouns is differential argument marking. The phenomenon, which is particularly well-studied in Greek, has been assumed to be pan-Anatolian (Janse, 2004; Key, 2012), although it is in fact quite widespread in the languages of the world, and especially in the neighbouring areas (Balkans, Caucasus, Iran, Russia). In Cappadocian, it concerns the indefinite object (DOM) (Janse, 2004) (4), and in Pontic Greek there are even reported cases of differential subject marking (DSM) (5):

4. a. ama tranižne to liko (Axos; Cappadocian)
 when see.3PL the.ACC wolf
 ‘as soon as they see the wolf’
 (Axos; Dawkins, 1916, p. 396)

Armenian definite article *-n* which is also a central unrounded *-ə*. The material convergence is particularly troubling in the possessive construction:

6. a. *çocuğ-un kitab-i* (Turkish)
 child-3SG.GEN book-EZ
 b. *yerexay-i-n girik'-ə* (Western Armenian)
 child-GEN-DEF book-DEF
 'the book of the child'

This 'echo' effect results from the double material affinity of *Tk ezafə* and *Arm article* on the one hand, and *Tk genitive* and *Arm genitive+article* on the other, despite the fact that these examples are subject to different morphosyntactic analysis. Anyway, such convergences open doors for reanalysis of markers and deeper convergence mechanisms, including both material similarity and category isomorphism.

Overall, category isomorphism as an aspect of language contact in Anatolia, should be better understood through bringing together the in-depth study of categories which are not assumed to be identical, but deal with similar linguistic operations or domains (this would improve the achievements of linguistics typology as well).

3.3 Syntactic isoglosses (with focus on Western Anatolia)

A wide range of phenomena has been suggested in the literature as potential syntactic isoglosses demarcating different language areas within Anatolia. In what follows we survey the most important such phenomena encountered in Western Anatolia – the least discussed in the literature – leaving aside the discussion of basic constituent word order and head-finality for Section 4.

First, let us consider complementation patterns, which separate (a) languages in which the verb 'want' precedes the subordinate verb (7a) versus languages in which the verb 'want' follows the subordinate verb (7b) often with substantial intra-linguistic variation (7a–7c); (b) languages in which modals select for infinitives (8a/b/c) versus languages in which the subjunctive is selected instead (8d/e) although there is also significant intra-linguistic variation (8a) despite some being syntactically conditioned (as is the case in Romyeika, in which a negated present tense modal selects a subjunctive complement (8b) whereas a negated past tense modal selects an infinitive (8c) (see Sitaridou 2014a/b); (c) languages with nominalization (9a/b/c/d) versus languages without (9e); (d) languages with final complementizers (10a/b) and languages without (10c/d). Interestingly, phenomena (a), (c), (d) have also been associated with head-final languages, an issue to which we return in Section 4.

7. a. *Ara-n g'uz-e dun mē šinel* (Standard Western Armenian)
 Ara.NOM-DEF want-3SG house INDEF build.INF
 'Ara wants to build a house'
 b. *Amedi cari-şe na mo-xt'-a-s-ere* (Laz)
 Ahmet.NOM dinner-ALL SUB PV-come-SUBJ-PRS.3SG-AUX
 b-gor-um-Ø.
 I.SBJ-want-TS-PRS.1SG
 'I want Ahmet to come to dinner'
 c. *Ara-n dun mē šinel g'uz-e* (Standard Western Armenian)
 Ara.NOM-DEF house INDEF build.INF want-3SG
 'Ara wants to build a house' (the subordinate verb is in focus position)

8. a. Bedk e kam (Standard Western Armenian: in some cases only)
 need be.3SG come.SUBJ.1SG
 ‘I have to come’
- b. Tši 0elo na porpato (Romeyka and all other AMG)
 not want.1SG SUB walk.1SG
 ‘I don’t want to walk’
- c. utš eporesa almeksini (Romeyka only)
 not could.PST.1SG milk.INF
 ‘I couldn’t milk’ (Sitaridou 2014b, p. 37)
- d. Amedi oxori-s ort’-a-s-ere (Laz)
 Ahmet.NOM house-DAT COP-SUBJ-PRS.3SG-AUX
 ‘Ahmet must/should/ . . . be at home’
 (Öztürk and Pöchtrager, 2011, p. 84)
- e. Grévu na ipáu (Pharasiot)
 want.IPFV.NPST.1SG SUB go.PFV.NPST.1SG
 ‘I want to go.’
 (Bağrıaçık, 2018, p. 108)
9. a. Ara-i-n dun šinel-ë des-a?r (Standard Western Armenian)
 Ara-GEN-DEF house build.INF-DEF saw-2SG?
 ‘Did you see Ara building a house?/Did you see that Ara built a house?’
- b. ap ađa so spitin ts Aişes to panimon 0elo (Romeyka)
 from here to.the house the Ayşe.GEN the going want.PRS.1SG
 ‘I want Ayşe to make her way from here to the house.’
- c. Amedi-şi cari-şe mo-lva-muşi b-gor-um-Ø. (Laz)
 Ahmet-GEN dinner-ALL PV-come-POSS.3SG 1SBJ-want-TS-PRS.1SG
 ‘I want Ahmet to come to dinner’ (Lit.: I want Ahmet’s coming to dinner’)
 (Öztürk and Pöchtrager, 2011, p. 79)
- d. Ayşe-nin bu ev-den git-me-sin-ı ist-iyor-um (Trabzon Turkish)
 Ayşe-GEN this house-ABL go-VN-POSS-ACC want.IMPF-1SG
 ‘I want Ayşe to leave this house.’ (lit. ‘I want Ayşe’s going (away) from this
- e. Dóka émri na ipámi. (Pharasiot)
 give.IPFV.PST.1SG order.N.NOM.SG SUB go.PFV.NPST.1PL
 ‘I ordered (that) we leave.’ (Bağrıaçık, 2018, p. 127)
10. a. Uz-es ne (informal Western Armenian)
 want-2SG if
 ‘If you want’
 (Donabedian, 2018, p. 128)
- a’. yet’e uz-es (Standard Western Armenian)
 if want-2SG
- b. 0aripsin cimati dei (Cappadocia; Malakopi)
 think.PST.3SG sleep.3SG COMP
 ‘He thought that she was asleep.’
 (Dawkins, 1916, p. 404)
- c. Hini-k do-p-xed-a-Ø -t ya do t’k’v-es (Laz)
 they-ERG PV-1SBJ-sit-SUBJ-PRS.1-PL SUB say-PST.3PL
 ‘They said (that) let us sit, too.’
 (Öztürk and Pöchtrager, 2011, p. 77)

- d. Enespala na leyo ti mami ta xaberæ. (Romeyka)
 forget.PST.1SG SUB say.1SG the.ACC grandmother.ACC the news.ACC
 ‘I forgot to tell the news to the grandmother.’
 (Sitaridou, 2014a, p. 125)

Second, let us consider the left periphery in Western Anatolia; in particular, we note (1) the availability of multiple *wh*-fronting (MWF), which is found in all the languages of the area (11); crucially, the syntax of MWF cannot be taken to be identical to the Turkish MWF (see Michelioudakis and Sitaridou, 2016, 2019 for a detailed discussion of MWF in Romeyka and Turkish); and (2) the emergence of discourse particles (with different pragmatic values ranging from topicalization marker to contrastive topic marker to just contrastive marker), which is attested in almost all the languages of the region, as shown in Table 21.2. They are often the result of reanalysis of adverbs, as is the case for Pontic Greek *pa*, which originates in *palin* ‘again’ and which currently functions as a contrastive topic marker (see Sitaridou and Kaltsa, 2014).

11. a. O?v i?nč ěs-av (Western Armenian)
 who what said-PST.3SG
 ‘Who said what?’
- b. Pios tinan ayapai? (Romeyka)
 who.NOM who.ACC.HUM love.3SG
 ‘Who loves whom?’
 (Michelioudakis and Sitaridou, 2016 p. 4)
- c. Mík muya asere? (Pazar Laz)
 who what make?
 ‘Who will do what?’
 (Öztürk and Pöchtrager, 2011, p. 156)

Importantly, syntactic isoglosses discussed in this section are not restricted to Western Anatolia; in fact, some have been put forward as potential pan-Anatolian features (see Haig (2017) for the position of the complement of the verb ‘want’ as a relevant isogloss for Eastern Anatolian and Stilo (2005, 2012) for Iran-Araxes). Thus, once again, it is a blurred picture as to what counts as an isogloss and what counts as a pan-Anatolian feature.

4. Methodological and theoretical issues

In this section we assess the claim whereby word order and other related syntactic properties, as briefly mentioned in Section 3.3, constitute a common diasystem or not. In doing so, we point out various methodological, empirical and theoretical issues as to why this does not seem to be the case or at least cannot be taken for what it seems to be.

Table 21.2 Discourse particles in Western Anatolian languages

| | |
|--------------|--------------------------|
| <i>se</i> | Aramaic/NENA |
| <i>el/al</i> | Eastern/Western Armenian |
| <i>pa</i> | Pontic Greek |
| <i>ti</i> | Laz |
| <i>da/de</i> | Trabzon Turkish |

4.1 *A pan-Anatolian word order?*

Possibly the most discussed topic in the literature is word order. In fact, its consideration has reignited the discussion about a new Anatolian Sprachbund (see Section 1) because of a seemingly emerging isogrammatism. Indeed, *prima facie*, such a claim seems to hold true for the languages of Araxes-Iran, as argued by Stilo, 2015 – consider (12):

12. Apparent isogrammatism in a range of Araxes-Iran languages (a. Georgian, b. Colloquial Armenian, c. Colloquial Azerbaijani, d. Northern Talyshi):
- | | | | | | | |
|----|-----|-------------|--------|---------|----------|--------------|
| a. | šen | ginda | sami | (chali) | vašli | ø-iqid-o |
| | | | | | | |
| b. | du | uzumes | yerekh | haṭ | xənzor | ø-añ-es |
| | | | | | | |
| c. | sæn | istirsæn | üç | dana | alma | al-a-san |
| | | | | | | |
| d. | tī | peday | se | gīla | sef | bī-san-iš |
| | you | want.PRS.2s | three | grain | apple.SG | buy-SUBJ-2SG |
- ‘You want to buy three apples’
(Stilo, 2015, p. 345)

Moreover, word order isogrammatism may even seem to hold true of Western Anatolian languages – consider (13):

13. Limited isogrammatism in a range of Western Anatolian languages (a. Laz, b. Western Armenian c. Turkish of Trabzon, d. Kurmanji e. Zazaki, f. Romeyka, g. Țuroyo (Midin, near Mardin), h. Lower Tyare (western Hakkari)):
- | | | | | | |
|----|--------------------------|-------------|--------|------------|---------------------------------------|
| a. | sum | | uškuri | eçopinu | ginon |
| | three | | apple | buy | want.2SG |
| | | | | | |
| b. | yerek ^h (hat) | | xənzor | añ-el | g’uzes (also possible: g’uzes añ-el)} |
| | three (CLASS) | | apple | buy.INFIN | want.2SG |
| | | | | | |
| c. | ûc | | elma | almak | istiysun |
| | three | | apple | buy.INFIN | want.2SG |
| d. | tu | dixwazî | sê | sêvan | bikirrî |
| | you | want.2SG | three | apples | buy.INFIN |
| | | | | | |
| e. | ti | wazenê | hirî | sa | biherinê |
| | you | want | three | apples | buy.SUBJ.2SG |
| | | | | | |
| f. | esi | thelis | tria | mila n’ | eperis |
| | you | want.2SG | three | apples SUB | take.2SG |
| g. | ko-bŷ-ət | zon-ət | tre | ħabuše | |
| | IND-want.IPFV-2MS | buy.IRR-2MS | two | apples | |
| | | | | | |
| h. | băyət | zon-ət | tre | xabuše | |
| | want.IPFV-2MS | buy.IRR-2MS | two | apples | |
- ‘You want to buy three apples’

Leaving aside a major methodological issue, namely that sometimes languages are compared with dialects and/or registers of different periods, the pertinent question is whether we should consider this apparent isogrammatism in (12) (SOV) and (13) (SO-‘want’-V) as evidence for a Sprachbund or not. The answer to this question is subject to both definition and interpretation, as already discussed in Section 2.2: (a) Anatolia, where Altaic, Indo-European, Kartvelian, Semitic and North-Western Caucasian language families have symbiosed for millennia, cannot constitute anything other than a Sprachbund in the original Trubetzkoy (1928) sense (see also Thomason, 2001, p. 99); (b) if however, Anatolian Sprachbund implies the emergence of a mixed language or morpho-syntactic isogrammatism, i.e., the same grammar for all the languages concerned through contact rather than inheritance, the answer is by no means straightforward. For many authors including Haig (2001), Stilo (2015) and Tzitzilis (to appear) (among others), full isogrammatism (for an elaboration on the term vis-à-vis other equivalence relations, see Heine and Kuteva, 2005), the most extreme manifestation of convergence, would probably be needed in order to legitimize the use of the term Anatolian Sprachbund. However, as we will show next in Section 4.2, such isogrammatism does not exist.

4.2 Patchy isogrammatism

Surface orders such as the ones in (12) and (14) (a) do not necessarily reflect deeper structural properties, namely base-generation of the object to the left of the verb versus discourse-sensitive (epiphenomenal) movement of the object to the left periphery; in other words it is unclear/unchecked/not argued for in the relevant literature whether the preverbal object in (12) and (14) bears focus or not, and if so, what type (informational, contrastive, etc.); (b) nor is it discussed whether OV order is associated with the same cluster of head-final properties (see (15)–(19)) – to put differently all languages displaying head-finality in simple declarative sentences will also be displaying N-finality and C-finality; or, (c) in fact, transmit the same historical signal. For instance, word order in Romeyka could be considered either as a case of (a) homoplasy (also called back-mutation) whereby the new feature is not shared by the immediate ancestor – if the immediate ancestor is Medieval Greek – in which case it is the result of contact; or as an instance of an (b) an inherited feature from Hellenistic Greek, which allowed discourse-sensitive movement of the object in the left periphery (see Sitaridou, 2016) without excluding the possibility that contact with Turkish further exacerbated the existing pattern and in fact leading to a reanalysis from discourse-sensitive movement to base-generation to the left of the verb. In fact, the only head-finality related properties uniformly found in Western Anatolia are: prenominal genitives and preposed predicatives, as shown in (15) and (16), respectively:

14. Aramaic/NENA: S-O-V-IO

- a. brota axəst-aw yawó-la l-jambali
 girl ring-POSS.3SG.F give.PST-3SG.F to-Jambali
 ‘The girl gave her ring to Jambali.’

but also in Aramaic/NENA: SVO

- b. Sabre-ni kebé-la bratexun
 son-POSS.1PL love-3SG.M.OBJ daughter.POSS.2PL
 ‘Our son loves your daughter.’
 (Fox, 2009 in Haig, 2015, p. 410)

Western Armenian: SOV

- c. Ara-n dun më šinec’
 Ara.NOM-DEF house INDEF build.PST.3SG
 ‘Ara built a house.’

Romeyka: SOV/VO

- d. (tšorban) alis efaen (tšorban)
 soup.ACC Alis.NOM eat.PST.3SG soup.ACC
 ‘Alis ate a soup.’

Laz: SOV

- e. Bere-k çitab-i do-nç’ar-u
 child-ERG book-NOM PV-write-PST.A.3SG
 ‘The child wrote the book.’
 (Öztürk and Pöchtrager, 2011, p. 56)

Trabzon Turkish: SOV

- f. Ali çorba ye-di
 Ali soup eat-PST. 3SG
 ‘Ali ate a soup.’

15. Preposed predicatives

Western Armenian:

- a. Ara-n šad keyec’ig e
 Ara.NOM-DEF very beautiful be.3SG
 ‘Ara is very beautiful.’

All AMG:

- b. atos o pap^hos=m en
 he the.NOM grandfather.NOM my be.3SG
 ‘This is my grandfather.’

Laz:

- c. Aiše msk’va o-n
 Ayşe. NOM beautiful be-3SG
 ‘Ayşe is beautiful.’
 (Öztürk and Pöchtrager, 2011, p. 120)

Trabzon Turkish:

- d. Ayşe güzel-ø/-dir
 Ayşe. NOM beautiful-be.3SG/-be.3SG
 ‘Ayşe is beautiful.’

16. Prenominal genitives (see Guardiano et al., 2016)

Western Armenian:

- a. Aram-i-n mama-n des-a
 Aram-GEN-DEF mother-DEF see-PST.1SG
 ‘I saw Aram’s mother.’

Romeyka:

- b. to zo to ылtsi
 the animal.GEN the milk.NOM
 ‘The milk of the animal.’

Cappadocian:

- c. rantsa Jorika *(du) klatš
 see.PST.1SG Jorikas.GEN the kid.ACC
 ‘I saw Jorikas’ child.’

Laz:

- d. Atina-şi çoye-pe
 Atina-GEN village-PL.NOM
 ‘Atina’s villages.’
 (Öztürk and Pöchtrager, 2011, p. 28)

Trabzon Turkish:

- e. Ali-nin kitab-ı
 Ali-GEN book-3SG.POSS
 ‘Ali’s book.’

Interestingly, other syntactic properties traditionally linked to head-finality do not display uniformity across Western Anatolian – these are verb-final auxiliaries (17), prenominal relative clauses (18), and possessor/adjective – noun order (19).

17. Verb-final auxiliaries (found in Western Armenian, Turkish, Laz, and Ulaghátsh variety of Cappadocian, but not in Romeyka although they have been reported to be have been possible in Amisos Pontic Greek):

Western Armenian:

- a. Ara-n dun mē šinaj e
 Ara.NOM-DEF house INDEF build.PTCP.3SG AUX.3SG
 ‘Ara has built a house.’

Ulaghátsh variety of Cappadocian:

- b. vava t na to pçaş ton
 father.NOM his SUB him catch.PNP.3SG be.IMP.3SG
 ‘His father would have caught him.’
 (Dawkins, 1916, p. 366)

Pazar Laz:

- c. Ma past’a v-i-ç’v-i donu
 I.ERG cake.NOM A.1-REFL-bake-PST.1SG AUX
 ‘I have baked a cake for myself.’
 (Öztürk and Pöchtrager, 2011, p. 66)

18. Prenominal relative clauses (found in Western Armenian, Laz, Pharasiot and Turkish, but in Romeyka there is variation between prenominal and postnominal ones):

Western Armenian:

- a. Ara-i-n šinaj dun-ē šad keyec’ig e
 Ara.GEN-DEF build.PTCP house-DEF very beautiful be.3SG
 ‘The house Ara has built is very beautiful.’

Pharasiot:

- b. iða tu yorases to tomafili
 saw.1SG that buy.PST.2SG the.ACC car.ACC
 ‘I saw the car that you have bought.’

Trabzon Turkish:

- c. [mera-da otlā-yan] koyun-lar
 meadow-LOC graze-SUBJ.REL sheep-PL
 ‘(the) sheep which are grazing/graze/have grazed in the meadow’

19. Possessor /adjective – noun order (strictly prenominal adjectives in Western Armenian, Romeyka, Laz and Trabzon Turkish) (see Guardiano et al., 2016):

Western Armenian:

- a. Ir yergu xelac'i yerexa-ner-ë (*yerexaner xelac'i)
 his two smart child-PL-DEF

Romeyka:

- b. (to tranon) (askemon) to muxteron (*to tranon) (*t=askemon)
 the big the=ugly the pig the big the=ugly
 'the big, ugly pig'

but Romeyka allowing both prenominal and postnominal possessors (also in Pharasiot):

- b'. temon to peði
 my the child

- b''. to peðim
 the child=my
 'my child'

Laz:

- c. Ts'ulu bere /*bere ts'ulu
 little child/child little
 'little child'
 (Öztürk and Pöchtrager, 2011, p. 35)

Trabzon Turkish:

- d. iki büyük yağız Arap at-ı
 two big swart Arabic horse-COMPOUND MARKER
 'two big swart Arabian horses'

Overall, although the empirical generalization about an emergent uniform OV (neo)Anatolian syntax may hold true if Anatolia is viewed with binoculars from afar (simple declarative sentences), it crumbles when the area is examined with a microscope (specific syntactic structures) thus indicating a range of possible explanations: (1) discourse-sensitive movement can also move the object to the left of the verb; (2) macro-parameters may be convergent, but micro-parameters – to be revealed by zooming in – are not thus suggesting that bottom-up change (from v-to-T-to-C, see Biberauer and Roberts, 2012) may be hijacked in word order change (possibly due to prosody); (3) inheritance cannot be ruled out. For all these reasons, a common OV diasystem (cf. Matras, 2009, p. 270) cannot be proposed here albeit there is, undoubtedly, a great deal of mutual convergence.

5. Future directions

Despite the recent surge in fieldwork in Turkey and the wider region (see Khan, 2007a, 2007b, 2008; and others for NENA, Donabedian (2018) for Western Armenian, Lacroix (2009) for Laz, Vaux (2007) for Homshetsma (West Armenian), Herin (2016) for Domari of Aleppo, Haig and Öpengin (2014) and Haig and Öpengin (2014) for Kurmanji Kurdish, Sitaridou (2013, 2014a, 2014b, 2016) for Romeyka), more work is needed on as many spoken varieties as possible using a combination of methods including observation, (semi-)spontaneous data, structured questionnaires. Also, the fieldwork needs to be urgently carried out in displaced/diasporic communities across the world in what should be the last concentrated effort to capture the Anatolian Babel before the communities become entirely attrited—the latter is by no

means less interesting; in fact, it can reveal new ‘buffer’ zones equally revealing and reveal a lot the mechanism of language contact.

Moreover, phylogeny-oriented methods are undoubtedly very promising, especially if we consider (a) a sufficient number of parameters, possibly subject to reciprocal interactions, but relatively isolated from interaction with parameters external to the set; (b) a sufficient number of languages; (3) a sufficiently fine-grained analysis of the data, as dictated by the ‘Modularized global parameterization’ model (Longobardi, 2003; Longobardi and Guardiano, 2009). Likewise, phylogenetic analyses of versions of Swadesh list from the earliest language attestations to the current ones is also believed to yield useful insights. Finally, development of a version of the Swadesh list with cultural vocabulary items (a proposal originally made by Ross, 2001) will open up further possibilities for socio-syntax approaches (consider, for instance, the possibility of studying the development of spatial deixis).

6. Further reading

Csató E.A., Isaksson N. and Jahani C., eds. (2005). *Linguistic convergence and areal diffusion. Case studies from Iranian, Semitic and Turkic*. London and New York: Routledge Curzon.

This volume presents a number of studies on language contact in three of the major language families found in Anatolia.

Haig, G. and Khan G., eds. (2018). *The languages and linguistics of Western Asia. An areal perspective*. Berlin and Boston: De Gruyter Mouton.

This book presents descriptions of the modern language of Anatolia spanning across all language families while evaluating similarities across the languages that may have arisen as a result of areal contact.

Johanson, L. (2002). *Structural factors in Turkic language contacts*. London: Routledge.

This study proposes a global approach of phenomena induced by language contact with Turkish with different languages. It also offers a useful introduction to Lars Johanson’s code-copying framework.

Tzitzilis, Ch. (In press). Sprachbund and etymology: Turkish etymology and the Anatolian Sprachbund.

In: G. Selz, ed., *The Tietze symposium proceedings (Vienna 14–17 July 2014)*. Wien: Wiener Offene Orientalistik.

This is one of the few papers discussing Asia Minor Greek within the wider Anatolian context, competently addressing issues pertaining to the existence of a Sprachbund on the basis of etymology.

7. Related topics

Borrowing, convergence, Balkans

Abbreviations

| | |
|-----|------------------------------|
| 1SG | 1st person singular |
| 2SG | 2d person singular |
| 3SG | 3d person singular |
| 2MS | 2d person masculine singular |
| ABL | ablative |
| ACC | accusative |
| ALL | allative |
| AMG | Asia Minor Greek |
| AUX | auxiliary |

| | |
|-------|-------------------------------|
| COMP | complementizer |
| COP | copula |
| CVB | converb |
| DAT | dative |
| DOA | differential object agreement |
| DOF | differential object flagging |
| DOM | differential object marking |
| DP | determiner phrase |
| DSM | differential subject marking |
| EArm | Eastern Armenian |
| ERG | ergative |
| EZ | ezafe |
| GEN | genitive |
| HUM | human |
| INF | infinitive |
| IMPF | imperfect |
| INDEF | indefinite |
| IPFV | imperfective |
| IRR | irrealis |
| LOC | locative |
| MWF | multiple wh-fronting |
| N | noun |
| NEG | negative |
| NENA | North-Eastern Neo-Aramaic |
| NOM | nominative |
| NP | noun phrase |
| NPST | non-past |
| PFV | perfective |
| PL | plural |
| PNP | perfective non past |
| POSS | possessive |
| PRS | present |
| PST | past |
| PTCP | participle |
| PV | preverb |
| REFL | reflexive |
| SBJ | subject |
| SUB | subordinator |
| SUBJ | subjunctive |
| TS | thematic suffix |
| VN | verbal noun |
| WArm | Western Armenian |

Note

- 1 Anaïd Donabedian gratefully acknowledges support by a public grant overseen by the French National Research Agency (ANR) as part of the program ‘Investissements d’Avenir’ (reference: ANR-10-LABX-0083). It contributes to the IdEx Université de Paris – ANR-18-IDEX-0001. Maps

are elaborated in collaboration with Emmanuel Giraudet, Laboratoire Monde Iranien et Indien, FRE 2018 – Labex EFL.

References

- Adjarian, H. (1909). *Classification des dialectes arméniens* [*Classification of Armenian dialects*]. Paris: Librairie Honore Champion.
- Akkus, F. (2018). Peripheral Arabic. In: E. Benmamoun and R. Bassiouney, eds., *The Routledge handbook of Arabic linguistics*, 1st ed. London and New York: Routledge, pp. 454–471.
- Arnold, W. (1998). *Die arabischen Dialekte Antiochiens.*(= *Semita Viva*; Bd. 19, zugl. Habilitationsschrift Erlangen 1996). Wiesbaden: Harrassowitz Verlag.
- Arnold, W. (2000). The Arabic dialects in the Turkish Province of Hatay and the Aramaic dialects in the Syrian mountains of Qalamūn: Two minority languages compared. In: J. Owens, ed., *Arabic as a minority language*, 1st ed. Berlin: De Gruyter Mouton, pp. 347–370.
- Arnold, W. (2006). Antiochia Arabic. In: K. Versteegh ed., *Encyclopedia of Arabic language and linguistics*, 1st ed. Leiden and Boston: Brill, pp. 111–119.
- Bağrıaçık, M. (2018). *Pharasiot Greek: Word order and clause structure*. PhD. Université de Gand.
- Bağrıaçık, M. and Janse, M. (2016). Reduplication with fixed segmentism in Cappadocian and Armenian as a contact-induced phenomenon. In: A. Ralli, ed., *Contact morphology in modern Greek dialects*, 1st ed. Newcastle upon Tyne: Cambridge Scholars Publishing, pp. 179–209.
- Balta, E. and Ölmez, M. (2011). *Between religion and language: Turkish-speaking Christians, Jews and Greek-Speaking Muslims and Catholics in the Ottoman empire*. Istanbul: Eren.
- Biberauer, T. and Roberts, I. (2012). Towards a parameter hierarchy for auxiliaries: Diachronic considerations. *Cambridge Occasional Papers in Linguistics*, 6, Article, 9, pp. 267–294.
- Blanc, H. (1964). *Communal dialects in Baghdad*. Cambridge, MA: Harvard University Press (Harvard Middle Eastern Monographs X).
- Bornes-Varol, M.C. (2008). Un processus endogène de categorisation metasystemique: le transfert du médiatif du turc au judeo-espagnol (Istanbul). In: F. Alvarez- Pereyre, ed., *Catégories et catégorisation – une perspective interdisciplinaire*. Société d’Etudes Linguistiques et Anthropologiques de France, 448(33). Paris: Peeters, pp. 95–122.
- Brendemoen, B. (2002). *The Turkish dialects of Trabzon. Their phonology and historical development* [Turcologica, Band 50]. Wiesbaden: Harrassowitz Verlag.
- Chatzikyriakidis, S. (2012). A dynamic account of the Cypriot Greek clitic positioningsystem. *Lingua*, 122, pp. 642–672.
- Chyet, M.L. (1995). Neo-Aramaic and Kurdish – an interdisciplinary consideration of their influence on each other. In: I. Shlomo and D. Rina, eds., *Language culture in the near east, Israel oriental studies XI*, 1st ed. Leiden and Boston: Brill, pp. 219–252.
- Coghill, E. (2014). Differential object marking in neo-Aramaic. *Linguistics*, 52(2), pp. 335–364.
- Coghill, E. (2018). The neo-Aramaic dialect of Telkepe. In: N. Vidro, R. Vollandt, E.M. Wagner and J. Olszowy-Schlanger, eds., *Semitic linguistics and manuscripts: A liber discipulorum in honour of Professor Geoffrey Khan*, 1st ed. Uppsala: Acta Universitatis Upsaliensis, pp. 234–271.
- Höfler, C.M. (2016). Group belonging beyond language boundaries –Language, religion and identity in the multilingual Greek community of Georgia. *Language Typology and Universals* 69(2): 213–234.
- Corbett, G. (2001). *Gender*. Cambridge: Cambridge University Press.
- Csató, E.A., Isaksson, N. and Jahani, C., eds. (2005). *Linguistic convergence and areal diffusion. Case studies from Iranian, semitic and Turkic*. London and New York: Routledge Curzon.
- Dahl, Ö. (2004). *The growth and the maintenance of linguistic complexity*. Amsterdam and Philadelphia: John Benjamins.
- Dawkins, R.M. (1916). *Modern Greek in Asia Minor*. Cambridge: Cambridge University Press.
- DeGraff, M. (2005). Morphology and word order in ‘creolization’ and beyond. In: G. Cinque and R. Kayne, eds., *The Oxford handbook of comparative syntax*, 1st ed. Oxford: Oxford University Press, pp. 293–372.

- Demir, N. (2018). Turkish reduplicative adjectives and adverbs. *Proceedings of the Linguistic Society of America*, 3(19), pp. 1–14.
- Donabedian, A. (2010a). Неоформленный субъект в западноармянском языке – к вопросу о неаккузативности [Bare subject in Western Armenian, contribution to the issue on inacusativity]. *Voprosy Jazykoznanija*, 1/2010, pp. 24–45.
- Donabedian, A. (2010b). Nom nu et tropisme typologique: le cas de l'arménien. In: F. Floricic, ed., *Essais de typologie et de linguistique générale, mélanges offerts à Denis Creissels*. Lyon: Ecole Normale Supérieure Editions, pp. 403–416.
- Donabedian, A. (2018). Middle East and beyond – Western Armenian at the crossroads: A sociolinguistic and typological sketch. In: C. Bulut ed., *Linguistic minorities in Turkey and Turkic-speaking minorities of the peripheries: Turcologica III*. Wiesbaden: Harassowitz Verlag, pp. 89–148.
- Drettas, G. (1997). *Aspects Pontiques*. Paris: Association de Recherches Pluridisciplinaires.
- Dumézil, G. (1965). *Documents anatoliens sur les langues et les traditions du Caucase. (Travaux et mémoires de l'institut d'ethnologie, LXXI)*. Paris: Institut d'Ethnologie.
- Eloeva, F. (1998). Les Grecs turcophones de Géorgie. Territoires et tradition orale à Tsalka et Tetriskaro. In: M. Bruneau, ed., *Les Grecs pontiques. Diaspora, identité, territoires*, 1st ed. Paris: CNRS, pp. 137–141.
- Fallmerayer, J.P. (1845). *Fragmente aus dem orient*. Stuttgart: J.G. Gotta'scher Verlag.
- Friedman, V.A. (1996). Typological and areal features linking and separating the Balkans and the Caucasus. In: J.I. Press and F.E. Knowles, eds., *Papers in Slavonic linguistics (Occasional Series)*, III, 1st ed. London: University of London, pp. 99–111.
- Greenblatt, J. (2011). The Jewish Neo-Aramaic dialect of Amədyā. In: *Studies in Semitic languages and linguistics, Vol. 61*. Leiden and Boston: Brill.
- Guardiano, C., Michelioudakis, D., Ceolin, A., Irimia, M., Longobardi, G., Radkevich, N., Silvestri, G. and Sitaridou, I. (2016). South by Southeast. A syntactic approach to Greek and Romance microvariation. *L'Italia dialettale LXXVII*, pp. 95–166.
- Haig, G. (2001). Linguistic diffusion in modern East Anatolia: From top to bottom. In: A. Aikhenvald and R. Dixon eds., *A real diffusion and genetic inheritance: Problems in comparative linguistics*. Oxford: Oxford University Press, pp. 195–224.
- Haig, G. (2004). Das Genussystem in der kurdischen Sprache: Strukturelle und soziolinguistische Aspekte. In: S. Hajo, C. Borck, E. Savelsberg and D. Şukriye, eds., *Gender in Kurdistan und der Diaspora*, 1st ed. Münster: Lit, pp. 33–58.
- Haig, G. (2008). *Alignment change in Iranian languages: A construction grammar approach*. Boston and Berlin: De Gruyter Mouton.
- Haig, G. (2014a). East Anatolia as a linguistic area? Empirical and conceptual issues. In: L. Behzadi, P. Franke, G. Haig, C. Herzog, B. Hoffmann, L. Korn and S. Talabardon, eds., *Bamberger Orient Studien, Vol. 1*, 1st ed. Bamberg: Bamberg University Press, pp. 13–36.
- Haig, G. (2014b). Regional variation in Kurmanji: A preliminary classification of dialects. *Kurdish Studies*, 2, pp. 143–176.
- Haig, G. (2015). Verb-Goal (VG) Word order in Kurdish and Neo-Aramaic: Typological and areal considerations. In: G. Khan and L. Napiórkowska, eds., *Neo-Aramaic and its linguistic context, Gorgias neo-Aramaic studies 14*, 1st ed. Piscataway, NJ: Gorgias Press, pp. 419–437.
- Haig, G. (2017). Western Asia: East Anatolia as a transition zone. In: R. Hickey, ed., *The Cambridge handbook of areal linguistics*, 1st ed. London: Cambridge University Press, pp. 396–423.
- Haig, G. and Khan, G., eds. (2018). *The languages and linguistics of Western Asia. An areal perspective*. Berlin and Boston: De Gruyter Mouton.
- Haig, G. and Öpengin, E. (2014). Introduction to special issue – Kurdish: A critical research overview. *Kurdish Studies*, 2(2), pp. 99–122.
- Haig, G. and Öpengin, E. (2015). Gender in Kurdish: Structural and socio-cultural dimensions. In: M. Hellinger and H. Motschenbacher, eds., *Gender across languages, Volume 4*, 1st ed. Amsterdam: John Benjamins, pp. 247–276.

- Haig, G. and Öpengin, E. (2018). Kurmanji Kurdish in Turkey: Structure, varieties and status. In: C. Bulut ed., *Linguistic minorities in Turkey and Turkic-speaking minorities of the peripheries: Turcologica III*. Wiesbaden: Harrassowitz Verlag, pp. 157–230.
- Heine, B. and Kuteva, T. (2005). *Language contact and grammatical change*. London: Cambridge University Press.
- Herin, B. (2016). Elements of Domari dialectology. *Mediterranean Language Review*, 23, pp. 33–73.
- von Heusinger, K. and Kornfilt, J. (2017). Partitivity and case marking in Turkish and related languages. *Glossa: A Journal of General Linguistics*, 2(1), 20, pp. 1–40.
- Janse, M. (2004). Animacy, definiteness and case in Cappadocian and other Asia Minor Greek dialects. *Journal of Greek Linguistics*, 5, pp. 3–26.
- Jastrow, O. (1980). Das mesopotamische Arabische. In: W. Fischer and O. Jastrow, eds., *Handbuch der Arabischen Dialekte*, 1st ed. Wiesbaden: Harrassowitz Verlag, pp. 140–154.
- Jastrow, O. (1985). Mlahsō: An unknown neo-Aramaic language of Turkey. *Journal of Semitic Studies*, XXX, pp. 265–270.
- Jastrow, O. (1988). *Der neuaramäische Dialekt von Hertevin (Provinz Siirt)*. *Semitica Viva* 3. Wiesbaden: Harrassowitz Verlag.
- Jastrow, O. (2006). Anatolian Arabic. In: K. Versteegh, ed., *Encyclopedia of Arabic language and linguistics*, 1st ed. Leiden and Boston: Brill, pp. 86–96.
- Jastrow, O. (2015). The position of Mardin Arabic in the Mesopotamian-Levantine dialect Continuum. In: L. Edzard, ed., *Arabic and semitic linguistics contextualized*, 1st ed. Wiesbaden: Harrassowitz Verlag, pp. 177–189.
- Johanson, L. (2002). *Structural factors in Turkic language contacts*. London: Routledge.
- Johanson, L. and Bulut, C., eds. (2006). *Turkic-Iranian contact areas. Historical and linguistic aspects (Turcologica 62)*. Wiesbaden: Harrassowitz Verlag.
- Johanson, L. and Utas, B., eds. (2000). *Evidentials (Turkic, Iranian and neighbouring languages)*. Berlin and New York: De Gruyter Mouton.
- Jügel, T. and Samvelian, P. (2016). Les pronoms enclitiques dans les langues ouest- iraniennes. Fonctions et distribution géographique. *Bulletin de la Société de Linguistique de Paris*, CXI, pp. 391–432.
- Karatsareas, P. (2009). The loss of grammatical gender in Cappadocian Greek. *Transactions of the Philological Society*, 107(2), pp. 196–230.
- Karatsareas, P. (2011). *A study of Cappadocian Greek nominal morphology from a diachronic and dialectological perspective*. PhD. University of Cambridge.
- Karatsareas, P. (2014). On the diachrony of gender in Asia Minor Greek: The development of semantic agreement in Pontic. *Language Sciences*, 43, pp. 77–101.
- Key, G. (2012). Differential object marking in Turkic and Persian as a contact phenomenon. In: *Proceedings of the annual meeting of the Berkeley linguistics society* 38. Berkeley: Berkeley Linguistics Society and Linguistic Society of America.
- Khan, G. (2007a). Remarks on the historical background of the Modern Assyrian language. *Journal of Assyrian Academic Studies*, 21(2), pp. 4–11.
- Khan, G. (2007b). The North-Eastern neo-Aramaic dialects. *Journal of Semitic Studies*, 52(1), pp. 1–20.
- Khan, G. (2008). *The neo-Aramaic dialect of Barwar*. Leiden: Brill.
- Kim, R. (2010). Towards a historical phonology of Modern Aramaic: The relative chronology of Turoyo sound changes. In: F.M. Fales and F. Grassi, eds., *Camsemud 2007: Proceedings of the 13th Italian meeting of Afro-Asiatic linguistics*. Padua: S.A.R.G.O.N. Editrice e Libreria, pp. 229–238.
- Komondouros, M. and McEntee-Atalianis, L. (2007). Language attitudes, shift and the ethnolinguistic vitality of the Greek orthodox community in Istanbul. *Journal of Multilingual and Multicultural Development*, 28(5), pp. 365–384.
- Lacroix, R. (2009). *Description du dialecte laze d'Arhavi (Caucasique du Sud, Turquie): Grammaire et textes*. PhD. Université Lyon 2, Lyon.
- Longobardi, G. (2003). Methods in parametric linguistics and cognitive history. *Linguistic Variation Yearbook*, 3, pp. 101–138.

- Longobardi, G. and Guardiano, C. (2009). Evidence for syntax as a signal of historical relatedness. *Lingua*, 119(11), pp. 1679–1706.
- Luraghi, S. (2010). *Linguistique historique et indo-européenne*. Louvain-la-Neuve: Peeters.
- MacKenzie, D. (1961, 1962). *Kurdish dialect studies I, II*. Oxford: Oxford University Press.
- Mackridge, P. (1987). Greek-Speaking Moslems of North-East Turkey: Prolegomena to a study of the Ophitic sub-dialect of Pontic. In: *Byzantine and modern Greek studies*. Cambridge: Cambridge University Press, 11, pp. 115–137.
- Matras, Y. (2000). How predictable is contact induced change in grammar? In: C. Renfrew, A. McMahon and R.L. Trask, eds., *Time depth in historical linguistics*, 1st ed. Cambridge: McDonalds Institute for Archaeological Research, pp. 563–583.
- Matras, Y. (2002). *Romani: A linguistic introduction*. Cambridge: Cambridge University Press.
- Matras, Y. (2009). *Language contact*. London: Cambridge University Press.
- Matras, Y. (2010). Contact, convergence and typology. In: R. Hickey, ed., *Handbook of language contact*, 1st ed. Oxford: Blackwell, pp. 66–85.
- Matras, Y. (2011). Universals of structural borrowing. In: P. Siemund, ed., *Linguistic universals and language variation*, 1st ed. Berlin: De Gruyter Mouton, pp. 200–229.
- Matras, Y. (2012). *A grammar of Domari*. Berlin and Boston: De Gruyter Mouton. Available at: www.degruyter.com/view/product/183441 [Accessed 23 Jul. 2019].
- Michelioudakis, D. and Sitaridou, I. (2016). Recasting the typology of multiple *wh*-fronting: Evidence from Pontic Greek. *Glossa: A Journal of General Linguistics*, 1(1), pp. 1–33.
- Michelioudakis, D. and Sitaridou, I. (2019). Towards a formal model of transfer under contact: Contrasting Asia Minor Greek to mainland Greek and Turkish in search of syntactic borrowings. In: M. Georgiafentis, G. Giannouloupoulou, M. Koliopoulou and A. Tsokoglou, eds., *Theoretical approaches to contrastive linguistics. Morphological and syntactic perspectives*. London: Bloomsbury Studies in Theoretical Linguistics, pp. 245–261.
- Napiórkowska, L. (2013). The two neo-Aramaic dialects from Urmi in the light of language contact. *Rocznik Orientalistyczny/Yearbook of Oriental Studies*, Z. 2, pp. 86–101.
- Oeconomides, D.E. (1908). *Laulehre des Pontischen*. A. Leipzig: Deichert'sche Verlagsbuchhandlung Nachf.
- Öztürk, B. and Pöchrager, M.A., eds. (2011). *Pazar Laz*. München: Lincom Europa.
- Paul, L. (1998). The Position of Zazaki among West Iranian Languages. In: N. Sims-Williams, ed., *Proceedings of the third European conference of Iranian studies held in Cambridge, 11th to 15th September 1995. Part I: Old and Middle Iranian Studies*. Wiesbaden: Ludwig Reichert, pp. 163–177.
- Poizat, B. (2016). *Parlons Soureth: Araméen Contemporain*. Paris: L'Harmattan.
- Pooth, R., Kerkhof, P., Kulikov, L. and Barðdal, J. (in press). The origin of non-canonical case marking of subjects in Proto-Indo-European: Accusative, ergative, or semantic alignment. *Indogermanische Forschungen*.
- Poplack, S. and Levey, S. (2010). Contact-induced grammatical change. In: P. Auer and J.E. Schmidt, eds., *Language and space – An international handbook of linguistic variation: Volume 1 – theories and methods*, 1st ed. Berlin: De Gruyter Mouton, pp. 391–419.
- Procházka, S. (1999). From language contact to language death: The example of the Arabic spoken in Cilicia (Southern Turkey). *Orientalia Suecana*, 48, pp. 115–125.
- Procházka, S. (2014). Feminine and masculine plural pronouns in Modern Arabic dialects. In: Davidovich et al. eds., *From Tur Abdin to Hadramawt: Semitic studies. Festschrift in honour of Bo Isaksson*, 1st ed. Weisbaden: Harrassowitz Verlag.
- Renault, R. (1987). Genre grammatical et typologie linguistique. *Bulletin de la Société de Linguistique de Paris*, 82(1), pp. 69–117.
- Romagnolo, D. (2015). The Greek-Anatolian area in the 2nd millennium B.C.: Between language contact, Indo-European inheritance and typologically natural tendencies. *Studi e Saggi Linguistici*, [S.I.], 53(2), pp. 429–446. [online] Available at: www.studiesagginguistici.it/index.php/ssl/article/view/135 [Accessed 15 Jul. 2019].
- Ross, M.D. (2001). Contact-induced change in Oceanic languages in North-West Melanesia. In: A. Aikhenvald and R.M.W. Dixon, eds., *Areal diffusion and genetic inheritance: Problems in comparative linguistics*. Oxford: Oxford University Press, pp. 134–166.

- Scala, A. (2014). The mixed language of the Armenian Boshha (Lomavren) and its inflectional morphology: Some considerations in light of Armenian dialectal variation. In: *Annali del dipartimento di studi letterari, linguistici e comparati. Università degli studi di Napoli L'Orientale. Sezione linguistica 3*, pp. 233–250.
- Sitaridou, I. (2013). Greek-speaking enclaves in Pontus today: The documentation and revitalization of Romeyka. In: M. Jones and S. Ogilvie, eds., *Keeping languages alive. language endangerment: Documentation, pedagogy and revitalization*, 1st ed. Cambridge: Cambridge University Press, pp. 98–112.
- Sitaridou, I. (2014a). Modality, antiveridicality, and complementation: The Romeyka infinitive as a negative polarity item. *Lingua*, pp. 118–146.
- Sitaridou, I. (2014b). The Romeyka infinitive: Continuity, contact and change in the Hellenic varieties of Pontus. *Diachronica*, 31(1), pp. 23–73.
- Sitaridou, I. (2016). Refraining the phylogeny of Asia Minor Greek: The view from Pontic Greek. *CHS Research Bulletin, Center for Hellenic Studies, Harvard University*, 4(1), pp. 1–17.
- Sitaridou, I. and Kaltsa, M. (2014). Contrastivity in Pontic Greek. *Lingua*, pp. 1–27.
- Skopeteas, S. (2014). Caucasian Urums and Urum language. *Journal of Endangered Turkish Languages*, 3(1), pp. 333–364.
- Stilo, D. (2005). Iranian as a buffer zone between the universal typology of Turkic and Semitic. In: E. Csato, B. Issakson and C. Jahani, eds., *Linguistic convergence and areal diffusion. Case studies from Iranian, Semitic and Turkic*. London and New York: Routledge Curzon, pp. 35–63.
- Stilo, D. (2007). An introduction to the “Atlas of the Araxes linguistic area”. In: Presentation at the *Conference on the languages of the Caucasus*. Leipzig. Available at: www.eva.mpg.de/lingua/conference/07-CaucasusConference/pdf/final%20abstracts%20english/StiloAbstractl-Engl.pdf.
- Stilo, D. (2012). Intersection zones, overlapping isoglosses, and ‘fade-out/fade-in’ phenomena in central Iran. In: B. Aghaei and M.R. Ghanooonparvar, eds., *Iranian language and culture: Essays in honor of Gernot Ludwig Windfuhr*, 1st ed. Costa Mesa: Mazda, pp. 3–33.
- Stilo, D. (2015). An introduction to the Atlas of the Araxes-Iran linguistic area. In: B. Comrie and L. Golluscio, eds., *Language contact and documentation/contacto lingüístico y documentación*, 1st ed. Berlin and Boston: De Gruyter Mouton, pp. 343–356.
- Thomason, S. (2001). *Language contact: An introduction*. Georgetown University Press.
- Thomason, S. and Kaufman, T. (1988). *Language contact, creolization, and genetic linguistics*. Berkeley: University of California Press.
- Trubetzkoy, N.S. (1928). Proposition 16. In: *Actes du Premier Congrès International des Linguistes à la Haye*. Leiden: A.W. Sijthoff, pp. 17–18.
- Trudgill, P. (2011). *Sociolinguistic typology: Social determinants of linguistic complexity*. Oxford: Oxford University Press.
- Tzitzilis, C. (1989). Zu den griechisch-türkischen Sprachbeziehungen. *Linguistique Balkanique*, 32, pp. 185–197.
- Tzitzilis, C. (to appear). Sprachbund and etymology: Turkish etymology and the Anatolian Sprachbund. In: G. Selz, ed., *The Tietze symposium proceedings (Vienna 14–17 July 2014)*. Wien: Wiener Offene Orientalistik.
- Vaux, B. (2007). Homshetsma: The language of the Armenians of Hamshen. In: H. Simonian, ed., *The Hemshin: History, society and identity in the highlands of Northeast Turkey*, 1st ed. Abingdon: Routledge, pp. 257–278.
- Vaux, B., LaPorta, S. and Tucker, E. (1996). Ethnographic materials from the Muslim Hemshinli with linguistic notes. *Annual of Armenian Linguistics*, 17, pp. 25–46.
- Watkins, C. (2001). An Indo-European linguistic area and its characteristics: Ancient Anatolia. Areal diffusion as a challenge to the comparative method. In: A.V. Aikhenvald and R.M. W. Dixon, eds., *Areal diffusion and genetic inheritance*. Oxford: Oxford University Press, pp. 44–63.
- Xanthopoulou-Kyriakou, A. (1991). The diaspora of the Greeks of the Pontos. Historical background. *Journal of Refugee Studies*, 4, pp. 357–363.

Language contact in the Asian region

Umberto Ansaldo and Lisa Lim

1. Introduction

The task of covering Asia in terms of any linguistic phenomena within a single chapter calls for a great deal of synthesis as well as selection, in order to be able to both paint an overview, as well as provide more detailed sketches of this vast and linguistically diverse area. This is even more so when it comes to language contact, a field which is intrinsically interdisciplinary, and which requires historical, sociological as well as linguistic perspectives. In this chapter the focus is on South Asia, i.e., India and Sri Lanka, and Southeast Asia, which includes Mainland and peninsular Southeast Asia as well as southern China. Central Asia, in itself a diverse, rich, and complex area, is in many ways the meeting point of Western Asian, Middle Eastern, as well as Eastern European cultures, and deserves a treatment of its own. At the other end of the Asian map, northeastern China, Korea, and Japan are, due to a history of relative historical isolation, somewhat less salient from the point of view of language contact.

There are three broad phenomena that define the areas of South and Southeast Asia in terms of language contact, and they are interrelated: migration, climate, and trade. Migration is obviously one of the prerequisites for contact to occur, and the region is rich in migration patterns. We see this very clearly in the deep history of South Asia, where Indo-Aryan and Dravidian cultures cross paths, as well as in later movements of peoples during colonial exploration and settlement, and the contact languages that evolved in those contexts. Climate is one of the fundamental factors that determines the nature and direction of such migrations. The Indian Ocean here plays a fundamental role, defined as it is by two seasonal wind patterns, the monsoons that allow travel in southwesterly direction for half of the year, and then reverse to command travel in northeasterly direction for the other half. Along these travel routes, we observe in particular the nature of Sino-Javanese contacts. Trade is one of the most powerful dynamics of cultural exchange, whose interaction necessitates communication amongst diverse parties. It is along the trade routes and especially within the commercial enclaves that arise to sustain them that multilingualism flourishes, and with it language contact, shift, and admixture. This is pervasive in the region, evidenced in the regional networks pre-dating European discovery of the region, and in encounters between the languages of Western colonial powers such as Portuguese and English, and languages of the local powers such as Malay and Sinitic.

With the disparate regions of South and Southeast Asia having diverse histories and presenting different lessons in contact issues, this chapter discusses each region separately, highlighting, for each of them, a selection of contact language situations and the particular processes and outcomes which are instructive for each.

Finally, because a study of language contact cannot only be looking back in time, this chapter rounds off its offering by highlighting new contact dynamics and outcomes in contemporary ecologies.

2. Historical overview, critical issues, and current research

2.1 *South Asia*

2.1.1 *Indo-Aryan and Dravidian admixture*

India (here not including the Himalayas) and Sri Lanka have traditionally been occupied by Indo-Aryan and Dravidian populations, with areas of Austroasiatic people. There are reasons to believe that Austroasiatic populations were the earliest inhabitants: today residual pockets of them are found in the East and the Centre of India, often Munda speakers, while Sri Lanka's Vedda peoples, aboriginal populations of unclear genetic origin, are now extinct. It is quite likely that Dravidian populations occupied most of the Indian subcontinent and Sri Lanka until the arrival of Indo-Aryan populations that led to their partial displacement (Cavalli-Sforza, Menozzi and Piazza, 1994).

It is crucial to recognize, however, that it is hardly possible to distinguish genetically between Dravidians and Indo-Aryans in this region, as discussed in Ansaldo (2015). This means that a deep and prolonged admixture between the two populations has taken place over the centuries – according to Reich et al. (2009), the admixture of Northern and Southern Indian genetic traits could go back as far as 3,500 years, with Lankan people likewise extremely mixed.

Linguistic evidence supports this scenario. The presence of Dravidian features can be found in Old Indo-Aryan, which could indicate that a language shift took place in speakers of Dravidian languages towards Indo-Aryan varieties (Thomason and Kaufman, 1988; Erdosy, 1995). This shift may be explained by cultural assimilation with a more prestigious neighbour, or it could be interpreted as the result of different waves of migration.

The degree of admixture witnessed in the peoples does naturally extend to language. This language admixture is what lies beneath the origins of one of the most illustrative examples of Sprachbund: Emeneau (1956), who defined a linguistic area as one which includes languages of different families with shared traits that set them apart from other members of those same families, famously identified India as such an area, a place where Indo-Aryan and Dravidian resemble each other more than Indo-Aryan resembles other Indo-European languages. A rich inventory of features is recognized that display 'Indianization,' i.e., traits that are definitely not Proto-Indo-European and mostly Dravidian in origin (Emeneau, 1956, synthesized in Ansaldo, 2015, p. 576).

In phonology: retroflex or cerebral consonants in contrast with dentals series. The earliest Sanskrit records already display phonemes of this class, which are definitely not Proto-Indo-European and generally not found in other Indo-European languages. In Dravidian, retroflexes in contrast with dentals are Proto-Dravidian in origin, leading to the conclusion that pre-Indo-Aryan and pre-Dravidian bilingualism might have provided the conditions under which pre-Indo-Aryan allophones became redistributed as retroflex phonemes.

In grammar (based on earlier work by Bloch, 1965):

- (a) Loss of dual: starting from Sanskrit, then paralleled by the rest of the Indo-European domain;
- (b) Loss of infixation;
- (c) Disuse of verbal prefixes: starting in Modern Indo-Aryan and tied up with the general shift of accent to initial syllables;
- (d) Absence of prepositions in favour of compounded prefixes;
- (e) Two themes of personal pronouns: i.e., double stems in personal pronouns in Indo-European languages and the same phenomena found in some Dravidian languages;
- (f) Constructions in which verb stems or non-finite verb forms are strung together in series which are closed by a finite verb form (or other predicate ender): a prominent feature in Dravidian and Munda languages, but a feature that makes Sanskrit stand out from other Indo-European languages;
- (g) Constructions based on a nominalized form of a verb (or rather of a predication ending in a verb) followed by a postposition: common in pan-Dravidian languages, Buddhist Hybrid Sanskrit, Pali, and Asian languages outside India, but unavailable in Munda languages, e.g., [Marathi] *tujhī āī vārlyā-pāsūn*, ‘your mother died-since’;
- (h) Echo-word construction, in which a basic word formulated as CVX is followed by an echo-word in which CV is replaced by a morpheme *gi-* or *u-* or the like (or C is replaced by *m-* or the like), and X echoes the X (or VX echoes the VX) of the basic word; the meaning of the echo-word is ‘and the like’;
- (j) Classifiers or quantifiers: a shared feature among all the three families, spread from Indo-Aryan but not an Indo-European phenomenon.

It is notable that Dravidian has a deeper impact on Indo-Aryan than vice-versa. The influence from Dravidian into Indo-Aryan is of a structural nature, suggesting its role there as a substrate, a view held by Emeneau, in the sense of a pre-existing language. It is possible that Dravidians, as L2 speakers of an early form of Sanskrit, were the agents of a restructuring of that language, possibly a lingua franca, in the development leading Middle Indo-Aryan (or Middle Indic) (Krishnamurti, 1969). Dravidian, on the other hand, shows mostly superficial, if extensive lexical borrowings from Indo-Aryan.

While Indo-Aryan languages are not predominant in South India, in Sri Lanka they constitute the dominant variety of the island, in the form of Sinhala; the closely related Maldivian (Dhivehi) is spoken on the island of Maldives. But Sinhala is rather different from the Indo-Aryan varieties in the north, revealing a history of contacts of a different kind, as well as possibly independent innovations. While retaining a strong Indo-Aryan feel, Sinhala shows influences of indigenous Vedda as well as Dravidian languages in its lexicon, as well as Sinhala-Tamil compounds, having been in contact with these for a prolonged period of time, (Gair, 1998). Possible independent innovations in grammar, which cannot be traced to neighbouring languages, are noted in the gender system, including the loss of the neutral present in both Indo-Aryan and Dravidian, as well as anomaly distinctions. Beyond the basic animate/inanimate distinction, also found in some contact languages of the island, there is a further distinction between human and animal, and within these the additional dimensions of general and feminine (Gair, 1998, p. 9).

Today Indo-Aryan languages are dominant in North India and most of Sri Lanka, while Dravidian languages are dominant in South India. Notwithstanding this distribution, and also noting the character of Sinhala as mentioned earlier, a further, more specific linguistic area

has also been identified – the ‘South-South Asia’ (SSA) area (Gair, 1994) which includes the southern states of India characterized by Dravidian varieties, and Sri Lanka. This is substantiated by genetic studies (Kshatriya, 1995) which show that present-day Sinhalese and Tamils of Sri Lanka are closer to Indian Tamils and South Indian Muslims than they are to, for example, Gujaratis and Punjabis of north-west India.

2.1.2 Contact languages in Sri Lanka

Sri Lanka is also home to two contact languages, Sri Lanka Malay (SLM) and Sri Lanka Portuguese (SLP; also known as Sri Lanka Creole Portuguese or Burgher Portuguese), both of which have gained prominence in language contact and creole studies in recent decades. Both are of great historical interest, being contact languages which came about as a consequence of population movements and contact during the colonial period, and which developed along the same processes that typically characterize linguistic areas. Both are also especially instructive for language contact studies, with SLM in particular being used to describe processes of typological convergence, metatypy, or admixture, that lead to the emergence of an innovative, grammatically mixed variety.

Sri Lanka Portuguese (SLP) is a Portuguese-lexified creole language spoken by the Portuguese Burgher community of Eastern Sri Lanka and historically also by the Kaffir and the Dutch Burgher communities. The formation of the creole is connected with Portuguese colonial domination of large tracts of Sri Lanka (then Ceylon) between the early sixteenth and the mid-seventeenth century, with the original speakers either Portuguese or Portuguese slaves from African colonies, descendants of African slaves who were brought over. When the Dutch took over the island from the Portuguese, these groups started becoming more marginal; today there are only a few pockets of SLP speakers remaining, notably on the east coast around Batticaloa, the Kaffirs in Puttalam, and in the northern area of Trincomalee. It is generally believed that SLP emerged in admixture of primarily varieties of Portuguese and Sri Lankan Tamil (Smith, 1979a, 1979b). Its grammar is characterized by basic V-final word order, postpositions, sentence final particles to mark quotative case, as well as a typical South Asian case system (see Table 22.1); also notable are particular genres of music and dance (Cardoso, 2017). SLP is considered gravely endangered with much language loss in middle and younger generations, and often only ritual usages of language in the older generations, especially in the musical tradition.

Much more vital are the communities of Sri Lanka Malay that occupy the central areas around Kandy, the south-eastern hamlet of Kirinda, and, to some extent, pockets in the capital, Colombo. The majority of the Malays¹ of Sri Lanka trace their ancestry to the communities of people from the region encompassing Malaysia through to the easternmost provinces of

Table 22.1 Recent Lankan contact features

| | <i>SLM</i> | <i>SLP</i> | <i>Lankan</i> |
|--------------------------|---|---|---|
| Case morphology | Case suffixes and postpositions | Case suffixes and postpositions | Case suffixes |
| Discourse markers | Enclitics, verbal suffixes of particles | Enclitics, verbal suffixes of particles | Enclitics, verbal suffixes of particles |
| Verbal morphology | Suffixes only | Suffixes, preverbal | Preverbal |

Indonesia (the Dutch East Indies) who were brought over to the island during Dutch (1656–1796) and British rule (1796–1948); the community based in Colombo’s Slave Island district may have been there already during Portuguese rule. The different communities included: exiled nobility and political dissenters deported with their families from different corners of the Indonesian archipelago and beyond, including Java, the Maluku and Goa, during Dutch occupation of the East Indies; a large contingent of soldiers with their wives coming from disparate places such as Bali, Java, Riau, Ambon and peninsular Malaysia, imported first by the Dutch to form a ‘Malay’ garrison to fight against native rulers, becoming the Ceylon Rifle Regiment under the British who continued the same practice; a third group comprising convicts, slaves and indentured labourers, probably from Portuguese occupation, through both Dutch and British rule. Contacts between the groups were quite frequent, due among other reasons to the practice of employing noblemen as officers of the troops, master-servant relations, and a common, Islamic faith. As a small minority in contact with the major ethnic groups in Ceylon, it is not surprising that, over the centuries, various aspects of the Malay/Indonesian community evolved to become ‘Lankan.’ While their religious practices were maintained in the Muslim tradition, the community assimilated numerous cultural traditions of the two dominant ethnic groups of the island, the Sinhalese and the Tamils. For instance, until only very recently, the Sri Lankan Malay women have worn the South Asian sari as their traditional dress, rather than baju kurong or sarong kebaya as in Malaysia and Indonesia, and weddings involve payment of a dowry as in Hindu practice, in contrast with Islamic tradition which only involves the groom’s payment of mahar to the bride’s father. (For a fuller historical and socio-linguistic account of the SLM community, see e.g., Lim and Ansaldo, 2007.)

The formation of the vernacular Sri Lanka Malay (SLM) has been described as the outcome of language contact between, on the one hand, early Malay varieties, including vernaculars and lingua francas, brought over from the Indonesian archipelago – Bazaar Malay, illustrated in (1) (also see next section), would have allowed the early Malay settlers to communicate across their vernacular varieties as diverse as Javanese, Balinese, Ambonese – with, on the other hand, two typologically distinct adstrates: colloquial Sinhala, the dominant language of the population of Sri Lanka, and Lankan Tamil, spoken by, amongst others, traders and plantation workers, illustrated in (2, 3). What evolved was a restructured mixed language of trilingual base, with lexical items predominantly from Bazaar Malay and grammatical features, including V-final word order, number and case morphology, and agglutination, from Sinhala and Tamil (Ansaldo, 2009), illustrated in (4), contrasted with Standard Malay in (5). SLM is usually considered a creole in the literature, with more recent work (e.g., Ansaldo, 2015) capturing the process of language creation as one of metatypy leading to a hybrid profile of Lankan grammar and Malay-derived lexicon, as will be elaborated on later.

1. saya tak tahu cakap melayu (Trade/ Bazaar) Malay (Austronesian)
I NEG know speak Malay
‘I can’t speak Malay.’
2. eyaa-te hungak salli tiu-na Sinhala (Indo-European)
he-DAT much money exist-PAST
‘He had a lot of money.’
3. ongal-ukku ayar-e teriyumaa? Tamil (Dravidian)
you-DAT he-ACC know
‘Do you know him?’

4. samma anak-pada manahari skul-nang ar-pi Sri Lanka Malay
 all child-PL everyday school-DAT DUR-go
 ‘All the children go to school every day.’
5. kanak-kanak semua sehari-hari pergi sekolah (Standard) Malay
 child-2 every one.day-2 go school
 ‘All the children go to school every day.’

There are two main significant points to note of language contact outcomes observed in Sri Lanka. The big picture first is that both contact languages of SLP and SLM have evolved away from their original structural type, to display strong typological affinities of SSA type. They are verb-final, morphologically agglutinative or fusional, phonologically ‘Indian,’ and only retain traces of, respectively, their European and Malay heritage in the lexicon; several features are summarized in Table 22.1 (based on Bakker, 2006, pp. 141–144). The point to be underscored is that typological convergence deeply affects even newcomer languages, such as SLM and SLP.

Focusing on SLM more specifically, common traits shared with Dravidian and Sinhala can be seen in the following list of more convincing cases of areal features (a selection from Gair, 2013):

- (a) Post-verbal sentence-final question marker when unmarked, question-internal when the topic is in focus. This feature is found in Sinhala, SLM, as well as some Dravidian languages, though not Tamil. In Sinhala, for example, (6a) reads as ‘That man gave Gunapaala money yesterday?’, while (6b) would be rendered as ‘Was it **that man** that gave Gunapaala money yesterday?’

- 6a. ee minihaa iiye Gunapaala-ṭa salli dunnaa=da?
 that man yesterday Gunapaala-DAT money gave=INTERR
 6b. ee minihaa=da iiye Gunapaala-ṭa salli dunn-e?
 that man=INTERR yesterday Gunapaala-DAT money gave-FOC

- (b) Subordinate clause marking in the form of final verbal affixes or conjunctive forms. This is an areal feature of South Asia as a whole, but it distinguishes itself in SSA in two ways: the marker is placed at the right margin, rather than the left, and these markers tend to be multifunctional. This is illustrated in Sinhala in example (7) ‘Siri ate and went home.’

7. siri koeoema kaalaa gedara giyaa
 Siri food eat\CP home went

- (c) Preposed relative clauses occur in Sinhala, Tamil, and SLM as main or exclusive form.
 (d) Sentence-final quotative markers ‘say’ are used in all SSA varieties to mark in/direct speech.
 (e) Nominalization without genetivization through verbal affixes, found in all SSA varieties, e.g., *eka* in Sinhala.

8. Silva mahattaya ma-ta eeka kiww-a-eka oetta
 Silva gentleman I-DAT that say\PAST-ADJ-NMLZ truth
 ‘It is true that Mister Silva said that to me.’

- (f) A variety of complex negative forms including verbal, equational, existential, and subordinate negative verbs, found in Sinhala, Tamil, and SLM.
 (g) Sentence-final hearsay particle that function as evidential clitics.

Table 22.2 Core case in SLM, Sinhala and Lankan Tamil

| Case | Function | | |
|-------------------|--------------------------------------|--------------------------------------|--------------------------------------|
| | SLM | Sinhala | Tamil |
| Dative | Experiencer/ Genitive/ Possession | Experiencer/ Genitive/ Possession | Experiencer/ Genitive/ Possession |
| Nominative | Agent | Agent | Agent |
| Accusative | Patient | Patient | Patient |

This overwhelming typological restructuring has been explained through the typological pressure that Sinhala and Tamil combined exercise over the emblematic original variety, in a typical metatypic scenario. Looking more closely at the case system, metatypic effects are observed that render Sinhala, Lankan Tamil, SLM, and SLP comparable (Bakker and Mous, 1994; Ansaldo, 2009, 2011). One of the most striking features in this respect is the marking of dative subjects in all SSA varieties, i.e., dative marking is used for agents who are perceived as not being particularly in control of an action. As Table 22.2 shows, dative case is used to mark experiential roles as well as goal, beneficiary, and possessive in SLM, Sinhala and Lankan Tamil.

The second significant point pertains to the processes involved. The development of the case system is part of three noteworthy syntactic-semantic alignments happening in the emblematic language (Ansaldo, 2011):

- (a) General VO > OV shift; from Austronesian to Lankan word-order;
- (b) Following (a), the grammaticalization of [PDM Lex + N] into [N-Case Marker];
- (c) In parallel with (b), the semantic obligation to express (core) cases.

Elaborating on (b), the development of the case markers have been speculated to have etymologies as follows: dative (DAT) *nang* coming from Malay *nang* ‘towards’; accusative (ACC) *yang* from Malay-*nya*, a definiteness marker; and instrumental/ ablative (INST/ABL) *dering* (often reduced to *ring*) from Jakarta *bikin*, ‘to make’ (Ansaldo, 2009, p. 129). These are illustrated in examples (9) to (11) (adapted from Nordhoff, 2009, p. 483).

9. go=dang karang bannyak thàràsiggār
I=DAT now very sick
‘I am very sick now.’
10. Titanic kappal=yang su-thinggālam
Titanic ship=ACC PAST-sink
‘The ship Titanic sank.’
11. Police=dering su-dhaatang
police=INSTR PAST-come
‘The police came.’

What is also important to note is that both contact languages have developed morphological *complexity* in the nominal and verbal domains, in conjunction with shifting to a verb-final,

right-headed typology. In addition to the historical scenario of contact, given that these varieties emerged in a rich multilingual pool typical of contact environments, where Lankan features dominate, this is also strong evidence of the role of typological frequency in action in shaping the direction of restructuring (Ansaldò, 2009). If one assumes the typical simplified scenario of language genesis starting in a pidgin, then the change is even more dramatic and shows an overwhelming dominance of the typological pool.

The metatypic scenario that accounts for this state of affairs has been explained as originating in bilingual individuals possessing an enriched linguistic system which they are capable of appropriately adapting to the context in which they function (Matras, 2010, p. 66). In this sense, language contact phenomena can be seen as function-driven interferences that speakers are subject to in goal-oriented communicative interaction. Convergence thus emerges as a process that ‘offers speakers the opportunity to accommodate and generalise and yet still hold on to a mental demarcation between subsets of word forms within their repertoire’ (Matras, 2010, p. 76). The type of contact-induced change that defines the region is of the kind that leads to emergence of complex structure rather than simplification. While this may be related to a substantial role of young learners in the restructuring, it is impossible not to note the role of typological ecology in which this contact takes place.

It is clear that prolonged contact between unrelated languages has promoted a number of significant common developments in the region. The scenario advocated for the emergence of a SA linguistic area, namely restructuring through second language acquisition, can also be applied to the SSA area. Just as the Tamils of Sri Lanka are closer to the Sinhalese because they were always in close proximity to each other historically, linguistically, and culturally, the Malay and Portuguese of Sri Lanka, despite being more recent arrivals, have similarly undergone convergence to Lankan language.

Recent and ongoing research, documentation and online resources of these contact languages include, inter alia, the SLM corpus (principal investigator U. Ansaldò; Ansaldò et al., 2004) at the Documentation for Endangered Languages (DOBES) archive, and the SLP corpus (principal investigator H. C. Cardoso; Cardoso, 2017) in the Endangered Languages Archive.

2.2 *Southeast Asia*

2.2.1 *Climate, geopolitics, trade*

To appreciate the contact patterns that characterize maritime Asia, one must recognize at the outset a particularly defining feature of the region: its wind patterns. Unlike the trade winds of the Atlantic, the Indian Ocean, the China Sea, and the Sea of Japan are ruled by the seasonal winds known as monsoons, from Malay *musim* ‘season.’ These winds have for the longest time carried vessels from the southwest during the period from April to September, from the Arabian Sea all the way to Japan. From October to February, on the other hand, the direction of the monsoon reverses, facilitating sailing from Northeast to Southeast Asia and beyond. For the longest time, these patterns determined trade and navigation in and around Southeast Asia: in order to travel with the most favourable conditions, a common practice was for merchants to sail in one direction during one part of the year and, once a good trading port was established, to spend the months waiting on land for the change in monsoons.

Another factor defining Southeast Asia’s contact phenomena is its position historically as a long-standing hotspot of trade, for two major geopolitical reasons: it lies between two significant classical centres of power, India and China, and it houses the fabled Spice Islands – the Indonesian archipelago of the Moluccas (or Maluku Islands) – so-called for their then exclusive source of

spices of cloves, nutmeg and mace, which were one of the most precious and contested sources of wealth, especially once the Europeans discovered them in the sixteenth century.

In the period between the seventh and the fourteenth century – long before European colonization – we see the establishment and development in this region of a vibrant cultural sphere characterized by networks of trade, small sultanates, large empires, and self-sufficient city-ports (see Ansaldo, 2009 for a detailed historical account). Along these routes and within these centres, multicultural, multilingual, and religiously pluralistic societies developed within which exchange of ideas, goods, and services was at the very centre of social activity. Not surprisingly, this also led to intense language contact and the emergence of new languages.

One of the earliest well-recorded contacts, in the tenth century, in Southeast Asia saw the involvement of traders based in the Indonesian archipelago, around Java, Sumatra, Sulawesi, Borneo, on one hand, and the coastal areas of southern and eastern China on the other. As already mentioned, one of the focal points of these activities was the spice trade, especially in cloves, pepper, nutmeg, camphor, and cinnamon from the Maluku Islands – the earliest evidence of the use of cloves can be found in Chinese historiography. By far the largest network was established by Chinese, Arab, and Javanese traders in the port of Melaka (from Arabic *malakat* ‘market’), on the southwest coast of peninsular Malaysia, which was in full bloom in the fourteenth century. Over some 500 years of trade, multicultural and multilingual communities such as Melaka, characterized by religious freedom, shared political power between the dominant groups, mercenary armies, and clear legal codes, developed across the region. These trading centres proved particularly supportive to the emergence of contact languages of various types. These linguistic outcomes are the product not only of Malay-Chinese exchanges, but also the influences of Indian trading powers, Arabian networks, and the diverse migrant and indentured populations involved in the trade.

One further prominent player not only in the Southeast Asian region was and continues to be the British. Their arrival, following the Portuguese and Dutch, and the establishment of their exploitation colonies in Asia – in India, Malaya and the Pearl River Delta – introduced the English language into the mix. Its spread and penetration – through regular use and appropriation by the local multilingual communities – meant that it was only a matter of time before contact effects were evident.

2.2.2 *Malay contact varieties*

BAZAAR MALAY

Melaka was but one very significant city port, of the many trade centres spanning all the way from the west coast of Peninsular Malaysia through the Indonesian archipelago. Along these routes, a plethora of contact varieties emerged. Many of these new contact languages are believed to have evolved from a centuries-long widespread lingua franca of the region known as Bazaar Malay (Adelaar and Prentice, 1996; Reid, 2000). Numerous sources offer detailed accounts of Malay-speaking interpreters who routinely took part in the voyages across the region. For example, an Italian-Malay list compiled by Antonio Pigafetta in 1522 while travelling through Brunei and Tidore proves that colloquial Malay was in fact used as far as 2,000 kilometres away from the cultural centre of Melaka (Collins, 1996, p. 17). It is for these reasons that the Malay of the time has been compared to Latin in Europe. Throughout the seventeenth century, Malay was the dominant language of the Malay/Indonesian world and beyond, and it was in fact a prerequisite that any person engaged in the region have some understanding of it.

In discussing contact Malay varieties, it is important to note at the outset two main issues. First, it is doubtful whether a single well-defined pidgin Malay ever existed (Reineke et al., 1975). It is obvious that, in its role of lingua franca, it would be difficult to capture ‘Malay’ as one single uniform variety: after all, any language of interethnic communication spread over such a vast territory, and spoken by people of so many different linguistic backgrounds, might display a small core of similar features, but would otherwise be expected to show a lot of regional variation through diverse contact. Second, with regard to Bazaar Malay, one needs to recognize two different linguistic phenomena: (1) the historically attested diffuse Bazaar Malay lingua franca spoken across the Malay/Indonesian archipelago as a code of trade and interethnic communication; and (2) specific varieties of Bazaar Malay spoken today in different cities across Southeast Asia, which show different degrees of focusing and different typological traits depending on the speech community.

In constructing a structural profile of the Bazaar Malay lingua franca, the notion of Pidgin-derived Malay (PDM) developed by Adelaar (2005) is a useful construct for referring to what appear to be local Malay vernaculars and restructured varieties. The notion of PDM does not imply any firm claim as to the existence of an actual Malay pidgin as a precursor to other Malay contact languages, but must be interpreted in the sense of PDM *varieties* – in line with the diffuse nature of the lingua franca mentioned earlier – which are related by a cluster of mostly grammatical features that seem to be shared by a number of contact varieties across a vast geographical region and by many different ethnolinguistic groups. These include:

- (a) Possessive constructions based on *punya* (‘owner’ > ‘to possess’) [possessor + *punya* + possessee];
- (b) Very few productive morphological affixes compared to Literary Malay (i.e., isolating typology);
- (c) Pronouns of Sinitic, Hokkien origin, especially first and second person;
- (d) Plural pronouns formed with singular form + Malay *orang* ‘person’;
- (e) Reduced forms of demonstratives that precede the noun: *ini* ‘this’ and *itu* ‘that’;
- (f) Use of the Malay existential marker *ada* to indicate progressive aspect;
- (g) ‘Give’ and ‘make’ causative constructions, based on Malay *kasi* ‘to give’ and *buat* ‘to make’ preceding the verb;
- (h) Polyfunctional preposition *sama*.

The notion of PDM varieties has several merits, as follows:

- (i) It moves us away from the trap of ‘constructing’ a single linguistic code that perhaps never existed, as might also be the case for Lingua Francas of the Mediterranean (Selbach, 2007). Instead, PDM suggests the existence of a variety of similar linguistic codes, which arose for the purpose of commerce between speakers of different languages. This picture is confirmed by the variation found between speakers of different ethnic groups in accounts and analyses of Bazaar Malay (Lim, 1988, pp. 50–52).
- (ii) PDM features constitute an important ingredient of many contact languages of the region that evolved in precolonial and colonial times, such as Sri Lanka Malay, Cocos Malay and Baba Malay. These features are likely to be of ancient origin, as predicted by the founder principle, according to which, under particular evolutionary circumstances, older features gain in robustness and tend to be maintained in the histories of some languages (Mufwene, 1996).
- (iii) The significant Sinitic influence in some PDM features corroborates the idea that a not insignificant degree of Malay-Sinitic contact was involved in the early evolution of

PDM varieties. In this sense, the notion of PDM is useful because, by not assuming the existence of one unitary code as language of trade and interethnic communication, it allows us to consider a more sophisticated and realistic option.

- (iv) PDM varieties may be regarded as the outcome of various processes of language contact between different populations. Given that this was a region that, in precolonial times, had a very small population (Reid, 1980), the number and range of Malay contact languages, trade jargons, and pidgins as well as registers that are known to have existed to date, as well as their geographical spread, attest to the vast and central role that Malay must have had in the region, to the intensity and geolinguistic diversity of contacts between different populations, and to the complex sociolinguistic functions it served.

Several BZM varieties are still spoken predominantly at markets of the region, and when examining these contemporary BZM varieties, it is crucial to consider the specific ecologies in which they exist. For example, the BZM in use in Malaysia has been reported to fulfil a number of diverse functional needs beyond basic communication between speakers of mutually unintelligible languages, often by the non-English-educated as a language of interethnic communication, and shows interesting syntactic and phonological innovation, with lexical influences from different languages and much individual variation (Collins, 1984; Lim, 1988). BZM as spoken in Singapore has also been documented to have a wide range of functional uses, fully allowing for sophisticated types of communication (Khin Khin Aye, 2006).

BABA MALAY

Related to but in contrast with Bazaar Malay are contact Malay varieties, which are indigenous languages of specific communities, and which, as in-group vernaculars, are much more focused as a linguistic system. A classic example of this is found in the Babas or Peranakan Chinese communities – descendants of southern Chinese seafaring traders who settled in the Malay archipelago, primarily in Malacca and Penang in peninsular Malaysia, and in Singapore (British colonies on the Strait of Malacca and Strait of Singapore which were amalgamated in 1826 to form the Straits Settlements) from at least the seventeenth century and who married non-Muslim natives of the region, such as Balinese or Batak slaves. The Peranakans comprised one of the earliest and largest groups of the influential class of Chinese capitalists in the region. The new hybrid culture that emerged in this context shows unique traits that set the Peranakans apart from other Chinese, the more indigenous local populations, and other ethnically mixed groups, including a mixed cuisine consisting of Chinese culinary practices largely influenced by Malay traditions, and the wearing of Malay/Indonesian sarong and kebaya, instead of the Chinese dress, by the women, but the retention of Chinese rituals, such as religious practices and traditional wedding customs. (For detailed historical and sociolinguistics accounts of the community, see Tan, 1988; Rudolph, 1998; Ansaldo, Lim and Mufwene, 2007; Lim, 2010a).

Their vernacular, known as Baba Malay (BM), is a restructured variety of Malay with substantial southern Chinese, primarily Hokkien, influence, showing variation across regional BM communities (Pakir, 1986; Lim, 1988; Ansaldo and Matthews, 1999; Lee, 2018). Examples include a pronominal system deriving from Hokkien, and Sinitic word order, illustrated in (12), contrasting with Malay in (13).

12. *Gua punya bilik* Baba Malay
 1SG POSS room
 ‘My room.’

13. *Bilik saya* Malay
 room 1SG
 ‘My room.’

It has sometimes been said that drawing a clear distinction between Baba Malay (BM) and Bazaar Malay (BZM) varieties is not easy. In fact, early work (Shellabear, 1913; Reineke et al., 1975) have suggested that they are essentially the same language, which could have derived from a ‘restructured’ Malay developed by the Chinese in Malaysia, and with BM seen as a nativized variety of BZM. It is certainly acknowledged by most that they are closely related. This is not unexpected as both BZM and BM are closely associated to the history of Melaka and of the Peranakan communities found in such trading ports, with cross-fertilization between the varieties surely having occurred. Moreover, as merchants and mediators between Chinese interests and regional powers in precolonial times, and later as middlemen between those same regional powers and colonial interests, the Babas were an influential element in the ecology of BZM. Structurally, BZM and BM both manifest PDM features – though this in itself attests, as mentioned earlier, to the fact that PDM features may be recognized as forming the core, i.e., more stable and widespread, properties of contact Malay varieties across the region – with Singapore BZM and BM differentiated only in terms of phonology and lexicon, being comparable grammatically. The crucial factor may thus be identified as its ecology: while BM evolved as an in-group vernacular of communities of a homogeneous ethnic origin and thus is more focused and more readily identifiable as a discrete linguistic system, BZM varieties are typically used as lingua franca for interethnic communication, in different kinds of social networks than those associated with a specific ethnic group, and show high degrees of variation.

Recent online resources for these contact languages include, inter alia, the BZM data set in *APiCS* (Khin Khin Aye, 2013), and recordings of BM (Lee, 2013 (note: no files seem to be yet uploaded)) in the Kaipuleohone University of Hawai’i Digital Language Archive.

OTHER CONTACT MALAY VARIETIES

To provide a sense of the diversity of contact varieties of this region, and the roles they fulfilled, a few brief sketches of other Malay contact varieties are offered next (adapted from a synthesis in Ansaldo, 2009; a detailed and fuller list is found in Adelaar and Prentice, 1996, pp. 676–693), ranging from communities’ vernaculars to transient lingua francas, from the contemporary and widely used to the extinct. The differences between these varieties are often most marked at the phonological level, and peculiar regional grammatical influences can at times also be identified.

- (i) Chitty Malay: the vernacular of mixed South Indian-Malay descendants in Malaysia and Singapore;
- (ii) Sabah Pidgin Malay: used between Chinese and Sabahans, as well as between Sabahans of different ethnolinguistic backgrounds in East Malaysia;
- (iii) Jakarta/ Batavian/ Betawi Malay: spoken between different Indonesian groups, between Chinese and Indonesians, between masters and slaves, amongst Chinese belonging to different Sinitic groups, between Chinese and non-Chinese, and also by Eurasians when shifting from Portuguese to Malay. This was possibly already the mother tongue of Indonesian Peranakans, and may have been in existence since the eighteenth century; it is spoken today by one million people;

- (iv) Java Malay: used to be the language of lower administration; still widespread in Java;
- (v) Java Port Malay: trade language of the ports on the northern coast of Java, now extinct with no records available; it may have been the ancestor of Jakarta Malay;
- (vi) Tangsi Malay ('Barracks Malay'): used to be spoken by the Dutch East Indian army, and contained Dutch, and Ambon, Java, and Riau Malay elements; it was brought to the Netherlands by the Moluccan soldiers who had served in the Dutch colonial army and their families as *Melaju Sini* ('the Malay from here');
- (vii) *Papia Kristang* ('Christian Speak'): a Malay-Portuguese vernacular spoken in Melaka which still persists today and is also found in Singapore. The mixed language is likely to have arisen in mixed Malay-Portuguese households as the Portuguese had the widespread and state-encouraged practice of marrying their males to local women.

Two major observations should be made. First, clearly, the geographical reach of Malay in the region has been vast. It went as far east as Papua New Guinea, where, in the nineteenth century, it became the *lingua franca* of Chinese and Javanese workers on several plantations. There are also Malay loanwords in Hokkien and in Taiwanese texts from the seventeenth century. Its presence is attested also to the south, on the Cocos (Keeling) Islands and Christmas Island off the northwestern coast of Australia. To the west, features of PDM can be found in Sri Lanka Malay, in Malagasy, and in South Africa, where Cape Malay is claimed to have influenced Afrikaans. Second, it is also evident that these varieties were used by all ethnic groups. The picture that emerges of pre-modern Southeast Asia is one of highly diverse, multilingual, and interrelated linguistic ecologies. These are ecologies where speakers navigate between different languages, registers and *lingua francas* depending on the context and the purpose of their interaction, and in which both continuous transmission of traditional varieties coexists with contact-induced language change and language creation.

2.2.3 *Restructured English varieties*

Another group of language varieties in the Asian region of great significance in contact linguistics comprise the restructured varieties of English. These are the outcome of the introduction of the English language, in the numerous settlements of the British in Asia (as well as in Africa) during their expansion of empire in the seventeenth and eighteenth centuries. Crucially, such exploitation colonies involved not large-scale immigration as in the settlement colonies, but smaller numbers of British personnel in positions in government and education, with the local populations continuing to maintain their own vernaculars, and English introduced primarily in formal, educational contexts (but also see Section 2.3 for an example of an informal contact scenario).

One such restructured variety which has garnered much attention, for good reason, both in the fields of World Englishes and creole studies is Singapore English or Singlish, having evolved in a context which holds several critical ingredients for a keen observing of contact dynamics. This includes: (a) an ecology which has encompassed languages of diverse typologies, including Austronesian (several Malay varieties as already described), a range of Sinitic languages (both southern varieties as well as Mandarin), as well as several Dravidian and Indo-Aryan varieties from the South Asian subcontinent, a result of pre-colonial trading networks, as well as of the immigrants that were drawn to Singapore once it became a free

port during British colonization from 1819 onwards, and (b) post-independence language and immigration policies which have altered the ecology. In what follows we highlight two features which have come to characterize SgE and are also found in other contact English varieties in Asia – particles and tone.

Particles have long been recognized as a discourse-prominent feature, and consequently very easily transferred in contact-induced change (Matras, 2000). The substrates in the majority of Asian ecologies include particles in their grammars – languages such as Cantonese, Hokkien, Mandarin, Malay, Tagalog, and Hindi abound with discourse particles, which are used widely in those languages to communicate pragmatic functions of various types. It is not surprising then that, as a consequence of contact, particles figure as a characteristic feature in the New Englishes (Lim and Ansaldo, 2012; Lim, 2020). For example, several of Tagalog's 18 enclitic particles are consistently found in Philippine English, including *na*, which signals a relatively new or altered situation, *pa*, which denotes a relatively old or continuing situation, and *ba*, a question marker obligatory in formulaic yes-no questions (Lim and Borlongan, 2011), and many of the particles from the rich inventory found in Cantonese are widely used in Hong Kong English, such as *a*, *ma*, and *wor* (James, 2001; Chen, n.d.). Most revealingly, Hindi particles such as *yaar* and *na* occur in Indian English regardless of a speaker's mother tongue (Lange, 2009).

But it is the particles in Singapore English which are especially instructive for several reasons, as investigated comprehensively in Lim (2007) and summarized here. Two strands of evidence – the sociohistorical and the structural – come together to account for their origins. On the one hand there are the *lah* and *ah* particles, which have been documented in SgE (and Malaysian English) since the 1970s, and demonstrated to have origins in Bazaar Malay and Hokkien (Gupta, 1992; Platt, 1987; Lim, 2007), which were the languages dominant in the ecology in that era, as inter- and intra-ethnic lingua francas respectively. A larger set of particles, comprising *hor*, *leh*, *lor*, *ma*, and *meh*, only started being observed in SgE in the late 1980s, several illustrated in examples (14) and (15).

14. My parents old fashion *a21*? Then your parents *le55*? (Lim, 2007, p. 451)
 'Are you saying that my parents are old-fashioned? Then what about your parents?'
15. No *la21*! He's using Pirelli, you don't know *me55*? (Lim, 2007, p. 451)
 'No, he has Pirelli tyres; didn't you know that?' [incredulously]

While the forms and functions of particles of the main dominant southern Chinese varieties are comparable, it is Cantonese which shows a perfect match with these later SgE particles in all three domains of (a) segmental and (b) suprasegmental forms, and (c) functions. Moreover, it is in the 1980s and 1990s that Cantonese can be seen to attain prominence in Singapore's ecology, both in popular culture – it being the golden years of Cantopo and Hong Kong cinema – and in demography – when Hong Kong was a target source for skilled migrants (Lim, 2010b). Together these stand to explain Cantonese's dominance in the feature pool and thus the acquisition of Cantonese particles into SgE in this later era.

The second feature in SgE worth examining is tone. Recognition of the evolution of tone in a contact language variety due to the presence of tone in the broader linguistic environment, viz. in the tone-language substrate(s), has for some time now been recognized, but mostly in contact situations involving European accent languages and African tone languages, including

Saramaccan (Good, 2004a, 2004b, 2006), Papiamentu (Kouwenberg, 2004; Rivera-Castillo and Pickering, 2004), Pichi or Fernando Po Creole English (Yakpo, 2009), and Nigerian English (Gut, 2005). Although most Southeast and East Asian contexts boast ecologies where tone languages are dominant, such contact outcomes in Asian contact varieties have only recently been documented. In Singapore, Hokkien was prominent as the Chinese intraethnic lingua franca and a widely used interethnic lingua franca in colonial and early independence eras, with Mandarin subsequently assuming importance as one of the nation's four official languages, and Cantonese seeing a resurgence in the late 1980s and 1990s, thanks to Cantonese popular culture and significant immigration from Hong Kong. Due to demographic, political, and social factors, tone languages have thus been dominant in the ecology, making tone salient in the feature pool, and SgE exhibits (Sinitic-type) tone, at the level of the word and phrase² (Lim, 2009, 2011).

But the significance of tone in the contact dynamics extends further. In SgE is noted a systematic pattern of word- and phrase-final prominence, as illustrated in (17) and (18). This, notably, contrasts with the prosodic patterning found in all other contact varieties in which tone has evolved, such as Nigerian English and Hong Kong English, where the general pattern locates High (H) tones on what would be stressed or accented syllables at word- or phrase-level, and Low (L) tones on unstressed ones, as shown for HKE in (18) and (19).

- | | | |
|------------------------------------|------------|--------------------------|
| 16. `manage, `teacher | MH | (SgE, Wee, 2008, p. 490) |
| in `tend, a `round | LH | |
| `origin, bi `lingual | LMH | |
| o `riginal, se `curity | LMMH | |
| 17. I think happier | LHLLM | (SgE, Lim, 2004, p. 44) |
| 18. in `tend | LH | (HKE, Wee, 2008, p. 488) |
| `origin, `photograph | HLL | |
| o `riginal | LHLL | |
| 19. I saw the manager this morning | LHHHHHHHL! | (HKE, Luke, 2008) |

SgE's apparently uncharacteristic prosody can be explained – and this is the other instructive element of this variety. This involves both acknowledging the Founder Principle in the ecology paradigm (Mufwene, 2001, 2008), viz. that the founder population in an ecology exerts a strong influence on features, an influence which persists in the emergent variety, as well as recognizing the Peranakans as a founder population, one with significant economic and social prominence in Singapore's ecology. Structural analysis aligns with evolutionary and sociolinguistic facts. The Peranakans' vernacular Baba Malay displays word- and phrase-final prominence – the right-edge phrasal prominence that is documented in many Malay varieties. This prosodic pattern also developed in the Peranakans' variety of English (Lim, 2011, 2014, 2016). The Peranakans, though a small minority, were clearly dominant in the ecology, as outlined earlier, due to their political, economic and social status, and their position as intermediaries, and later as teachers. As early English adopters, crucially during the British colonial period, theirs would have been the early features influencing the particular prosody in the emerging variety of SgE.

2.3 East Asia

2.3.1 Contact in the Pearl River Delta

In the Far East of the region under discussion, we see a shorter period of intense multicultural exchange as well as much more restricted networks of interaction. Japan was largely unaffected, in cultural and linguistic terms, by the Western colonial powers that ravaged the Chinese mainland. Despite being humiliated and abused by Western powers for centuries, China also managed in its typical fashion to remain highly unaffected by external influences. There is however one case of language contact in this region – in southern China – that is unique in terms of the wealth of data that have been gathered and the descriptions that have been put forward: Chinese Pidgin English, or China Coast Pidgin, is perhaps the best known variety of its type. Its development may be related to the earliest Western presence in China, the Portuguese, and by those that followed, in particular the British. (For a comprehensive, incisive account of history and grammatical description, see Ansaldo, Matthews and Smith, 2010.)

The Portuguese established themselves in Melaka in 1511, and used it as a base for further expansion east, in particular to Macau on the Pearl River Delta in southern China (Guangdong). By the sixteenth century, there were approximately 500 to 600 ‘Portuguese’ merchants, and by 1635, it was estimated that there were 1,700 *casados* in Macau, i.e., married Christians, of which half were ‘white’ and half were ‘black.’³ From Macau the Portuguese started trading up the Pearl River Delta and were a pioneering force in the China-West trade.⁴ From the latter part of the seventeenth century, the British presence started growing in southern China as a consequence of the expansion of the East India Company, and this, together with the arrival of other Western companies and the shift in power from Portugal to Britain, led to the establishment of a new base of trade, in which English would replace Portuguese as *lingua franca*.

In the first half of the eighteenth century, the city of Canton (today’s Guangzhou) and the river delta on which it sits were a zone of intense multicultural exchanges. In order to exercise a form of control over the activities of Western trading companies and their local Chinese partners, the ‘Canton System’ was developed as a set of protocols to safeguard the passage from Macau to Canton, which required boats to sail up the Pearl River into Canton, providing authorities with an ideal passage that was easy to control (Van Dyke, 2005). Chinese merchant families were appointed by the government to act as brokers and superintendents of the foreign traders, with each foreign ship the responsibility of one Chinese family. These ‘security’ merchants formed a guild at the command of a customs superintendent who answered directly to the Emperor. Their control and taxing of the foreign cargoes proceeded in three steps: foreign boats were first required to call at Macau, and were then escorted up the river to Canton; 20 kilometres from Canton, at the harbour of Whampoa (Huangpu), the boats stopped, and cargo was unloaded onto smaller vessels; from Whampoa, the chief trader of the foreign boat, the Taipan, and the cargo finally moved to Canton. In Canton, the factories, i.e., foreign merchant houses, took care of the trade between Chinese and Western merchants.

These contexts of navigation and trading involved contact between parties with no common language. The evidence of this lies in the existence of what is possibly the best documented pidgin known to scholars, Chinese Pidgin English, also known as China Coast Pidgin. Attestations of CPE are found in English-language sources, and in the form of a series of booklets written by Chinese in Chinese characters for the purpose of teaching and learning this trade language.⁵

2.3.2 *A sketch of Chinese Pidgin English*

A number of features characterize Chinese Pidgin English (CPE); a selection is summarized here.

- (a) Lexicon are derived mostly from English, with some Portuguese, Malay, Hindi and, as would be expected from the immediate ecology, Cantonese influences, as seen in the examples in (20) (Martino, 2003).

20. pidgin < business (English)
 catchee < catch ‘fetch’ (English)
 two muchy < too much ‘extremely’ (English)
 joss < Deus (Portuguese ‘God’)
 sabbee < saber (Portuguese ‘to know’)
 chop < chapa (Malay, from Hindi ‘chop, stamp’)
 taipan (Cantonese ‘supercargo,’ lit. ‘big class’)
 fankuei (Cantonese ‘Westerner,’ lit. ‘foreign devil’)

- (b) The structure of the NP reveals variation between Sinitic and non-Sinitic constituent order, as seen in the typically Sinitic patterns [NUM-CL-N] and [DEM-CL-N], as in Cantonese (21); also a frequent pattern is the typical Sinitic use of the classifier, in CPE manifested as *piece*, illustrated in (22).

21. Yāt go yàhn Nī go yàhn
 1 CL man DEM CL man
 ‘One man.’ ‘This man.’

22. You wantchee catchee one piecee lawyer (*Instructor* IV.32)

- (c) The pronouns of CPE have attracted some attention in the literature. While second and third person forms were almost invariant, there was considerable variation in first person forms. In particular, besides what appear to be regular English forms, in a first phase, three first person singular forms were found, *I*, *my* and *me*, all three of which could be found in subject position, while the latter two were also used in object position (Baker and Mühlhäusler, 1990, p. 104; Tryon, Mühlhäusler and Baker, 1996, p. 488). Eventually, only *my* emerged in both functions – this may be seen as evidence of stabilization in the grammatical system of CPE.

- (d) Copula constructions are relatively rare, as CPE tends to be zero-copula, like Sinitic languages in general, as in (23) (Hall, 1944; Selby and Selby, 1995, p. 138).

23. Englishman very good talkee; all heart bad, – no talkee true – too much a proudy

Where copula is used, there are at least two different types: *have* in (24), usually realized as *hab*, which can also be used as a possessive verb as well as an aspectual marker, marking perfective in (25), and *belong*, with the latter taking over as a copula while the aspectual usage of *hab* increased (Baker and Mühlhäusler, 1990, p. 103). The use of *hab* has a basis in Cantonese as well as English grammar, where Cantonese uses existential verb *yáuh* ‘have.’ Eventually, *hab* and *habgot* were used in possessive constructions. Most occurrences of *belong* involve its use in its lexical function ‘to belong to,’ as in (26) and (27).

24. Chinese man very great rogue truly, but have fashion, no can help
‘Chinese men are real rogues but that’s how it is, can’t help it.’
(Anon 1748, in Baker and Mühlhäusler, 1990, p. 103).
25. My hap go court one time (*Instructor* IV.4)
‘I have been to court once.’
26. These belong you? (*Instructor* IV.53)
‘Is this yours?’
27. The tea belong first crop (*Instructor* VI.14)
‘This is first crop tea.’
- (e) The verb *got* is very similar to the Cantonese verb *yáuh*, which can indicate location, possession and existence, as in (28).
28. You got how muchee piecee children (*Instructor* IV.55)
‘How many children have you?’
- (f) The ‘X-side’ construction is calqued on Cantonese, illustrated in (29) and (30).
29. bring that egg come thisee side (*Instructor* VI: 40)
‘Bring the eggs here.’
30. come Sydney side (*Instructor* VI: 32)
‘(She) comes from Sydney.’

This is an interesting feature as it is one of the few cases in which CPE expressions have found their way into the present-day Hong Kong variety of English: commonly heard in Hong Kong English are phrases such as ‘Kowloon side,’ i.e., ‘over in Kowloon.’

3. Future directions

3.1 Globalization and contact

Several phenomena associated with globalization have been recognized in recent decades, including the decline of the (British) Establishment, developing ethnic pluralism, especially in urban settings, and the proliferation and speeding up of communication technologies (a full list and discussion can be found in Lim and Ansaldo, 2016). While such processes are not new in substance, they are acknowledged as being new in intensity, scope and scale, and hold implications for bringing communities and their languages into contact with each other, in new, contemporary contexts distinct from those of the past, with significant outcome. In this regard, the Asian region presents itself prominently for two main reasons, as argued in Lim and Ansaldo (2016). The first factor regards the phenomenon that, with the persistence of certain demographic and economic trends, the twenty-first century will be dominated by Asian politics and culture.⁶ Indeed, as is well recognized, there has been a great shift in the global economy’s centre of gravity from West to East, which entails various phenomena of globalization and economic growth in Asia, such as the pursuit of linguistic capital, mobility, trade, communication technology, etc. The second factor concerns the fact that, with English one of the major languages of global communication today, Asia holds a prominent position as the site for the largest and most quickly growing number of users of English: the total English-using population in Asia is some 600 million, far more than English speakers in the ‘Inner Circle.’ These

factors together present the scenario of multilingual speakers of diverse languages plus Asian Englishes (themselves contact varieties) interacting with other such speakers, and the significant implications for language contact in terms of bringing together communities and their languages in new dynamics, providing new sites and potential for consequences of contact. In what follows, two contemporary ecologies are highlighted: the dynamics and outcomes in computer-mediated communication (CMC) and in modern trading posts (more comprehensive and detailed accounts in Lim and Ansaldo, 2016).

3.2 *Computer-mediated communication*

Already some two decades ago, scholarship started recognizing the emergence of new ways of expression as a consequence of the rise of electronic media: new literacies which are shaped by the opportunities and constraints of the electronic medium. The platform has afforded flexibility and creativity of expression, including the exploiting of multilingual repertoires, notably in situations involving an emergent English and languages using different orthographic traditions. This is because, although advances have certainly been made and continue to be made in developing keyboards for various scripts, such as Chinese characters or Devanagari script, users very often prefer to use a Latin-based keyboard, and/or English, due to the constraints of the keyboard or the comparative efficacy compared to using character keystrokes. This presents not so many challenges as opportunities for language contact, and the resultant evolution of varieties.

Young Hongkongers, for instance, who are, as all Hongkongers, normally Cantonese-dominant in all other domains, overwhelmingly agree that English is easier as an input in CMC than Chinese, and report a significant preference for using English, or English and Cantonese, rather than Chinese. Clearly, CMC promotes significantly greater English usage than what there would normally be for a community dominant in another language, and this has two major consequences.

First, CMC platforms comprise a site quite distinct from the community's usual communicative practices: here there is more widespread use of English than in non-CMC contexts, and consequently more frequent mixing of codes – in the case of Hong Kong, of Cantonese and English. The outcome of such language contact is illustrated in the following examples of online chat (from Wong, 2009). There is romanization of common Cantonese phrases, such as *mafan* 'troublesome' for 麻煩 *maa4faan4* (31, turn 5), and relexification, such as *gum is you dun ask* (32, turn 3), and *or . . . gum you continue lo* (32, turn 5).

31.

- 1A: head ask for resume??
 'The department head asked you for your resume?'
- 2A: how come ge
 'How come [ge2]?'
- 3B: yes ar
 'Yes [aa3].'
- 3B: he said he ask all people la wor
 'He said he had asked everyone for their resume already [aa3 wo5].'
- 4A: what for
 'What is that for?'
- 5A: ma fan
 'It's so troublesome.'

- 6B: not my head
'He is not my supervisor.'
- 7B: programmer head
'My supervisor is the head of the programming department.'

32.

- 1A: did u ask Wilson to pick you up in the train station?
2B: ah . . . not yet . . . hahaaa
3A: gum is u dun ask. . .
 咁係你唔問
 gam2 hai6 nei5 ng4 man6
 'Then it's you who don't ask him to pick you up.'
4A: dun say wt danger later ar . . . ghaa
 唔好話咩危險一陣呀
 ng4 hou2 waa6 me1 ngai4 him2 jat1 zan6 aa3
 'Don't say it is dangerous later (*laugh).'
- [. . .]
5A: or . . . gum u continue lo
 哦咁你繼續囉
 ngo4 gam2 nei5 gai3 zuk6 lo1
 'Ok . . . then you continue working on your assignment [lo1].'

One outcome of such contact is worth a closer look, the direct translation or calquing of Cantonese expression 加油 *galyau4* 'add oil,' into English 'add oil,' by younger Cantonese-English bilingual Hongkongers. As a general exhortation or cheer to persevere or to work hard, Cantonese 加油 *galyau4* is widely used, as seen in spoken Cantonese discourse in (33) and in CMC in (34) (Lim, 2015).

33.

- A: *Ngo chin gei yat sin tong kui lao yuen gao*
'I argued with him just a few days ago.'
B: *Hah?* Again? For what?
A: You know, just like *zi chin gor d lor*
'You know, just like what happened before.'
B: *Ai, kui d temper really. . . gayau ah!*
'Sigh, his temper is really bad . . . be strong!'

34.

- A: *Doin meh?*
'What are you doing?'
B: *Hea gun ah, u?*
'Just taking some rest, and you?'
A: Gonna finish some readings. Need slp earlier, tmr *faan gong*
'I'm going to finish some readings and need to sleep earlier. I need to work tomorrow.'
B: Oh *hai wor, ho chur ah, gayau!*
'Oh right, you're so busy. Just hang in there!'

While in CMC, Cantonese 加油 is used less 'regularly' than in spoken communication, the use of the English calque *add oil* increases significantly to 'quite often' whether texting

in Cantonese, or in English or Cantonese-English, in fact, more than its original Cantonese expression (Lim, 2015), illustrated in (35) (Wong, 2009) and (36) (Lim, 2015).

35.

- A: 7.00 am. . .
'I have to work at 7.00 am.'
A: very sh*t le
'It's very bad [ne].'
B: ahaha ~~~ *add oil!*
'[laugh] work hard!'
B: Then goodnight and sweet dreams la
A: talk to you next time

36.

- A: Are you ready for tomorrow's Chinese test?
B: Not yet. Mom's forcing me to drink bedtime milk.
A: Then you should probably sleep too. *Add oil* for the test.
B: Yeah.

This is a significant finding in itself – a CMC platform does enable language contact and prompt the development of HKE, in this case, in the use of particular HKE phrases, here calqued from Cantonese. An examination of microblogging sites such as Instagram, Twitter, and Tumblr for the hashtag #addoil turns up infinite numbers of posts. Clearly, such an innovation is already widely used in the community, in CMC.

Second, and even more significantly, more English-dominant bilinguals – e.g., Hongkongers who emigrated several years ago and then returned to Hong Kong, or Hongkongers of mixed parentage – exhibit a different pattern, where English *add oil* is used significantly more often when speaking. In other words, this feature appears to have spread from CMC to non-CMC domains. In effect, the increased use of English in the CMC domain comprises a drive in the direction of the community employing English in the bilingual mix to a greater extent, first in that domain, and then in others, which is the road to further nativization of a restructured New English in a contact context, and subsequent endonormative stabilization. CMC is identified as one of the forces in the knowledge economy that can drive the evolution of a new variety (Lim, 2015).

3.3 *Twenty-first-century trading relations*

While Sino-African trade dates as far back as the Tang Dynasty, the twenty-first century has witnessed increasingly strong economic ties between China and Africa, through trade relations, investments, and aid on the African continent. China is currently Africa's largest trade partner and the biggest investor in Africa. One million Chinese citizens are residing in Africa, and some 500,000 Africans are working in China, with the largest African diaspora found in Guangzhou, the commercial and trading centre of south China; it is estimated that there are some 100,000 Africans living and doing business in Guangzhou (Bodomo, 2010). Recent research on language use in these communities shows that, in a survey of 120 Africans and Chinese in Guangzhou, the most frequently used language between the two communities is overwhelmingly English (80.3%), with the majority saying they use English most or all of the

time (Liu, 2013). More crucially, it is African and Chinese varieties of English that are documented in use (Bodomo, 2010; Liu, 2013). One prominent feature documented is reduplication, as in the examples in (37) – a feature that is common cross-linguistically, found notably in African and Sinitic languages, which has emerged in these new Englishes.

37. big-big
 same-same
 talk-talk
 chop-chop

What is striking about these more contemporary contact dynamics, as observed in Lim and Ansaldo (2016) and Lim (2020), is that the impact of language contact extends further than the impact of the traditional substrates on a lexifier. The reduplicated phrase *chop-chop* has long been adopted in English, meaning ‘quickly,’ with its origins in Chinese Pidgin English meaning ‘quick’ (Holm, 1988, p. 516), from Cantonese 急急 *gāp gāp* ‘quick.’ In the linguistic practices observed amongst the traders in Guangzhou, however, it is used, by both Africans and Chinese alike, to mean ‘to eat’ (Liu, 2013). In various West African Pidgin English varieties, such as Nigerian Pidgin English, Ghanaian Pidgin English and Pichi, the meaning of (non-reduplicated) *chop* is ‘to eat, food’ (Faraclas, 1996; Blench, 2006; Yakpo, 2009), whose origin is suggested to lie in the obsolete English verb *chap* or *chop* ‘to take in the chops and eat’ (Huber, 1999, p. 99). It would appear that in this contemporary contact situation, the more stabilized African English varieties are exerting more influence on the newer Chinese Englishes (Lim and Ansaldo, 2016; Lim, 2020), and it remains to be seen how these dynamics will play out in the emerging contact varieties.

4. Closing remarks

The Asian region and the contact situations which have transpired therein through history are clearly numerous and diverse. This chapter has selected Sri Lanka, Melaka and Singapore, and Macau, as exemplifications of the ecologies of South, Southeast, and East Asia respectively, in order to highlight distinct and important lessons in contact dynamics, processes, and outcomes. The examples of contact that are occurring in the current contemporary era provide a taste of novel dynamics and outcomes that wait to be further explored in this rich and fascinating region.

5. Further reading

Ansaldo, U. (2009). *Contact languages: Ecology and evolution in Asia*. Cambridge: Cambridge University Press.

This book presents an evolutionary theory of contact languages and language formation, with focus on Monsoon Asia.

Lim, L. and Ansaldo, U. (2016). *Languages in contact*. Cambridge: Cambridge University Press.

This book comprises a more sociolinguistic account, with particular attention to Asia, and also encompassing contemporary contexts.

Matras, Y. (2009). *Language contact*. Cambridge: Cambridge University Press.

This book provides perhaps the most comprehensive and thorough approach to the theory, methodology and analysis of the field.

6. Related topics

Social factors, cognitive factors, typological factors, bilingual language acquisition, borrowing, convergence, creoles and pidgins, mixed languages, urban multilingualism

Abbreviations

| | |
|--------|---------------------------------------|
| 2 | reduplication |
| ABL | ablative |
| ACC | accusative |
| ADJ | adjective |
| BM | Baba Malay |
| BZM | Bazaar Malay |
| CL | classifier |
| CMC | computer-mediated communication |
| CP | conjunctive particle |
| CPE | Chinese Pidgin English |
| DAT | dative |
| DEM | demonstrative |
| DOBES | Documentation of Endangered Languages |
| DUR | durative |
| FOC | focus |
| H | high tone |
| HKE | Hong Kong English |
| INSTR | instrumental |
| INTERR | interrogative |
| L | low tone |
| Lex | lexeme |
| M | mid tone |
| N | noun |
| NEG | negation |
| NUM | numeral |
| NMLZ | nominalizer |
| PAST | past tense marker |
| PDM | Pidgin-derived Malay |
| PL | plural |
| POSS | possessive |
| SG | singular |
| SgE | Singapore English |
| SLM | Sri Lanka Malay |
| SLP | Sri Lanka Portuguese |
| SSA | South-South Asia |

Notes

- 1 It is well documented that ‘almost all the major ethnic groups from the region of the Eastern archipelago were represented’ Hussainmiya (1987, p. 48), and the peoples were known as *Ja Minissu* by the Sinhalese and *Java Manasar* by the Tamils ‘people from Java’ (Saldin, 2003, p. 3). It was the British

who, upon finding a community who spoke ‘Malay,’ attached the corresponding ethnic label to the group, and it is this designation ‘Malay’ that has persisted.

- 2 Here tone accents are used as in the sources, where L = Low tone, M = Mid tone, and H = High tone.
- 3 Such a classification was common throughout Portuguese Asia and distinguished between the married Christians of Portuguese origin, mostly of mixed ethnic descent, and the married Christians of indigenous origin (Bethencourt, 2005, pp. 116–117).
- 4 Chinese authorities imposed a strict segregation between Macau and the Chinese population in the surrounding areas – this explains why the local patois Makista (not within the purview of this chapter) has limited, and mostly lexical, influences from Chinese, and is otherwise very similar to Papia Kristang of Melaka.
- 5 Among the many sources, two are known to us today: (1) Hühng mòuh tùng yuhng fân wá (紅毛通用番話) ‘The Language of the Redhaired Foreigners’ (anonymous c.1835) published in Canton around the fourth decade of the nineteenth century, and (2) Ying yúh jaahp chyùhn (英語集全), the much less widely available six-volume work *Chinese and English Instructor* (henceforth *Instructor*) (Tong, 1862), written by Tong King-Sing around 1862.
- 6 The idea of the ‘Asian Century,’ a term attributed to a 1988 meeting with PRC leader Deng Xiaoping and Indian PM Rajiv Gandhi, is in counterpoint to the preceding ‘American Century,’ coined by *Time* publisher Henry Luce used to describe the dominance of the USA of much of the twentieth century, in contrast to the preceding nineteenth century as the ‘British Century.’

References

- Adelaar, K.A. (2005). Structural diversity in the Malayic subgroup. In: K.A. Adelaar and N.P. Himmelmann, eds., *The Austronesian languages of South East Asia and Madagascar*, 1st ed. London: Routledge Curzon, pp. 202–226.
- Adelaar, K.A. and Prentice, D.J. (1996). Malay: Its history, role and spread. In: S.A. Wurm, P. Mühlhäuser and D. Tryon, eds., *Atlas of languages of intercultural communication in the Pacific, Asia and the Americas*, 1st ed. Berlin and New York: De Gruyter Mouton, pp. 673–693.
- Anon (c. 1835). Hühng mòuh tùng yuhng fân wá (紅毛通用番話) [The common language of the red-haired foreigners]. Guangzhou.
- Ansaldò, U. (2009). *Contact languages*. Cambridge: Cambridge University Press.
- Ansaldò, U. (2011). Metatypy in Sri Lanka Malay. In: G. Sharma and R. Singh, eds., *Annual review of South Asian languages and linguistics 2011*. Berlin: De Gruyter, pp. 3–16.
- Ansaldò, U. (2015). Sri Lanka and South Asia. In: R. Hickey, ed., *Cambridge handbook of areal linguistics*, 1st ed. Cambridge: Cambridge University Press.
- Ansaldò, U., Lim, L., Bisang, W. and Nordhoff, S. (2004–2009). *Sri Lanka Malay corpus*. Documentation of Endangered Languages (DOBES) Archive. [online] Available at: <http://dobes.mpi.nl/projects/slm/>.
- Ansaldò, U., Lim, L. and Mufwene, S.S. (2007). The sociolinguistic history of the Peranakans: What it tells us about ‘creolization.’ In: U. Ansaldò, S. Matthews and L. Lim, eds., *Deconstructing creole*, 1st ed. Amsterdam and Philadelphia: John Benjamins, pp. 203–226.
- Ansaldò, U. and Matthews, S. (1999). The Minnan substrate and creolization in Baba Malay. *Journal of Chinese Linguistics*, 27(1), pp. 38–68.
- Ansaldò, U., Matthews, S.M. and Smith, G. (2010). China Coast Pidgin: Texts and contexts. In: U. Ansaldò, ed., *Pidgins and creoles in Asian contexts*, Special Issue, *Journal of Pidgin and Creole Languages*, 25(1), pp. 63–94.
- Bakker, P. (2006). The Sri Lanka Sprachbund: The newcomers Portuguese and Malay. In: Y. Matras, A. McMahon and N. Vincent, eds., *Linguistic areas: Convergence in historical and typological perspective*. Houndmills, Basingstoke, and New York: Palgrave Macmillan, pp. 135–159.
- Bakker, P. and Mous, M., eds. (1994). *Mixed languages: 15 case studies in language intertwining*. Amsterdam: IFOTT.
- Baker, P. and Mühlhäusler, P. (1990). From business to pidgin. *Journal of Asian Pacific Communication*, 1(1), pp. 87–115.

- Bethencourt, F. (2005). Lowcost Empire: Interaction between Portuguese and local societies in Asia. In: E. van Veen and L. Blussé, eds., *Rivalry and conflict: European traders and Asian trading networks in the 16th and 17th centuries*, 1st ed. Leiden: CNWS Publications, pp. 108–130.
- Blench, R. (2006). *Archaeology, language, and the African past*. Lanham, MD: AltaMira Press.
- Bloch, J. (1965). *Indo-Aryan: From the Veddas to modern times*. Paris: Librairie d’Amerique et d’Orient. Adrienne – Maisonneuve.
- Bodomo, A. (2010). The African trading community in Guangzhou: An emerging bridge for Africa-China relations. *China Quarterly*, 203, pp. 693–707.
- Cardoso, H.C. (2017). *Documentation of Sri Lanka Portuguese*. London: SOAS, Endangered Languages Archive. [online] Available at: <https://wurin.lis.soas.ac.uk/Collection/MPI1035102> [Accessed 2 Nov. 2018].
- Cavalli-Sforza, L.L., Menozzi, P. and Piazza, A. (1994). *The history and geography of human genes*. Princeton, NJ: Princeton University Press.
- Chen, K. (n.d.). *Multilingual Hong Kong corpus*. Available at: <https://sites.google.com/view/katherine-chen/research/multilingual-hong-kong>.
- Collins, J.T. (1984). Malaysian and Bazaar Malay: Polarity, continuity and communication. In: A.H. Omar, ed., *National language and communication in multilingual societies*, 1st ed. Kuala Lumpur: Dewan Bahasa dan Pustaka, pp. 151–174.
- Collins, J.T. (1996). *Malay, world language: A short history*. Malaysia: Dewan Bahasa dan Pustaka.
- Emeneau, M.B. (1956). India as a linguistic area. *Language*, 32(1), pp. 3–16.
- Erdosy, G. (1995). *The Indo-Aryans of ancient South Asia: Language, material culture and ethnicity*. Berlin: De Gruyter.
- Faraclas, N. (1996). *Nigerian Pidgin*. London: Routledge.
- Gair, J.W. (1994). Universals and the South-South Asian language area. *Keynote address, 1994 SALA meeting*. Philadelphia.
- Gair, J.W. (1998). *Studies in South Asian linguistics: Sinhala and other languages*. New York: Oxford University Press.
- Gair, J.W. (2013). Sri Lankan languages in the South-South Asia linguistic area: Sinhala and Sri Lanka Malay. In: S. Nordhoff, ed., *The genesis of Sri Lanka Malay*. 1st ed. Leiden: Brill, pp. 165–194.
- Good, J. (2004a). Tone and accent in Saramaccan: Charting a deep split in the phonology of a language. *Lingua*, 114, pp. 575–619.
- Good, J. (2004b). Split prosody and creole simplicity: The case of Saramaccan. *Journal of Portuguese Linguistics*, 3, pp. 11–30.
- Good, J. (2006). The phonetics of tone in Saramaccan. In: A. Deumert and S. Durrleman, eds., *Structure and variation in language contact*, 1st ed. Amsterdam and Philadelphia: John Benjamins, pp. 9–28.
- Gupta, A.F. (1992). The pragmatic particles of Singapore colloquial English. *Journal of Pragmatics*, 18, pp. 31–57.
- Gut, U. (2005). Nigerian English prosody. *English World-Wide*, 26(2), pp. 153–177.
- Hall, R.A. (1944). Chinese pidgin grammar and texts. *Journal of the American Oriental Society*, 64(3), pp. 95–115.
- Holm, J. (1988). *Pidgins and creoles, vol. 1: Theory and structure*. Cambridge: Cambridge University Press.
- Huber, M. (1999). *Ghanaian Pidgin English in its West African context*. Amsterdam and Philadelphia: John Benjamins.
- Hussainmiya, B.A. (1987). *Lost cousins: The Malays of Sri Lanka*. Bangi: Penerbit Universiti Kebangsaan Malaysia.
- James, G. (2001). Cantonese particles in Hong Kong students’ emails. *English Today*, 17(3), pp. 9–16.
- Khin Khin Aye (2006). *Bazaar Malay: History, grammar and contact*. PhD. National University of Singapore.
- Khin Khin Aye (2013). Bazaar Malay data set. In: S. Michaelis, P. Maurer, M. Haspelmath and M. Huber, eds., *The atlas of pidgin and creole language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology. [online] Available at: <http://apics-online.info/contributions/67>.

- Kouwenberg, S. (2004). The grammatical function of Papiamentu tone. *Journal of Portuguese Linguistics*, 3, pp. 55–69.
- Krishnamurti, B. (1969). Comparative Dravidian linguistics. In: *Current trends in linguistics*, vol. 5. The Hague: Mouton, pp. 309–330.
- Kshatriya, G.K. (1995). Genetic affinities of Sri Lankan populations. *Human Biology*, 67(6), pp. 843–866.
- Lange, C. (2009). Discourse particles in Indian English. In: T. Hoffmann and L. Siebers, eds., *World Englishes: Problems, properties, prospects*, 1st ed. Amsterdam and Philadelphia: John Benjamins, pp. 207–226.
- Lee, N.H. (2013). Nala Lee collection [Baba Malay media files, transcripts, and field notes]. *Kaipuleohone University of Hawai'i digital language archive*. Available at: <https://scholarspace.manoa.hawaii.edu/handle/10125/25950>.
- Lee, N.H. (2018). Baba Malay: Diverging trends in two ecologies. *Journal of Pidgin and Creole Languages*, 33(1), pp. 135–172.
- Lim, J.H. and Borlongan, A.M. (2011). Tagalog particles in Philippine English: The cases of *ba*, *na*, *no*, and *pa*. *Philippine Journal of Linguistics*, 42, pp. 59–74.
- Lim, L. (2004). Sounding Singaporean. In: L. Lim, ed., *Singapore English: A grammatical description*. (Varieties of English Around the World G33), 1st ed. Amsterdam and Philadelphia: John Benjamins, pp. 19–56.
- Lim, L. (2007). Mergers and acquisitions: On the ages and origins of Singapore English particles. *World Englishes*, 27(4), pp. 446–473.
- Lim, L. (2009). Revisiting English prosody: (Some) New Englishes as tone languages? *The Typology of Asian Englishes*, Special Issue, *English World-Wide*, 30(2), pp. 218–239.
- Lim, L. (2010a). Peranakan English in Singapore. In: D. Schreier, P. Trudgill, E.W. Schneider and J.P. Williams, eds., *The lesser-known varieties of English*, 1st ed. Cambridge: Cambridge University Press, pp. 327–347.
- Lim, L. (2010b). Migrants and ‘mother tongues’: Extralinguistic forces in the ecology of English in Singapore. In: L. Lim, A. Pakir and L. Wee, eds., *English in Singapore: Modernity and management*. (Asian Englishes Today), 1st ed. Hong Kong: Hong Kong University Press, pp. 19–54.
- Lim, L. (2011). Tone in Singlish: Substrate features from Sinitic and Malay. In: C. Lefebvre, ed., *Creoles, their substrates and language typology*. (Typological Studies in Language 95), 1st ed. Amsterdam and Philadelphia: John Benjamins, pp. 271–287.
- Lim, L. (2014). Yesterday’s founder population, today’s Englishes: The role of the Peranakans in the (continuing) evolution of Singapore English. In: S. Buschfeld, T. Hoffmann, M. Huber and A. Kautzsch, eds., *The evolution of Englishes*. (Varieties of English Around the World G49), 1st ed. Amsterdam and Philadelphia: John Benjamins, pp. 401–419.
- Lim, L. (2015). Catalysts for change: On the evolution of contact varieties in multilingual knowledge economy. Ms. The University of Hong Kong.
- Lim, L. (2016). Multilingual mediators: The role of the Peranakans in the contact dynamics of Singapore. In: Li Wei, ed., *Multilingualism in the Chinese diaspora worldwide: Transnational connections and local social realities*. (Routledge Critical Studies in Multilingualism), 1st ed. New York and Abingdon: Routledge, pp. 216–233.
- Lim, L. (2020). The contribution of language contact to the emergence of World Englishes. In: M. Hundt, D. Schreier, and E.W. Schneider, eds., *The Cambridge handbook of world Englishes*, 1st ed. Cambridge: Cambridge University Press, pp. 72–98.
- Lim, L. and Ansaldo, U. (2007). Identity alignment in the multilingual space: The Malays of Sri Lanka. In: E.A. Anchimbe, ed., *Linguistic identity in multilingual postcolonial spaces*, 1st ed. Newcastle upon Tyne: Cambridge Scholars Press, pp. 218–243.
- Lim, L. and Ansaldo, U. (2012). Contact in the Asian arena. In: T. Nevalainen and E.C. Traugott, eds., *The Oxford handbook of the history of English*, 1st ed. New York: Oxford University Press, pp. 560–571.
- Lim, L. and Ansaldo, U. (2016). *Languages in contact*. Cambridge: Cambridge University Press.
- Lim, S. (1988). Baba Malay: The language of the ‘Straits-born’ Chinese. In: H. Steinhauer, ed., *Papers in Western Austronesian linguistics No. 3*. Department of Linguistics, Research School of Pacific Studies, The Australian National University, pp. 1–61.

- Liu, Y. (2013). *Marketplace communication between Africans and Chinese in Guangzhou: An emerging pidgin?* MPhil. The University of Hong Kong.
- Luke, K.K. (2008). Stress and intonation in Hong Kong English. In: *14th conference of the international association for world Englishes (IAWE)*. Hong Kong, 1–5 December 2008.
- Martino, E. (2003). *Chinese Pidgin English: Genesi ed evoluzione* [Chinese Pidgin English: Genesis and evolution]. Tesi di Laurea. La Sapienza, Roma.
- Matras, Y. (2000). How predictable is contact-induced change in grammar? In: C. Renfrew, A. McMahon and L. Trask, eds., *Time depth in historical linguistics*, vol. 2, 1st ed. Oxford: MacDonal Institute for Archaeological Research, pp. 563–583.
- Matras, Y. (2010). Contact, convergence and typology. In: R. Hickey, ed. *Handbook of language contact*, 1st ed. Oxford: Blackwell, pp. 66–85.
- Mufwene, S.S. (1996). The founder principle in creole genesis. *Diachronica*, 13(1), pp. 83–134.
- Mufwene, S.S. (2001). *The ecology of language evolution*. Cambridge: Cambridge University Press.
- Mufwene, S.S. (2008). *Language evolution: Contact, competition and change*. London: Continuum.
- Nordhoff, S. (2009). *A grammar of upcountry Sri Lanka Malay*. PhD. University of Amsterdam.
- Pakir, A. (1986). *A linguistic investigation of Baba Malay*. PhD. University of Hawai‘i.
- Platt, J. (1987). Communicative functions of particles in Singapore English. In: R. Steele and T. Threadgold, eds., *Language topics: Essays in honour of Michael Halliday*, vol. 1, 1st ed. Amsterdam: John Benjamins, pp. 391–401.
- Reich, D., Kumarasamy, T., Patterson, N., Price, A.L. and Singh, L. (2009). Reconstructing Indian population history. *Nature*, 461(7263), pp. 489–494.
- Reid, A. (1980). The structure of cities in Southeast Asia, 15th to 17th centuries. *Journal of Southeast Asian Studies*, 11(2), pp. 235–250.
- Reid, A. (2000). *Charting the shape of early modern Southeast Asia*. Chiang Mai: Silkworm Books.
- Reineke, J.E., Tsuzaki, S.M., de Camp, D., Hancock, I.F. and Wood, R. (1975). *A bibliography of pidgin and creole languages*. Honolulu: The University Press of Hawai‘i.
- Rivera-Castillo, Y. and Pickering, L. (2004). Phonetic correlates of stress and tone in a mixed system. *Journal of Pidgin and Creole Languages*, 19, pp. 261–284.
- Rudolph, J. (1998). *Reconstructing identities: A social history of the Babas in Singapore*. Aldershot and Vermont: Ashgate Publishing Co.
- Saldin, B.D.K. (2003). *Portrait of a Sri Lanka Malay*. Dehiwala: B.D.K. Saldin.
- Selbach, R. (2007). Lingua Franca of the Mediterranean (1350–1830): Finding Nessie? In: S. Ureland, A. Lodge and S. Pugh, eds., *Language contact and minority languages on the littorals of Europe*, 1st ed. Berlin: Logos Verlag, pp. 149–160.
- Selby, A. and Selby, S. (1995). China Coast Pidgin English. *Journal of the Hong Kong Branch of the Royal Asiatic Society*, 35, pp. 113–141.
- Shellabear, G. (1913). Baba Malay: An introduction to the language of the Straits-born Chinese. *Journal of the Straits Branch of the Royal Asiatic Society*, 65, pp. 49–63.
- Smith, I.R. (1979a). Convergence in South Asia: A creole example. *Lingua*, 48, pp. 193–222.
- Smith, I.R. (1979b). Substrata versus universals in the formation of Sri Lanka Portuguese. *Papers in Pidgin and Creole Linguistics*, 2, pp. 183–200.
- Tan, C.B. (1988). Structure and change: Cultural identity of the Baba of Melaka. *Bijdragen tot de Taal, Land- en Volkenkunde*, 144, pp. 297–314.
- Thomason, S.G. and Kaufman, T. (1988). *Language contact, creolization, and genetic linguistics*. Berkeley: University of California Press.
- Tong, K.-S. (1862). Ying yúh jaahp chyúhn (英語集全) [*The Chinese and English instructor*], 6 vols. Guangzhou.
- Tryon, D., Mühlhäusler, P. and Baker, P. (1996). English-derived contact languages in the Pacific in the 19th century (excluding Australia). In: S.A. Wurm, P. Mühlhäusler and D. Tryon, eds., *Atlas of languages of intercultural communication in the Pacific, Asia and the Americas*, 1st ed. New York: De Gruyter Mouton, pp. 471–495.

- Van Dyke, P.A. (2005). *The Canton trade: Life and enterprise on the China coast, 1700–1845*. Hong Kong: Hong Kong University Press.
- Wee, L.H. (2008). More or less English? Two phonological patterns in the Englishes of Singapore and Hong Kong. *World Englishes*, 27, pp. 480–501.
- Wong, Y.T. (2009). *The linguistic function of Cantonese discourse particles in the English medium online chat of Cantonese speakers*. MA thesis. University of Wollongong.
- Yakpo, K. (2009). *A grammar of Pichi*. Berlin and Accra: Isimu Media.

Eastern Polynesia

Mary Walworth

1. Introduction

The linguistic area of Eastern Polynesia has undergone severe socio-political pressure from Europeans, North Americans, and South Americans since the late eighteenth century. This dominance has had a profound effect on the linguistic landscape and has led to numerous manifestations of language contact. In French Polynesia, the heavy influence of France has given rise to a local variety of French, ‘Tahitian French,’ and France’s centralization of the island of Tahiti has resulted in the prestige of the Tahitian language and subsequent formation of unique mixed languages with other indigenous Polynesian languages. In Hawai‘i, prominent sugar cane plantations in the early nineteenth century led to the eventual formation of Hawai‘i Creole English. Additionally, the spread of English in Hawai‘i over the last two centuries has led to a local variety of English, known as ‘Hawai‘i English.’ In New Zealand, due to the colonization by the British in the nineteenth century and the spread of English since, a local variety of English has also formed, known as ‘Māori English.’ In Rapa Nui, a local variety of Spanish, ‘Rapa Nui Spanish,’ is used by younger generations and is a product of their simultaneous acquisition of Spanish and Rapanui language.

This chapter introduces and examines the multitude of contact varieties spoken in Eastern Polynesia, highlighting how the region’s high linguistic diversity and manifold colonial histories have shaped a complex network of language contact. Primarily, this chapter discusses the resulting contact languages from the history of interaction between Polynesian speech communities and the linguistic intrusion of European languages during and after the colonial period. Through a description of the unique histories of contact in Rapa Nui, French Polynesia, Hawai‘i, and New Zealand; this chapter focuses on the social and linguistic processes that have formed the three primary types of contact languages found in Eastern Polynesia: creoles, mixed languages, and indigenized varieties of English, French, and Spanish. Here, creole is defined as a language that develops out of contact between a European language and several non-European languages in a particular socio-cultural situation – the colonial plantation (Mufwene, 2015). Mixed language as a language is defined here that has two identifiable source languages and develops out of bilingualism (Meakins, 2014; Matras and Bakker, 2003) to express a particular social identity (Meakins, 2013; Walworth, 2017b). A variety of a language

can be generally defined as the usage of a language in a specific geographic, social, or political context. In this chapter, I focus on what Mufwene calls ‘indigenised varieties’: major languages that have been adapted by local populations (Mufwene, 2001).

2. Historical overview

Eastern Polynesia encompasses most of the space within the Polynesian Triangle (Figure 23.1), an area that covers most of the central and eastern part of the Pacific Ocean.

The autochthonous languages of this area have been grouped into a subgroup of the Polynesian language family called Eastern Polynesian, first defined by Green in 1966 and developed further by Biggs in 1971. The model of Eastern Polynesian language diversification laid out by Biggs (1971) remained the prevailing theory until Walworth in 2014 presented a new model, based on intense prehistoric contact networks between the Eastern Polynesian languages. These languages are identified in the tree in Figure 23.2.

The Eastern Polynesian languages are very closely related and differ only slightly from one another through some lexical innovations, a few sporadic sound changes, and regular consonant changes (see Walworth, 2014). In order to demonstrate their genetic closeness, and to establish their difference from European languages, the comparative consonant inventories of many Eastern Polynesian languages are shown in Table 23.1.

Contact in Eastern Polynesia can be divided into two historical periods: the Pre-European (or Interactive) Period and the European (or Colonial) Period. The Pre-European period began with the arrival of Polynesian-speaking peoples into the area. These were the first settlers of the region, populating previously uninhabited islands, and they were the ancestors of the modern indigenous inhabitants of the area. The settlement of Eastern

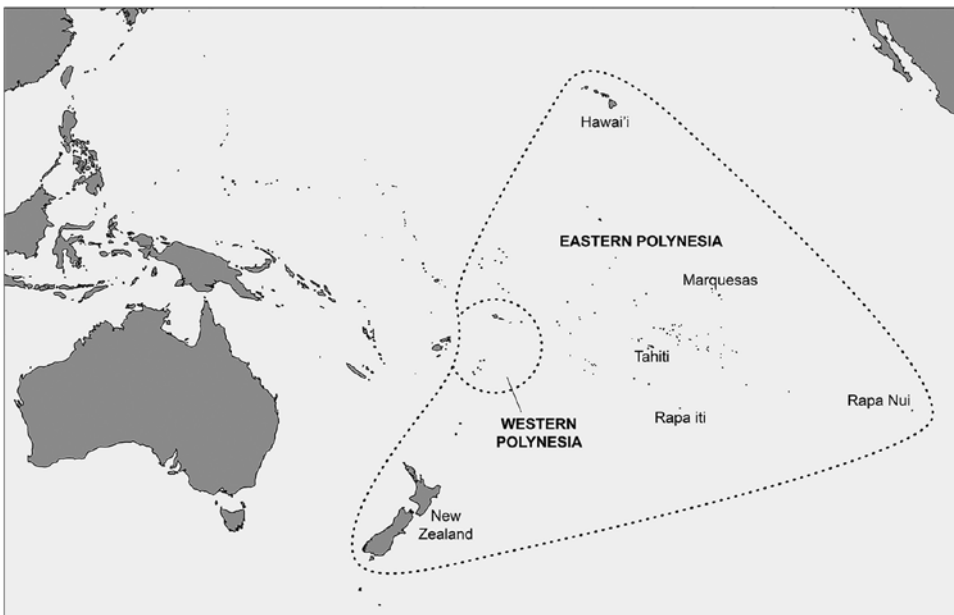


Figure 23.1 Map of the Polynesian Triangle

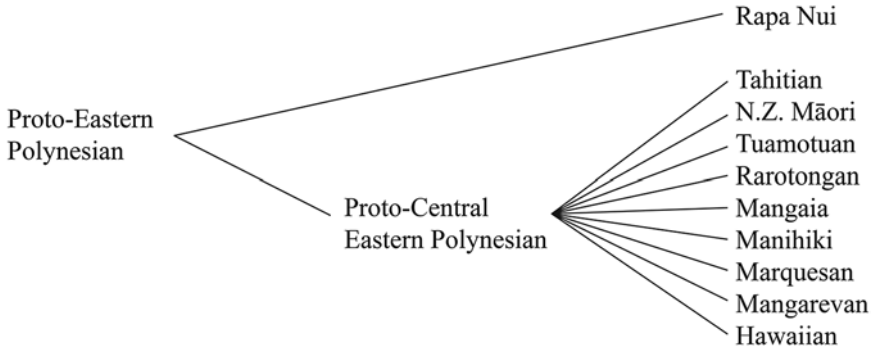


Figure 23.2 Eastern Polynesian linguistic relationships

Source: Walworth, 2014

Table 23.1 Consonant inventories of some Eastern Polynesian languages

| PEP | *p | *t | *k | *m | *n | *ʔ | *ŋ | *f | *s | *v | *r |
|-----|----|----|-----|----|----|----|-----|-----|----|----|----|
| HAW | p | k | ʔ | m | n | ∅ | n | h | h | w | l |
| TAH | p | t | ʔ | m | n | ∅ | ʔ | f/h | h | v | r |
| RPN | p | t | k | m | n | ʔ | ŋ | h | h | v | r |
| MGV | p | t | k | m | n | ∅ | ŋ | ʔ | ʔ | v | r |
| RAR | p | t | k | m | n | ∅ | ŋ | ʔ | ʔ | v | r |
| MAO | p | t | k | m | n | ∅ | ŋ | f | h | v | r |
| OR | p | t | k | m | n | ∅ | ŋ | ʔ | ʔ | v | r |
| MQS | p | t | k/ʔ | m | n | ∅ | k/n | f/h | h | v | ʔ |

Polynesia occurred in waves between 1,000 and 800 years ago (Kirch, 2017), followed by significant movement between island groups (Collerson and Weisler, 2007; Weisler, 1997; Hermann et al., 2017). Due to intensive and consistent exchange – for goods, resources, and to keep kin relationships active – the post-settlement period was highly interactive, with long-distance voyaging and localized networks of contact and influence characterizing the development of the Eastern Polynesian languages (Walworth, 2014, 2017a). Archaeologists report that this interaction slowed in the fifteenth century (Weisler, 1997; Rolett, 2002) and, as a result, speakers of early Eastern Polynesian languages became more isolated, allowing for linguistic diversity to emerge. It was not until the arrival of the Europeans that regular and maintained contact between languages began again, and in this case, Polynesian languages faced non-Polynesian languages in situations of dominance.

Europeans began actively exploring Eastern Polynesia in the 1700s. By the late eighteenth century, Europeans had descended upon nearly every part of the region for either religious missions or exploitation of natural resources. By the end of the nineteenth century, the entire area was under colonial rule: New Zealand by England, Hawai‘i by the United States, French Polynesia by France, and Rapa Nui by Chile. While the contact histories of the Colonial Period share a common theme of dominance, the nature of the European infiltration varies from island to island.

In New Zealand, whaling and sealing began in the 1790s. English Christian missionaries then began arriving in the early nineteenth century. New Zealand became a British colony in 1840, and, by 1858, there were more Europeans than Māori living there (King, 1993, p. 169).

By the turn of the twentieth century, non-Māori, or *pākehā* (Māori for New Zealanders of European descent), owned most of the land and controlled the government.

Europeans began arriving in Hawai‘i around 1775 when the islands became an important stopping point in the sandalwood, whaling, and fur trades (Drager, 2012; Reinecke, 1969). Both English and American missionaries began to arrive in Hawai‘i at the beginning of the nineteenth century, rapidly converting most of the population to Christianity. Shortly after the start of the Christian missions, American sugar cane companies began establishing plantations throughout the Hawaiian Islands. By the end of the nineteenth century, Hawai‘i was annexed to the United States, and the North American population increased significantly (Drager, 2012).

In French Polynesia, contact with Europeans began with the whaling and sandalwood trade in the later part of the nineteenth century (Fisher, 2013). Europeans began settling in various parts of French Polynesia in the early nineteenth century due to Christian missions. These included first the London Missionary Society followed shortly after by French Catholic priests (Fisher, 2013; Charpentier and François, 2015). Later in the 1800s, due to its centrality and large harbour, the French made Tahiti their political and commercial centre for the region, and in 1842, Tahiti became a protectorate of France (Fisher, 2013; Charpentier and François, 2015). By 1900, all islands of what is today known as French Polynesia had been annexed to France. The twentieth century saw continuous pressure and an increased French presence throughout the region due to economic interest (primarily for the extraction of phosphate and the pearl industry), and the growth of military presence as a result of over 30 years of nuclear testing (Fisher, 2013).

Colonial influence in Rapa Nui has been equally influential socio-linguistically, in spite of what might appear to be a less complex colonial past. Rapa Nui became a Chilean territory in 1888, but Kieviet (2017) and Makihara (2005) report that contact and regular interaction with Chileans has been a more recent phenomenon of only the past 50 years, since the arrival of a civil administration in 1966 (Makihara, 2005, p. 727). The installation of several hundred Chilean government workers and their families, at that time, had a significant cultural and linguistic impact on the population of indigenous Rapanui.

3. Critical issues and topics

Eastern Polynesia provides a unique opportunity to study a wide range of contact languages. The region’s relatively short population history in combination with 200 years of interaction with colonial outsiders has led to a diversity of language contact situations. While only beginning two centuries ago, this contact dominance incited rapid socio-cultural and linguistic change, resulting in the many language contact varieties spoken today. Three broader issues of language contact arise through the example of Eastern Polynesia: language contact is not always directly between local, indigenous languages and exotic, non-indigenous languages; contact languages that emerge from situations of dominance are often used to express covert prestige of the oppressed peoples; contact languages that are spoken alongside their source languages, in multilingual contexts, tend to exhibit significant variation.

3.1 *Indirect versus direct contact*

Contact language development in Eastern Polynesia is primarily due to the convergence of cultures during the Colonial period and is the result of two kinds of contact influence: direct (actual presence and first-hand dominance of European outsider) and indirect (centralization of religion and education, migrations to urban centres due to prestige) influence. These have led to two types of contact languages: (1) languages whose sources are unrelated, and (2) languages whose sources

are closely related. Type 1 is found throughout the region, where non-Polynesian colonial languages (English, French, and Spanish) directly dominated local Polynesian languages; Type 2 is seen primarily in French Polynesia, where indirect colonial pressure has prompted language shift and subsequent language mixing between closely related Eastern Polynesian languages. This indirect pressure has been defined as ‘trickledown endangerment’, when a dominant linguistic centre absorbs heavy linguistic influence, which is then then diffused to other areas (Walworth, 2015b).

3.2 Dominance and identity

Because issues of dominance play a central role in the way languages have converged in Eastern Polynesia, contact languages today strongly sociolinguistically mark identities among their speakers. This is usually due to covert prestige of ‘localness,’ where there is implicit value in being autochthonous and ‘local’ (Walworth, 2017b). Covert prestige refers to hidden values associated with non-standard speech (Labov, 1972, p. 249; Trudgill, 1972, p. 183). This prestige has an impact on domains of use and the speaker base for most of the contact languages discussed here within. These languages are context-driven and therefore spoken in very specific circumstances or domains, and because they are a way of marking a local indigenous identity, they are often spoken only among local populations or by those who seek to be identified as more ‘local.’ Furthermore, the covert prestige of ‘localness’ reveals and reflects positive attitudes toward indigenous languages even whilst the heritage language is endangered and the speaker-base decreases.

3.3 Variation

It is important to note that the contact languages of Eastern Polynesia are vernacular and non-standard. They therefore, exhibit significant variation. All of the languages described in this chapter demonstrate some degree of variation between individual speakers because they are spoken alongside their lexifying languages. Speakers are often multilingual, speaking at least two languages. Thus, the source languages for these contact varieties are often in regular contact, and the varieties are consistently changing and developing.

4. Current contributions and research

The following sections present and define several contact varieties throughout Eastern Polynesia while addressing the nature of the contact context that led to their emergence. Contact here is discussed according to sub region (New Zealand, Hawaii, French Polynesia, and Rapa Nui). The sociolinguistic context for each language’s development, the demography of the speaker-base, the language’s use in society, and a summary of each language’s marked features are all described. Because issues of dominance play a central role in the way these languages have developed in Eastern Polynesia, each language description also confronts how the language marks identity among its speakers. It is important to note that the descriptions that follow serve as summaries and do not account for the full range of variation between speakers of different age groups, different genders, and different socio-economic status.

4.1 New Zealand and Māori English

The Polynesians who settled in New Zealand eventually came to be known as Māori, and their various lects are today known collectively as the Māori language. The influence of English, beginning from the European period, has led to it being the dominant language in nearly all

domains, for both Māori and Pākehā. In the Colonial period, as Pākehā grew to be the majority in New Zealand, schools became English-only, and use of Māori language experienced a steady decline in all domains outside of the home (Maclagan, King, and Gillon, 2008). After World War II, urbanization led to a rapid decline in Māori speakers, as the majority of Māori moved to city-centres for work (Maclagan, King and Gillon, 2008:659). English was then viewed as necessary to participate in the growing economy and to obtain employment (Maclagan, King and Gillon, p. 659). Today, Māori has been almost completely replaced by English in every domain, except for very specific, formal situations (Holmes, 1997, p. 66). In their 2008 paper, Maclagan, King and Gillon (2008, p. 659) reported that only 9% of the Māori population were fluent speakers of Māori and that all Māori speakers are bilingual with English. According to a 2013 survey on Māori culture and well-being, Te Kupenga, only 50,000 people (1% of the total population) reported to be fluent in Māori (Statistics New Zealand, 2013).

While Māori and non-Māori interact extensively in New Zealand society, and both primarily use English, there are post-colonial socio-economic realities that push many Māori to the margins of society and have led to social segregation (Holmes, 1997). Furthermore, the dedication to the cultural revival of Māori tradition in the face of linguistic and socio-political dominance by non-Māori has led to some cultural segregation (Holmes, 1997). These two factors are the primary social driving forces behind the development of a uniquely Māori-style English (Holmes, 1997; Szakay, 2007). Holmes (1997) and Szakay (2007) both describe that the Māori variety of English has formed to emphasize Māori culture. This is evident in the fact that the speaker base of Māori English is not restricted to Māori people and sometimes includes Pākehā, who identify as part of Māori communities (King, 1993). Holmes (1997) calls these ‘Māori contexts’ and describes Māori English as used to ‘signal solidarity,’ observing that this variety of English serves to signal a positive Māori identity in an English dominated society. It can be heard in mainstream media but is most commonly found spoken in areas where there is a higher Māori population (Maclagan, King and Gillon, 2008 , p. 661). Maclagan, King and Gillon, 2008 additionally wrote that Māori English is commonly heard in specific occupational groups, by Māori students at university, and by non-Māori children who go to school in areas with higher Māori populations (2008, p. 659). Māori English is generally spoken by people under the age of 50 and is less often used by Māori elders (Maclagan, King and Gillon, 2008 , p. 661).

Māori English exhibits influence from the Māori language through morphological, lexical, pragmatic, and phonological features. In the following sections, following the more marked characteristics that have been described for Māori English are listed and briefly described as they relate to influence from the Māori language.

4.1.1 Marked phonological features of Māori English

Most of the features described for Māori English are phonological:

- Alternative pronunciation of dental fricatives
- Deaspiration of English /t/
- Devoicing of English /z/
- Fronting of back vowels
- Less variation in syllable length
- Difference in syllable timing

These features highlight differences in the sound systems of Māori and English and are clear markers of Māori language input.

Māori English consonant qualities are demonstrably affected by Māori's consonant inventory. First, Māori does not contain dental fricatives /h/ and /ð/. In Māori English, these English dental fricatives are therefore stopped to /t/ and /d/, affricated, or fronted to /tʃ/ and /v/ (Maclagan, King and Gillon, 2008, p. 663; Bell, 2000). Second, Māori English shows clear influence from Māori language in speakers' deaspiration of initial /t/ (Holmes and Ainsworth, 1996; Bell, 2000; Maclagan and King, 2007). English initial plosive stops are usually aspirated in English (Maclagan, King and Gillon, 2008, p. 663). However, plosives in Māori are described as unaspirated (Bauer et al., 1997). Third, English /z/ becomes devoiced in Māori English (Holmes, 1997; Benton, 1966; Bell, 2000; Maclagan, King and Gillon, 2008, p. 663). Māori contains very few voiced consonants and importantly does not recognize a distinction between voiced and unvoiced alveolar fricatives (*s* and *z*). English, on the other hand, does. King (1993, p. 35) describes that devoicing of otherwise voiced English consonants is common in Māori English as a result of Māori influence.

Concerning vowels, Māori English exhibits fronting of back vowels, in particular, /u/ (Bell, 2000). Māori English furthermore demonstrates less variation in syllable length than standard varieties of English. It tends to exhibit shortening of long vowels, reduction of diphthongs, and fully produced vowels in place of schwa in unstressed syllables (Bauer, 1994; Szakay, 2007). This is likely due to the fact that Māori makes a phonemic distinction between long and short vowels.

4.1.2 *Marked pragmatic features of Māori English*

Māori English's tag particle *eh* is one of the most salient features of Māori English and has been described by Bauer et al. (1997, p. 426) to originate from the Māori tag particle *ne*. Meyerhoff found that the tag marker serves as an in-group 'Māori' identity marker (1994, p. 375).

4.1.3 *Māori lexical tokens in Māori English*

Māori English employs many Māori words. These words are consistent across the speaker base and relate to specific aspects of Māori culture or the Māori family structure: *reo* 'language'; *whānau* 'family'; *kaupapa* 'philosophy'; *hui* 'assembly, gathering, to gather, to assemble'; *iwi* 'tribe'; *waka* 'canoe, vehicle or vessel'; *mahi* 'work or an activity' (Maclagan, King and Gillon, 2008, p. 664; Keegan, 2019). They, therefore, can be viewed as tokens incorporated into the English variety to mark a distinct Māori identity.

4.2 *Hawai'i*

The first non-Polynesian permanent settlers in Hawai'i were the English-speaking missionaries. Through the rapid conversion of much of the Hawaiian population, including Hawaiian royalty, and implementation of religious schooling, English spread very quickly (Kawamoto, 1993; Drager, 2012). The work of the missionaries had a profound effect on Hawaiian society, by shifting the traditionally oral linguistic culture of Hawaiian to a written one and eventually, through the production of written literacy materials such as English-Hawaiian dictionaries, encouraged English language learning (Kawamoto, 1993). Economic prestige was equally responsible for the dominance of the English language in Hawai'i, and this was particularly so in the domain of education where '[e]conomic incentives for teachers favoured English speakers, and consequently, the vernacular schools gradually lost ground' (Shi, 1990 in Kawamoto, 1993, p. 197). By 1900, English was the required language of schools (Stueber, 1964, p. 147 in Kawamoto, 1993, p. 200), thus thoroughly dominating all domains of life in Hawai'i.

In addition to the dominance of English, the development of the sugarcane and pineapple plantations in the 1830s had a incredible effect on language mixing in Hawaii (Siegel, 2010; Drager, 2012). The immigrant population who came to Hawai‘i in the mid and late nineteenth century to work in the plantations brought with them many different mutually unintelligible languages, notably Chinese, Portuguese, and Japanese. In attempts to communicate both with the English-speaking plantation owners and between each other, speakers of these languages, along with the local Hawaiian speaking workers, facilitated a mixing of their languages with English (often called jargon), eventually affecting a process of pidginization (Siegel, 2010). Initially lacking uniformity in its structure, stabilization occurred as the children of immigrants began to learn Hawai‘i Pidgin English as their first language, and the most critical period of Hawai‘i Pidgin English formation is said to have been between 1890 and 1910 (Siegel, 2010). As Hawai‘i Pidgin English continued to expand grammatically and develop stability in its inter-generational transmission, it became a creole; now officially known as Hawai‘i Creole English.

4.2.1 *Hawai‘i Creole English*

Hawai‘i Creole English is spoken by approximately half of the population in Hawaii, approximately 700,000 people (United States Census, 2018). It is primarily used in informal contexts and thrives in the domain of home and family life (Drager, 2012). By contrast, in Hawai‘i today, Standard English is the language of education and for economic success (Drager, 2012, p. 64). English is also visibly the language of government officials and policy-making.

Hawai‘i Creole English’s unique development history has led to its use as a marker of localness in Hawaii (Drager, 2012). Drager (2012, p. 64) has described a segregationist policy of the plantation owners that was intended to promote divisions between ethnic and linguistic in order to groups to maintain power and dominance over growing immigrant numbers. Hawai‘i Creole English’s formation can then be viewed as a means of unified identity formation among the immigrant groups in response to the dominance of the plantation owners, in spite of their ethnic and linguistic differences. Drager (2012, p. 65) wrote,

Hawai‘i Creole English was the ‘glue’ which bonded the descendants of the immigrants, who connected neither with their ancestral culture nor with the mainstream Anglo-Saxon American culture from which they were essentially locked out. By being ‘local,’ one could maintain a sense of ethnic identity while at the same time identifying with a larger, more encompassing culture.

Today, ‘localness’ is still very much attributed to speakers of Hawai‘i Creole English (Drager, 2012).

Turning to its structure, Hawai‘i Creole English is an English lexified creole, meaning that the majority of its lexicon is based on the English language. It is not, however, a variety of English; it maintains a different syntactic structure and sources both phonological and grammatical elements from other languages. In the next sections, I provide a very brief grammatical and phonological sketch of Hawai‘i Creole English. The language is well-described and more detailed descriptions including phonotactics and more on the grammatical structure can be found in the section on Further Reading.

4.2.1.1 PHONOLOGICAL CHARACTERISTICS OF HAWAI‘I CREOLE ENGLISH

Hawai‘i Creole English’s phonological system is often described as it compares to English (Bickerton and Odo, 1976; Sakoda and Siegel, 2004a) because its phoneme inventory is sourced mainly from English, with some phonetic differences. The consonant system is described by

Sakoda and Siegel (2004a) as the same as Standard American English, with some minor differences. Three significant examples are (1) that /t/ and /d/ are used where English contains /θ/ and /ð/; (2) that /r/ is rarely the coda of a syllable; and (3) Hawai'i Creole English employs the flap as an additional consonant. More examples and further description can be found in Sakoda and Siegel (2004a, pp. 743–744). The vowel phoneme inventory is as follows: /i e ə α ɔ o u/. There has been extensive research on the variation in the phonetic realization of Hawai'i Creole English vowels, described in detail by Sakoda and Siegel (2004a, pp. 739–744).

4.2.1.2 GRAMMATICAL CHARACTERISTICS OF HAWAI'I CREOLE ENGLISH

As stated previously, many features of Hawai'i Creole English syntax are derived from its non-English source languages. The most salient of these is found in Hawai'i Creole English's word order (Sakoda and Siegel, 2003). Word order is often verb-initial in Hawai'i Creole English, especially in descriptive phrases, mirroring that of Hawaiian (Sakoda and Siegel, 2003). For example, in Hawaiian, a descriptive sentence such as 'the house is big' would be *nui ka hale*, literally 'big the house.' Similarly, the phrase in Hawai'i Creole English would be *big, da house* (Sakoda and Siegel, 2003).

Hawai'i Creole English often employs English lexical forms that do not function the same as the identical form would in English (Sakoda and Siegel, 2003; 2004a; 2004b). In other words, while the English word-forms are being used, they carry a different meaning to their English meaning and source the grammatical function and total semantic value from other Hawai'i Creole English source languages. One example is Hawai'i Creole English *get*, which is used as both a possessive (Example 1) and an existential marker (Example 2). This comes from Cantonese, where the word *yáuh* has both possessive and existential functions (see Sakoda and Siegel, 2004a, p. 732 for examples).

1. *They get three sons.*
'They have three sons.'
2. *Get one student he very bright.*
'There's one student who is very bright.'

(Sakoda and Siegel, 2004a, p. 732)

Two major grammatical features come from Portuguese but also are realized through lexemes: *stay* and *for* (Sakoda and Siegel, 2003; 2004a; 2004b). In Portuguese, *para* introduces infinitive clauses (Example 3a). Similarly, Hawai'i Creole English uses *fo*, from English *for* (Example 3b). Hawai'i Creole English also has incorporated the copula functions of Portuguese *estar* (Example 4a) and applied them to the English lexeme *stay* (Example 4b).

3.
 - a. *Carlos é homem para fazer isso.*
Carlos is man for do that.
'Carlos is the man to do that.'
 - b. *Charles is da man fo do 'um.*
'Charles is the man to do it.'
4.
 - a. *água está fria.*
'The water is cold.'

- b. Da water stay cold.
'The water is cold.'

(Sakoda and Siegel, 2004a, pp. 732–733)

4.2.2 *Hawai'i English*

Hawai'i English is a variety of English spoken in Hawai'i that incorporates not only elements of the Hawaiian language, but also many elements of Hawai'i Creole English (Drager, 2012). Hawai'i English, like Hawai'i Creole English, is used in more familiar domains and is again used as a marker of local identity (Drager, 2012). Through the use of some local linguistic markers, Hawai'i English is a way for people to assert their local-ness, even without necessarily speaking Pidgin. Hawai'i English is primarily defined by lexical borrowings and some marked phonological characteristics, many of which are found to be part of the Hawai'i Creole English system (Drager, 2012).

4.2.2.1 MARKED PHONOLOGICAL FEATURES OF HAWAI'I ENGLISH

Four main characteristics have been identified as marking Hawai'i English, three of them related to vowel qualities of speakers. First, Hawai'i English speakers tend to employ full vowel qualities when Standard varieties of American English would exhibit schwa – for example, Hawai'i English speakers are more likely to pronounce the first vowel in the word 'today' as [u] rather than [ə] (Sato, 1993, p. 135; Drager, 2012, p. 65). Second, speakers tend to reduce diphthongs – for example, the diphthong vowel in English *goat* is realized as [o] (Sakoda and Siegel, 2004a, p. 742; Drager, 2012, p. 65), there is a merger in Hawai'i English of the /a/ and /ɔ/ vowels, whereas in Standard American English there is generally a distinction between them – for example, the vowels in *lot* and *thought* in American English are distinct, whereas in Hawai'i English, they are usually merged (Drager, 2012, p. 66), and finally, palatalization is often observed, when /t/ is preceded by /s/ (Drager, 2012, p. 66).

4.2.2.2 MARKED LEXICAL FEATURES OF HAWAI'I ENGLISH

Drager (2012, p. 67) has outlined a handful of marked lexical items that are associated with Hawai'i English; all of which, she wrote, are also used (and likely taken from) Hawai'i Creole English: *makai*, Hawaiian for 'toward the ocean,' *manapua*, Hawaiian for 'steamed pork bun,' *shi-shi*, Japanese for 'urinate.' Some additional Hawaiian terms may be added to this list: 'ono' 'delicious,' *aloha* 'greeting, love,' *ohana* 'family,' *mauka* 'toward the mountains (definitions from Ulukau, 2019)'

Drager (2012, p. 67) has also described additional marked terms that 'rely on cultural knowledge that is shared by people who were born and raised in the island chain.' She wrote that these words are particularly marked because their meaning can be derogatory, depending on the use. Drager has explained that speakers of Hawai'i English and Hawai'i Creole English recognize the contexts in which such words can be used without 'undesirable consequences' (2012, p. 67). One example listed by Drager is *moke*. In specific social contexts, and when used by Hawai'i English and Hawai'i Creole English speakers, this semantically modified British English term for 'donkey,' can mean simply 'local island guy' (Drager, 2012, p. 67). When used out of the Hawai'i English or Hawai'i Creole English context, *moke* can have derogatory connotations, taking on a meaning that approaches 'idiot local guy (Drager, 2012).'

4.1 French Polynesia

Language contact in French Polynesia falls into two categories: historic French intrusion and Tahitianization. These two channels of contact and influence have led to very different linguistic situations: (1) French intrusion has led to a variety of French spoken in Tahiti called Tahitian French, and (2) Tahitianization has led to significant mixing with other indigenous languages in the outer-islands of the region and the development of unique mixed languages between closely related Polynesian languages.

4.2.1 French intrusion and Tahitian French

The French have affected linguistic and cultural influence in Tahiti since the early nineteenth century. Beginning in the 1820s when Queen Pōmare IV made Pape'ete the capital of the region, French shipping and trading companies were based there (Charpentier and François, 2015). French Catholic missionaries arrived to Tahiti shortly after this time in the 1830s. Tahiti was officially claimed as a part of France by Captain Dupetit-Thouars in 1842 but was not fully annexed until 1880, along with the rest of the geo-political region known today as French Polynesia (Fisher, 2013). From the late nineteenth century, a French administration, functioning under French law, governed Tahiti. After the Second World War, France established the Centre d'Expérimentation du Pacifique for atomic testing. This intensified French infrastructure in Tahiti as well as marked growth of the French population (Love, 2006). It is no surprise that during the surge of French influence, education became French-centred and the French language required in schools (Love, 2006; Charpentier and François, 2015). With the increase of French-focused education and economic development, many people in Tahiti moved away from local language and traditional social practices in favour of Western, French lifeways (Love, 2006).

As a result of the movement toward the French language, a non-standard variety of French spoken in French Polynesia, Tahitian French, has developed. It is spoken throughout and serves as a lingua franca across the many islands of French Polynesia and is therefore highly variable (Love, 2006). Tahitian French is characterized primarily by Tahitian phonological and lexical elements. Notably in the phonology are the following features: (1) the Tahitian alveolar trilled /r/ in place of the French uvular trill /R/; (2) incorporation of Tahitian phonemes /h/ and /ʔ/, which are not present as phonemes in French; and (3) marked variation of nasalized vowels, reduction of consonant clusters, and de-voicing of voiced stops (Love, 2006) as Tahitian does not contain these features of French phonology. Turning to the lexicon, Tahitian French borrows many lexical items from Tahitian. These borrowings retain their Tahitian pronunciation and notably take a masculine article (Love, 2006, p. 116). Borrowings vary widely, but generally fall into the following categories (Based on Love, 2006): terms for kinship (*mo'otua* 'grandchild,' *tamari'i* 'child,' *fa'a'amu* 'adopt, lit. to feed,' *tāne* 'man, husband, boyfriend,' *vahine* 'woman, wife, girlfriend'); ethnic identity (*mā'ohi* 'Polynesian,' *farāni* 'French,' *tinitō* 'Chinese'); terms for local plants, particularly flowers (*tiare* 'flower'); fish names (*mahimahi* 'dorade'); food items (*mā'a* 'food, meal' *fe'i* 'banana,' *'uru* 'breadfruit,' *pua'a* 'pig, pork'), and environmental terms (*fenua* 'island'). Additional borrowings that are found in nearly every Tahitian French speaker's lexicon include: *fiu* 'fed-up, tired, bored, apathetic,' *fa'aitoito* 'good luck,' *haere mai* 'let's go; come on,' *oti* 'finished,' *ha'avitiviti* 'hurry-up,' *māuruuru* 'thank you,' *'ia ora na* 'hello (Definitions from Fare Vāna'a, 2017)

While Tahitian French is a consequence of French intrusion, it also serves as a marker of local identity and, due to the phonological and lexical borrowings from Tahitian, is often used to indicate one's connection with Tahitian, even while it is also responsible for the endangerment of Tahitian. It is employed in more informal situations – at home with family and

between friends – and it is rarely used in formal contexts such as government or school, where only Standard French and Tahitian are typically typically used.

4.2.2 Tahitianization and language mixing

Today in Tahiti, there exists a dominance of both French and Tahitian in all cultural domains. While Tahitian is itself dominated by French and thus endangered in Tahiti, it is the driving force of endangerment throughout the rest of French Polynesia. Nearly 90% of people living in the outer-islands speak Tahitian fluently (ISPF, 2012). Most adults speak and use Tahitian regularly, and Tahitian is used more frequently than local indigenous languages (Walworth, 2015b). Rutter (2006) was the first to remark of heavy tahitianization, or, Tahitian influence, in local speech throughout the outer islands.

The driving cause of this Tahitianization is the historic centralization of Tahiti. In the early nineteenth century, European whalers and traders made Tahiti the trading centre for the region (Charpentier and Francois, 2015). Following the traders, the missionaries made Tahiti the centre base for religious conversion and training. French colonizers then made Tahiti the political and commercial centre of the colony, and finally, the French military made it a base beginning in the 1960s due to nuclear testing in Tuamotu and Gambier Islands (Charpentier and Francois, 2015). As Tahiti became and stayed centralized, so did the Tahitian language become the central ‘indigenous’ language of the region. It was the first written and published local language in French Polynesia (by the Protestant missionary Davies in 1805 (Nicole, 1987)); it was the first language to have a language academy, from which it became the first language to have a grammar and standardized orthography; it is the most prominent Polynesian language in television and radio media throughout the region; and it is the language of the Bible. Missionaries, in fact, used Tahitian to translate the Bible for the entire region. The first Tahitian Bible was published in 1836 and was quickly adopted by most other Christian religions, who had a presence in every archipelago (Charpentier and François, 2015).

In the last 20 years, local policy-makers have made strides to increase indigenous language in education and government sectors (Saura, 2009), however, due to the region’s centralization in the island of Tahiti, these efforts have been primarily focused on Tahitian (Charpentier and François, 2015; Walworth, 2015a, 2015b). All schooling is conducted in French, with two hours and 40 minutes per week of ‘indigenous language learning’ required in all primary schools since 1984 (Walworth, 2015b). This generally occurs in Tahitian, even on islands where the indigenous language is not Tahitian. Furthermore, upper-level schooling (high school and university) is available only in Tahiti. As a result, Tahiti is viewed as a place of educational prestige, and often students travel incredibly long distances (sometimes days by boat) and live away from their immediate families to receive higher education (Walworth, 2015b).

At the government level, priority has been given to making Tahitian equally represented in the public sphere, without much acknowledgement of the other several languages of the region. For example, in the Constitution of French Polynesia, Tahitian is explicitly described as part of the collective French Polynesian identity, even while many other languages are spoken: ‘Even if the official language of French Polynesia is French, organic law affirms: the Tahitian language is a fundamental element of the cultural identity, the cement of social cohesion, a regular means of communication, it must be recognised and preserved’ (Constitutional Statute Article 57, 2004). A political movement to support Tahitian as *the* indigenous language of French Polynesia has effectively prioritized Tahitian over all other indigenous languages in the region. This has placed an importance on one single indigenous regional language, Tahitian, and developed a mono-indigenous Polynesian-cultural identity. It is thus creating a ‘mono-indigenous’ culture, where Tahitian has become ‘the’ indigenous language (Walworth, 2015b).

From this Tahitianization, mixing between the closely related Polynesian languages traditionally spoken on other islands and Tahitian has occurred, sometimes developing into stable mixed languages. These mixed languages combine source languages without any distinct division of particular linguistic features stems. The best example of this is a language spoken on the island of Rapa Iti called Reo Rapa. Reo Rapa exhibits a mix of features from both Tahitian and Old Rapa, the indigenous language of Rapa Iti. Reo Rapa is a language that began from a situation of bilingualism in Rapa Iti of Tahitian and Old Rapa and then developed out of language shift and subsequent fusion of linguistic features. This shift stalled due to anti-convergence sentiments, and Reo Rapa then became intergenerationally stable (Walworth, 2017b). Walworth (2017b) defined Reo Rapa as a *shift-break language*: a language that has resulted from a stalled shift due to a collective anti-convergence sentiment in the speech community. The halt in shift and retention of Old Rapa tokens in the lexicon alongside a solid comprehension of Old Rapa phonological features in Reo Rapa's phonology demonstrate a clear expression of a Rapa identity that was able to emerge when the language loss was perhaps too great to safeguard the language. The culturally specific tokens allowed for an explicit attachment to a Rapa linguistic identity regardless of actual language knowledge (Walworth, 2017b).

In Reo Rapa, lexical items are phonologically marked for their respective source languages, and speakers demonstrate an active awareness of what sounds are from Tahitian or Old Rapa. Phonemes that are absent in Tahitian, though present in Old Rapa, are the velar nasal /ng/ and the velar stop /k/; phonemes of Tahitian that are not found in Old Rapa are /h/ and /f/ (Walworth, 2017b). Reo Rapa speakers generally understand that where there is a glottal stop in Tahitian, there 'should' be a /k/ or /ng/ in Old Rapa, and where there is an /h/ or /f/ in Tahitian, there 'should' be a glottal stop in Old Rapa. Walworth (2017b) described Reo Rapa's phonological system as therefore layered or stratified, rather than demonstrating adaptive phonology. In Reo Rapa, all consonant phonemes from both Old Rapa and Tahitian have been retained. Regarding vowels, Reo Rapa exhibits the five-vowel system (a, e, i, o, u; with surface length contrast) found in both Old Rapa and Tahitian.

Regarding the lexicon, most of Reo Rapa's lexicon is Tahitian, however grammatical words come from both Tahitian and Old Rapa (Walworth, 2017b). Because Tahitian and Old Rapa are closely related, they have a high lexical similarity due to their many cognate reflexes from their shared mother language, Proto Eastern Polynesian. Thus, a clear identification of the source language for a particular Reo Rapa lexical item is only possible where either source language has developed separate innovations (Walworth, 2017b). Although much of the lexicon is the same in both languages, most of the unshared lexical items come from Tahitian (Walworth, 2017b). There are some very marked features of Old Rapa vocabulary in the Reo Rapa lexicon that differ from the Tahitian forms with the same meaning. These words are usually unique Old Rapa innovations, specialized vocabulary for traditional Rapa Iti activities, and some basic vocabulary (Walworth, 2017b).

4.4 Rapa Nui

Despite having been annexed to Chile in the late nineteenth century, Spanish language influence in Rapa Nui did not become prominent until the 1960s, which according to Kieviet (2017) when Rapa Nui peoples began actively participating in government, taking jobs for which proficiency in Spanish was necessary, and learning Spanish in school. According to the last census, as reported by Kieviet (2017), half of the island's population is of Rapa Nui origin; however, the number of actual speakers of the Rapa Nui language is much lower, Kieviet gives an estimate of 1000 speakers out of a population of nearly 6000. Importantly, Rapa Nui children speak predominantly Spanish, even if their home-context is Rapanui-speaking (Kieviet, 2017). Spanish is dominant and causing an extreme shift.

Kieviet (2017) reported that there has been a rapid shift to Spanish that has led to an infiltration of Spanish features into the Rapanui language and significant code-mixing between the two languages. Makihara (2005) has referred to this code-mixing as ‘Rapanui Spanish’ and described it as having increased variation dependent upon speaker-age, where younger speakers tend to include less Rapanui features than older speakers.

Based on Makihara’s 2005 description of it, Rapanui Spanish can be described as a variety of Spanish with a simplification of features to accommodate the Rapanui repertoire of a speaker. Regarding the phonological system, Spanish contains a larger consonant inventory than Rapanui sounds; therefore, there are many sounds that are not found in Rapanui’s inventory. These are typically replaced by Rapa Nui sounds: [r l d] > [r]; [b] > [v]; [g x] > [k]; [s] > t (Makihara, 2005, p. 733). Additionally, Spanish consonant clusters are reduced, and syllable-final consonants are deleted, or a vowel is added to adapt to Rapanui’s obligatory open syllable structure (Makihara, 2005).

Syntactically speaking, Rapanui Spanish exhibits mixing of the Spanish gender-specific pronouns because Rapanui does not have pronominal gender distinction (Makihara, 2005). In Example 5, the speaker was referring to a man but used the feminine form.

5. *A eia ro vió.*

‘She (He) saw him.

(Makihara, 2005, p. 734)

Rapanui Spanish also demonstrates a preference for present tense and third-person singular forms of verbs (Makihara, 2005).

6.

Intended: ‘I am seventy-five years old. But I always remember my great-great-grandfather.’

Actual:

Yo tiene setenta y cinco año. Pero yo siempre me acuerda de mi tatara abuera.

‘I **is** seventy-five years old. But I always **remembers** my great-great-grandmother.’

(Makihara, 2005, p. 736)

Furthermore, the future is typically expressed by the formula: Spanish *ir* ‘to go’ + preposition *a* + the infinitive form of the primary verb (Makihara, 2005). This construction represents a general preference in Rapanui Spanish to avoid inflectional morphology (affixation) likely due to the more analytic structure of Rapanui (Makihara, 2005).

Makihara (2005) has described the simplification processes in Rapanui Spanish as typical of contact language situations, particularly where the acquisition of one of the languages in contact language is truncated. In Rapa Nui, the Rapanui language is to a lesser and lesser extent transmitted to younger generations, and they are increasingly learning Chilean Spanish at school (Kieviet, 2017). The increased use of Rapanui Spanish represents this layered attrition of Rapanui (Makihara, 2005). However, simultaneously Rapanui Spanish speakers use the variety as a means to identify as ethnically Rapa Nui and speakers attach an identity to the Rapanui features. In this way, Rapanui Spanish implicitly acts to ‘underline solidarity and to authenticate Rapa Nui identity’ in an aggressively polarized ethnic and political environment (Makihara, 2005, p. 732).

5. Future directions

This summary of contact situations in Eastern Polynesia is not fully comprehensive – contact situations in the Cook Islands were not included here as there is minimal research about them.

Diaspora of other Pacific Islanders in major hubs in Eastern Polynesia (Auckland, New Zealand and, and Honolulu, Hawai‘i) also are of interest for modern language contact studies as they continue to affect language change and contribute to an ever-shifting linguistic landscape. These were not included here for limitations of space but certainly merit further attention.

Continued research is needed for understanding the structure and phonetics of Tahitian French and possible pre-pidgin contact languages that may be forming on both Tahiti and other islands of French Polynesia. Additionally, more investigative work in the outer islands of French Polynesia is required to identify other Polynesian mixed languages that may have developed as the definitive halt in shift, and intergenerational consistency is understudied outside of Rapa Iti.

6. Further reading

Love, S. (2006). *Tahitian French*. PhD. Australian National University.

This PhD thesis presents the first major description of the variety of French spoken in Tahiti and surrounding islands. It provides a socio-historical background and offers significant phonetic detail of Tahitian French.

Maclagan, M., King, J. and Gillon, G. (2008). Maori English. *Clinical Linguistics and Phonetics*, 22(8), pp. 658–670.

This article describes the linguistic situation of the Māori people in New Zealand, focusing on the increasing use of Māori English. This paper provides comprehensive details of the features of Māori English and its place in New Zealand society.

Makihara, M. (2005). Rapa Nui ways of speaking Spanish: language shift and socialization on Easter Island. *Language in Society*, 34(5), pp. 727–762.

This article provides a description of the history of colonial influence in Rapa Nui and the linguistic situation that has led to multiple varieties of Rapanui Spanish.

Sakoda, K. and Siegel, J. (2004). Hawai‘i Creole: phonology. In: B. Kortmann, E. Schneider, K. Burridge, R. Mesthrie and C. Upton, eds., *A handbook of varieties of English: A multimedia reference tool. Vol. 1: Phonology*, 1st ed. Berlin: De Gruyter Mouton, pp. 729–749.

This chapter offers a description of the phonology of Hawai‘i Creole English. It provides numerous examples alongside the detailed description of all phonological and phonetic features of the creole.

Sakoda, K. and Siegel, J. (2004). Hawai‘i Creole: Morphology and syntax. In: B. Kortmann, E. Schneider, K. Burridge, R. Mesthrie and C. Upton, eds., *A handbook of varieties of English: A multimedia reference tool. Vol. 2: Morphology and syntax*, 1st ed. Berlin: De Gruyter Mouton, pp. 742–769.

This chapter offers a detailed description of Hawai‘i Creole English’s morphological and syntactic features. Numerous examples are provided, which help the reader grasp the complexity of this creole.

Walworth, M. (2017b). Reo Rapa: A Polynesian contact language. *Journal of Language Contact*, 10(1), pp. 98–141.

This article presents the sociohistorical situation of Tahitian dominance on the island of Rapa Iti, and describes in detail the ways in which this situation has brought about language change. The article defines a new type of mixed language, Reo Rapa, and provides details of its structure and use as an in-group identity marker.

7. Related topics

Social factors, borrowing, creoles and pidgins, mixed languages

Abbreviations

PEP Proto Eastern Polynesian
HAW Hawaiian

| | |
|-----|------------|
| TAH | Tahitian |
| RPN | Rapanui |
| MGV | Mangarevan |
| RAR | Rarotongan |
| MAO | Māori |
| OR | Old Rapa |
| MQS | Marquesan |

References

- Bauer, L. (1994). English in New Zealand. In: R. Burchfield, ed., *The Cambridge history of the English language*, 1st ed. Cambridge: Cambridge University Press, pp. 382-429.
- Bauer, W., Parker, W., Evans, T. and Teepe, T. (1997). *The reed reference grammar of Maori*. Birkenhead, NZ: Reed Books.
- Bell, A. (2000). Māori and Pakeha English: A case study. In: A. Bell and K. Kuiper, ed., *New Zealand English*, 1st ed. Wellington: Victoria University Press, pp. 221-248.
- Benton, R. (1966). *Research into the English language difficulties of Māori schoolchildren 1963-1964*. Wellington: Māori Education Foundation.
- Bickerton, D. and Odo, C. (1976). *Change and variation in Hawaiian English Vol I: General phonology and pidgin syntax: Final report on national science foundation grant No. GS-39748*. University of Hawaii, Social Sciences and Linguistics Institute.
- Biggs, B. (1971). The languages of Polynesia. In: T. Sebeok, ed., *Current trends in linguistics: Linguistics in Oceania*, 1st ed. The Hague: Mouton, pp. 466-505.
- Charpentier, J. and François, A. (2015). *Linguistic Atlas of French Polynesia/Atlas linguistique de la Polynésie française*. Berlin: Mouton de Gruyter.
- Collerson, K. and Weisler, M. (2007). Stone Adze compositions and the extent of Ancient Polynesian voyaging and trade. *Science*, 317(5846), pp. 1907-1911.
- Constitutional Council of French Polynesia (2004). *Organic law no 2004-192*. Pape'ete.
- Drager, K. (2012). Pidgin and Hawai'i English: An overview. *International Journal of Language, Translation and Intercultural Communication*, 1, p. 61.
- Fare Vāna'a. (2017). Dictionnaire. <http://www.farevanaa.pf/dictionnaire.php>.
- Fisher, D. (2013). *France in the South Pacific: Power and politics*. Canberra: ANU E Press.
- Green, R. (1966). Linguistic subgrouping within Polynesia: The implications for prehistoric settlement. *Journal of the Polynesian Society*, 75(1), pp. 6-38.
- Hermann, A., Sauzéat, L., Guillou, H., Maury, R., Chauvel, C., Liorzou, C. and Conte, E. (2017). Combined geochemical and geochronological analyses of stone artefacts provide unambiguous evidence of intra- and inter-island interactions in Polynesia. *Journal of Archaeological Science: Reports*, 13, pp. 75-87.
- Holmes, J. (1997). Maori and Pakeha English: Some New Zealand social dialect data. *Language in Society*, 26(1), pp. 65-101.
- Holmes, J. and Ainsworth, H. (1996). Syllable-timing and Māori English. *Te Reo*, 39, pp.75-84.
- Institut de la statistique de la Polynésie française (ISPF). (2012). *Synthèses locales*. Available at: ispf.pf/bases/Recensements/2012.aspx.
- Kawamoto, K. (1993). Hegemony and language politics in Hawaii. *World Englishes*, 12(2), pp. 193-207.
- Keegan, P. (2019). *Maori language information, Maori words in New Zealand English*. [online] Maorilanguage.info. Available at: www.maorilanguage.info/mao_defns.html.
- Kieviet, P. (2017). *A grammar of Rapa Nui*. Berlin: Language Science Press.
- King, J. (1993). Māori English: A phonological study. *New Zealand English Newsletter*, 7, pp. 33-47.
- Kirch, P. (2017). *On the road of the winds*. Berkeley: University of California Press.
- Labov, W. (1972). *Sociolinguistic patterns*. Philadelphia: University of Pennsylvania Press.
- Love, S. (2006). *Tahitian French*. PhD. Australian National University.

- Maclagan, M. and King, J. (2007). Aspiration of plosives in Māori: Change over time. *Australian Journal of Linguistics*, 27(1), pp. 81–96.
- Maclagan, M., King, J. and Gillon, G. (2008). Maori English. *Clinical Linguistics and Phonetics*, 22(8), pp. 658–670.
- Makihara, M. (2005). Rapa Nui ways of speaking Spanish: Language shift and socialization on Easter Island. *Language in Society*, 34(5), pp. 727–762.
- Matras, Y. and Bakker, P. (2003). *The mixed language debate*. Berlin: De Gruyter Mouton.
- Meakins, F. (2013). Mixed languages. In: P. Bakker and Y. Matras, eds., *Contact languages: A comprehensive guide*, 1st ed. Boston: Mouton de Gruyter, pp. 159–228.
- Meakins, F. (2014). Language contact varieties. In: H. Koch and R. Nordlinger, eds., *The languages and linguistics of Australia: A comprehensive guide*, 1st ed. Berlin: Mouton de Gruyter, pp. 365–416.
- Meyerhoff, M. (1994). Sounds pretty ethnic, eh? A pragmatic particle in New Zealand English. *Language in Society*, 23(3), pp. 367–388.
- Mufwene, S. (2001). *The ecology of language evolution*. Cambridge: Cambridge University Press.
- Mufwene, S. (2015). Colonization, indigenization, and the differential evolution of English: Some ecological perspectives. *World Englishes*, 34(1), pp. 6–21.
- Nicole, J. (1987). Manuscript XIX – The first missionary text in a Polynesian language. *The Journal of Pacific History*, 22(2), pp. 94–101.
- Reinecke, J.E. (1969). *Language and dialect in Hawaii*. Honolulu: University of Hawai‘i Press.
- Rolett, B. (2002). Voyaging and interaction in ancient East Polynesia. *Asian Perspectives*, 41(2), pp. 182–194.
- Rutter, K. (2006). *Going French, going Tahitian: The Tahitianization of French Polynesia*. Tenth International Conference on Austronesian Linguistics, 17–20 January 2006, Palawan, Philippines.
- Sakoda, K. and Siegel, J. (2003). *Pidgin grammar: An introduction to the Creole Language of Hawai‘i*. Honolulu: Bess Press.
- Sakoda, K. and Siegel, J. (2004a). Hawai‘i creole: Phonology. In: B. Kortmann, E. Schneider, K. Burridge, R. Mesthrie and C. Upton, eds., *A handbook of varieties of English: A multimedia reference tool. Vol. 1: Phonology*, 1st ed. Berlin: De Gruyter Mouton, pp. 729–749.
- Sakoda, K. and Siegel, J. (2004b). Hawai‘i Creole: Morphology and syntax. In: B. Kortmann, E. Schneider, K. Burridge, R. Mesthrie and C. Upton, eds., *A handbook of varieties of English: A multimedia reference tool. Vol. 2: Morphology and syntax*, 1st ed. Berlin: De Gruyter Mouton, pp. 742–769.
- Sato, C.J. (1993). Language change in a creole continuum: Decreolization? In: Kenneth Hyldenstam and Åke Viberg, eds., *Progression and regression in language: Sociocultural, neuropsychological and linguistic perspectives*. Cambridge: Cambridge University Press, pp. 122–143.
- Saura, B. (2009). *Tahiti mā‘ohi*. Pirae: Ed. Au Vent des Iles.
- Siegel, J. (2010). Contact languages of the Pacific. In: R. Hickey, ed., *The handbook of language contact*, 1st ed. Chichester: Wiley-Blackwell, pp. 814–836.
- Statistics New Zealand (2013). *Te Kupenga*. Available at: http://archive.stats.govt.nz/browse_for_stats/people_and_communities/maori/TeKupenga_HOTP13/Commentary.asp
- Stueber, R.K. (1964). *Hawai‘i: A case study in development education 1778–1960*. PhD. University of Wisconsin.
- Szakay, A. (2007). *Identifying Māori English and Pākehā English from suprasegmental cues: A study based on speech resynthesis*. MA thesis. University of Canterbury.
- Trudgill, P. (1972). Sex, covert prestige and linguistic change in the urban British English of Norwich. *Language in Society*, 1(2), pp. 179–195.
- Ulukau. (2019). Nā Puke Wehewehe ‘Ōlelo Hawai‘i. <http://wehewehe.org/gsd12.85/cgi-bin/hdict?l=haw>.
- United States Census Bureau (2018). *Population estimates by state*. Available at: [census.gov/programs-surveys/popest.html](https://www.census.gov/programs-surveys/popest.html).
- Walworth, M. (2014). Eastern Polynesian: The linguistic evidence revisited. *Oceanic Linguistics*, 53(2), pp. 256–272.

- Walworth, M. (2015a). *The language of Rapa Iti: Description of a language in change*. PhD. University of Hawai'i at Mānoa.
- Walworth, M. (2015b). Trickle-down endangerment. Presented at *The New Zealand linguistics society meeting*. University of Otago, Dunedin.
- Walworth, M. (2017a). Classifying Old Rapa: Linguistic evidence for contact networks in South-east Polynesia. *Journal of the Southeast Asian Linguistics Society, Special Publication*, 1, pp. 102–122.
- Walworth, M. (2017b). Reo Rapa: A Polynesian contact language. *Journal of Language Contact*, 10(1), pp. 98–141.
- Weisler, M. (1997). Prehistoric long-distance interaction at the margins of Oceania. In: M. Weisler, ed., *Prehistoric long-distance interaction in Oceania: An interdisciplinary approach*, 1st ed. New Zealand Archaeological Association, pp. 149–172.

Linguistic Melanesia

Antoinette Schapper

1. Introduction and definitions

Linguistic Melanesia is a world hotspot of linguistic diversity and is home to around 1500 languages belonging to between 20 to 40 language families. Located to the north of the Australian continent, the area is centred on the island of New Guinea, extending eastward from the island of Sulawesi in Indonesia to the western fringe of Polynesia and stopping just of the islands of Fiji, Vanuatu, and New Caledonia. At its core Linguistic Melanesia is dominated by Papuan languages, but also takes in a large number of Austronesian languages (Figure 24.1). While Austronesian languages form a genealogical unit, Papuan languages do not. A language is said to be Papuan, if it is spoken on or near New Guinea, and is not Austronesian or Australian. The term ‘Papuan’ thus covers languages of numerous families. The alternative label ‘non-Austronesian’ is sometimes used, but presents its own difficulties in that Austronesian languages are in geographic contact with several language families, such as Austro-Asiatic or Sinitic, which are non-Austronesian but not Papuan. Throughout this chapter, I will use the term ‘Papuan.’

The antecedents of modern-day Papuan languages, along with the Australian languages, have been traced back to the earliest waves of migration out of Africa between 40,000–60,000BP (O’Connell, 2012; Tumonggor et al., 2013; Macaulay et al., 2005). While the vast majority of Papuan languages are located on New Guinea, there are around 60 so-called Papuan outliers scattered around New Guinea. Due to the limited documentation available for many Papuan languages, their classification into families remains controversial and largely a matter of conjecture. Trans-New Guinea (TNG) is a hypothesized macro-family of around 500 languages whose members are spread across the mountainous cordillera along the length of New Guinea and into many lowland regions, particularly on the south coast of New Guinea, as well as to the island of Timor and its satellites several hundred kilometres to the west of New Guinea. There is little agreement on the precise membership or higher subgroupings of the Trans-New Guinea family. The remaining Papuan languages, collectively labelled non-Trans-New Guinea, are classified as belonging to several dozen other language families that are not related to one another, although the exact number and constituency of these is not agreed upon. Figure 24.2 presents a relatively ‘lumping’ statement of Papuan families, that is, it represents proposed larger family groupings even where they are not definitively proven.

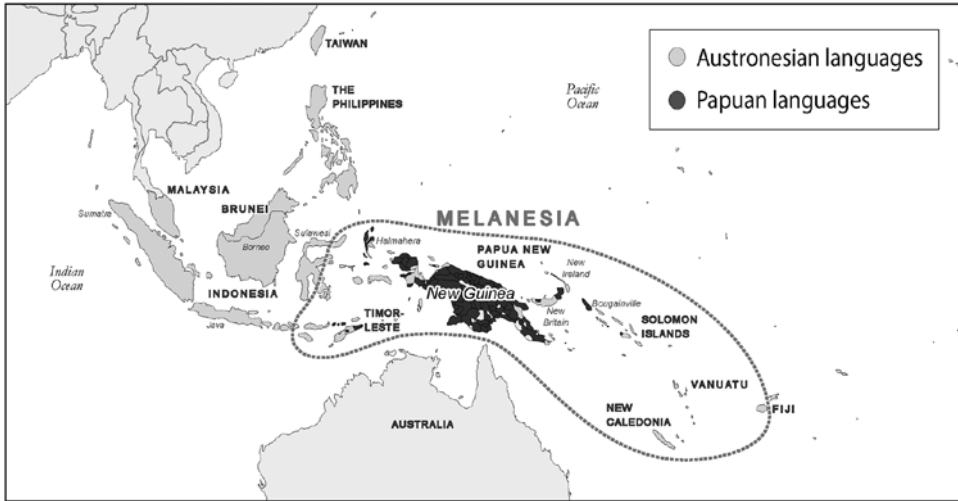


Figure 24.1: Approximate location of Linguistic Melanesia

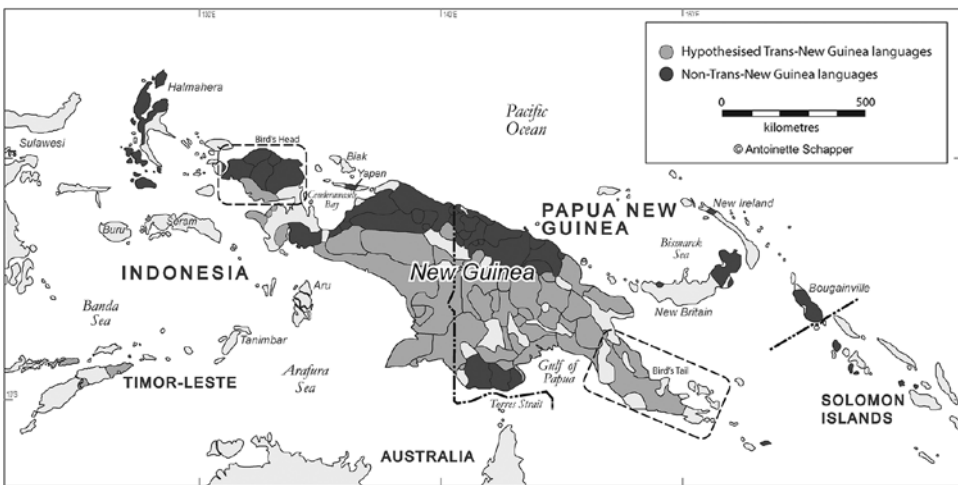


Figure 24.2: Distribution of Papuan languages on and around New Guinea†.

†The island of New Guinea is often conceived of as a bird and languages are frequently referred to by their location on this metaphorical bird. The so-called Bird's Head and Bird's Tail are marked on this figure for reference.

The spread of the Austronesian languages is associated with an out-of-Taiwan migration of Asians in the period between 8,000BP and 4,000BP (Bellwood, 2017; Hill et al., 2007; Tabbada et al., 2010). This transformative period, called the 'Southeast Asian Neolithic,' saw many pre-existing populations in the path of the expanding Austronesians overwhelmed and assimilated to become Austronesian-speaking. Groups of Austronesian speakers also moved out into the Pacific Ocean to be the first settlers of Fiji, Vanuatu, New Caledonia as well as the far-flung islands of Micronesia and Polynesia. On and around New Guinea, early speakers of

Austronesian languages met with more resistance from the already present Papuan-language speaking populations (Bellwood, 1998). Austronesian languages are found in only small, coastal enclaves on the New Guinea mainland, while numerous, albeit isolated, Papuan language families are scattered amongst the Austronesian languages in the insular areas around New Guinea.

Austronesian languages show a progressive convergence on the linguistic norms of Papuan languages the closer they are to New Guinea. This attenuation of Austronesian features to Papuan ones results in concentric circles of linguistic features clustering around New Guinea. On New Guinea, contact between Papuan languages has also led to smaller convergence areas in which linguistic features can be seen to have diffused between languages. This macro-area in which Papuan languages have converged with each other and Austronesian languages with Papuan languages is referred to here as ‘Linguistic Melanesia’ (or ‘Melanesian Linguistic Area’).

Whilst the label ‘Linguistic Melanesia’ as used here is not one that is yet well established in the literature, the concept of a sphere of Papuan linguistic influence around New Guinea will be familiar to all linguists working in the area. In its original use by Dumont d’Urville (1832) and as continued by many others, ‘Melanesia’ was a geographic term that was intended to denote a region of islands inhabited by dark-skinned peoples taking in all of New Guinea and the islands to its east (excluding Polynesia and Micronesia), but not those to its west (see Green, 1991 on the term Melanesia). In the earliest linguistic works in the region, ‘Melanesian’ languages are a subset of Austronesian languages, contrasting with the structurally different, but still related ‘Indonesian’ and ‘Polynesian’ types of Austronesian languages (see, e.g., von der Gabelentz, 1860; Codrington, 1885; Ray, 1907 for this obsolete usage; see Blust, 2009 for an up-to-date terminology for the Austronesian family). Neither the original geographic designation nor the early linguistic use of ‘Melanesia’ has much diachronic traction. Papuan languages and ‘Melanesian’ phenotypes are found in the region to the west of New Guinea (known as ‘Wallacea,’ Schapper, 2015), indicating that insular Papuan contact has also here played a formative role as to the east of New Guinea. The use of ‘Linguistic Melanesia’ here to define the whole sphere of Papuan languages and their inferred influences is thus intended to delimit a linguistic area in a principled and consistent way.

2. Historical overview

What is Linguistic Melanesia? In this section I present a synthesis of some key areal features of Linguistic Melanesia that have been identified in the literature. There are numerous overviews of typological characteristics of Melanesian languages (e.g., Foley, 2000; Aikhenvald and Stebbins, 2007), but the sheer number of languages in the area means these treatments are often skewed towards a small number (of subsets) of languages. Where broader linguistic features have been identified, they have typically been treated individually in disparate specialist publications on the region and have seldom been brought together to define an area of ‘Linguistic Melanesia.’ The approach taken here is to map the synchronic distribution of a range of morphosyntactic, phonological, and lexico-semantic features which can be used to demarcate Linguistic Melanesia or significant portions thereof as a convergence zone. Critical in defining the area and the numerous sub-areas of Linguistic Melanesia is the opposition between a feature found inside Linguistic Melanesia and its (non-)appearance in the geographically adjoining areas. As a result, each feature will be plotted in relation to the full spread of Austronesian languages both in and outside the Linguistic Melanesia. Occasional reference will also be made to the typological behaviour of Australian languages.

2.1 Phonological features

Cross-linguistic patterns in phonology have been relatively little explored for Linguistic Melanesia. Donohue and Whiting (2011) represents one of the few in-depth studies and is particularly valuable for its highlighting of the complexities of detecting areality for non-binary features in a linguistically diverse area like Melanesia. Nonetheless, there are several features of Melanesian phonology that are well-known from the literature.

One of the most salient phonological features of Linguistic Melanesia is phonemic tone (Figure 24.3). Tonal systems in Melanesia show a wide variety of types, from syllable tone through word tone to pitch accent (Donohue, 1997); due to insufficient descriptive materials for many, the variable we will consider here is merely the presence versus absence of tonal contrasts and not the types of contrasts made. More than 84% of Papuan languages in our sample have tonal contrasts, with the feature strongly concentrated on New Guinea. Tone is only found in less than 1% of Austronesian languages in our sample. Of these tonal languages, more than 80% are found inside Linguistic Melanesia and the vast majority are on found on New Guinea.

Tonal contrasts are overwhelmingly found in languages of cordillera of New Guinea and appear to have spread readily to languages which are in contact with them (Donohue, 2005a). At the same time, tone is notably absent across families in several subareas of Linguistic Melanesia. Tone is vanishingly rare in Papuan languages off-shore from New Guinea where Austronesian languages dominate; of the 60 Papuan outlier languages, only one, Abui, has been confirmed to have tonal contrasts, and even then, only in a small number of lexical items (Delpada, 2016), while one other language, Fataluku, possibly has a pitch accent system. On New Guinea, tone tends to be absent in distinct areas that are away from the central cordillera. In Figure 24.3, we see that exceptions to tonality in Papuan languages are concentrated in three sub-areas at the edges of New Guinea: (1) southwest New Guinea and Bird's belly

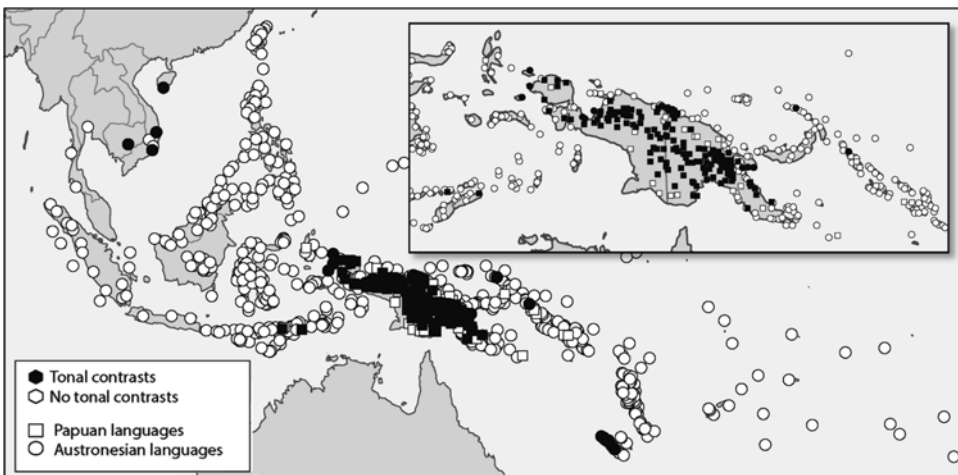


Figure 24.3 Presence of tonal contrasts in Linguistic Melanesia

Source: The map is adapted from Donohue et al. (2013) and Maddieson (2013), supplemented with data from Arnold (2018), Kamholz (2017) and Donohue (1997, 2005b)

Note: This map presents 861 language varieties (languages and language dialects), 666 Austronesian and 195 Papuan. Of these, 22 Austronesian and 164 Papuan languages have tonal contrasts.

area spanning Trans-New Guinea languages such as coastal varieties of Asmat and Marind as well as lowland Awyu-Dumut languages, and the various families of the Morehead and Fly river regions; (2) northeast New Guinea spanning a number of families ranging from many Madang languages of TNG family through middle Sepik region languages such as Yimas, Alamlak, Yessan-Mayo, and languages of the Ndu group to languages of the border region such as Imonda, Dla, and Momu, and; (3) the Bird's Tail area of New Guinea where toneless languages from several hypothesized Trans-New Guinea subgroups are concentrated. The geographical skewing and cross-familial character of both tone and toneless languages points strongly to diffusion as being an important part of the explanation of the distribution of tonality in Papuan languages.

Tonal contrasts are rare in the Austronesian language family, and their appearance is strikingly skewed to two areas. In the far west of the Austronesian area, a few Chamic languages have developed tonal contrasts in contact with tonal languages of Mainland Southeast Asia (Kirby and Brunelle, 2017). All other cases of tonal Austronesian languages in our sample are within Linguistic Melanesia. Tonal Austronesian languages are found in several pockets of Linguistic Melanesia, notably in Raja Ampat, in southern Cenderawasih Bay, in the Huon Gulf, and in New Caledonia, as well as in two isolated places of Island Melanesia, namely, Kara (New Ireland, Oceanic) and Tinputz (Bougainville, Oceanic). Tonogenesis in some of these groups has been seen to be the result of regular processes of historical change (e.g., out of voicing status and harmony of consonants within morphemes, Ross, 1993; out of geminates and aspirated consonants, Haudricourt, 1968). In various groups the development of tone has been argued to be the result of independent parallel innovations (Rivierre, 1993; Kamholz, 2017; Arnold, 2018). Whatever the specific historical sources of tone in the different groups, the repeated emergence of tone in Austronesian languages in Linguistic Melanesia suggests that areal pressure from Papuan languages has resulted in tonal contrasts erratically diffusing into Austronesian languages.

The lack of a contrast between a lateral and a rhotic phoneme is frequently observed as a particular feature of Papuan languages (e.g., Foley, 1986, pp. 55–56). Having a single liquid phoneme is indeed characteristic of much of the core Melanesian Linguistic Area (Figure 24.4). In our sample 71% of Papuan languages have a single liquid phoneme, while just 26% of Austronesian languages have a single liquid phoneme. The overwhelming absence of liquid contrasts in Papuan languages is particularly striking given the large number of liquid distinctions typically made in Australian languages.

In Austronesian languages, the correlation of a single liquid contrast with a location inside Linguistic Melanesia is only weak, with 57% of single liquid Austronesian languages in our sample occurring within Linguistic Melanesia. This indicates that the lack of a liquid contrast in Austronesian languages is not easily attributable to convergence with the norms of Papuan languages. By contrast, however, multiple liquid contrasts in Papuan languages does, with one exception, correlate with areas in which Austronesian influences are known. As seen also for the lack of tonal contrasts, Papuan languages with more than one liquid phoneme are concentrated in the maritime and coastal regions of New Guinea; it is again the multi-family subareas of southwestern New Guinea and of the north-central coast of New Guinea as well as all the Papuan 'outliers' that do not display the majority Papuan pattern for liquids. A further exceptional area is the eastern highlands where languages from different TNG families meet. Here we find unusual arrays of liquid phonemes that are not found elsewhere. For example, Huli has a trill /r/ and a retroflex lateral approximant /l̠/; Enga contrasts a retroflex flap /ɾ/ and a palatal lateral /ɕ/; Kobon distinguishes an alveolar lateral /l/, a palatal lateral /ɕ/, a subapical retroflex

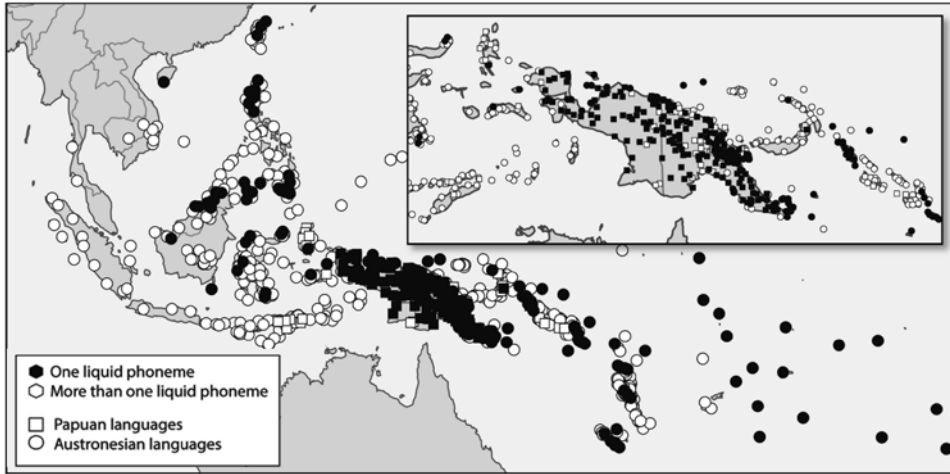


Figure 24.4 Presence of one liquid phoneme in Linguistic Melanesia

Source: The map is adapted from Donohue et al. (2013)

Note: This map presents 870 language varieties (languages and language dialects), 623 Austronesian and 247 Papuan. Of these, 166 Austronesian and 177 Papuan languages have one liquid phoneme.

lateral flap / \check{l} /, and a fricative trill / r /'. The small region where languages with these unusual lateral contrasts are found constitutes its own subarea distinct from the coastal and maritime Melanesian pattern where the typical contrast is between / l / and / r / [r ~ r].

The Melanesian Linguistic Area also stands out for its lack of a velar nasal phoneme (Figure 24.5). The vast majority of Austronesian languages have the velar nasal in their consonant phoneme inventories, amounting to more than 82% of the sample. Around New Guinea there is a clear concentration of Austronesian languages lacking the velar nasal. The lack of the velar nasal in Austronesian languages begins in Timor, moves through South-west Maluku, Central Maluku and the languages of Bomberai Peninsula and Cenderawasih Bay. On the north coast of New Guinea roughly half the Austronesian languages lack the velar nasal, while once on the Bird's Tail of New Guinea almost all Austronesian languages lack it. A smattering of Austronesian languages without velar nasals are then found in the Bismark archipelago and Northern Vanuatu. Beyond this, we find only two Austronesian outliers to the West (Nias and Enggano) in the Barrier Islands off the West coast of Sumatra and five outliers in remote Polynesia.

Papuan languages are divided roughly in half in terms of the velar nasal: a velar nasal phoneme is present in 221 Papuan languages, and absent in 293 (57%). Viewed in isolation then the lack of the velar nasal in Melanesia doesn't seem remarkable; however, taking a larger areal perspective encompassing Mainland Southeast Asia and Australia where the velar nasal is near-universally present, the absence of / η / becomes a highly marked feature of the Melanesian area (Anderson, 2013). What is more, velar nasal-lacking Papuan languages are concentrated in the maritime and coastal regions to the West of New Guinea and from the central northern region down off the Bird's Back and into the Bird's Tail region, precisely the regions where velar nasal-lacking Austronesian languages are most found.

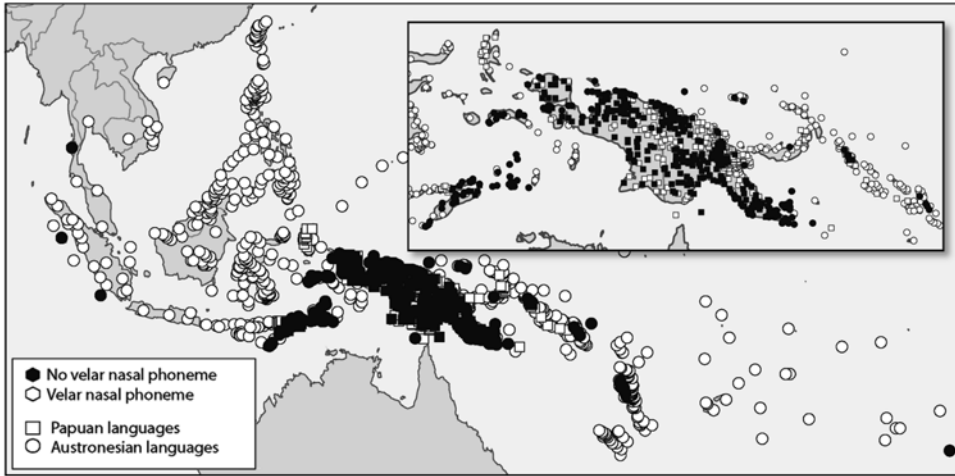


Figure 24.5 Absence of velar nasal phonemes in Linguistic Melanesia

Source: The map is adapted from Donohue et al. (2013)

Note: This map presents 1339 language varieties (languages and language dialects), 825 Austronesian and 514 Papuan. Of these, 142 Austronesian and 293 Papuan languages lack the velar nasal.

2.2 Word order features

Word order features are the most thoroughly studied and well-understood typological variable in Linguistic Melanesia. Differences in the positions of adpositions and clause joiners, and in the order of noun and demonstrative, of noun and adjective, of verb and negator and of verb and aspectual marker are among those that have been discussed for Linguistic Melanesia (Bradshaw, 1982; Donohue, 2007; Reesink, 2002; Klamer, Reesink and van Staden, 2008). Three features of Melanesian word order will be discussed here to illustrate general patterns of word order convergence over the area.

Linguistic Melanesia shows a high degree of consistency at its core in terms of the ordering of subject, object and verb. Papuan languages of almost all families are overwhelmingly SOV (Figure 24.6): nearly 85% of sampled Papuan languages have a dominant SOV word order. SOV is rare in Austronesian languages, but all instances of it are found in Austronesian languages spoken on the coast of New Guinea in regions contiguous with Papuan SOV languages. The Austronesian languages on New Guinea that lack SOV are typically found in regions where Papuan languages do not have a strong presence. Although absent in Austronesian languages outside New Guinea, Papuan outliers frequently display SOV: to the west of New Guinea, all Timor-Alor-Pantar languages, Kalamang off Bomberai peninsula, Yawa on Yapen, and Tobelo on Halmahera have SOV; to the east of New Guinea the Papuan languages of the Solomon Islands and Yele are almost entirely SOV. In short, SOV is the modal order for Linguistic Melanesia, even if Austronesian languages do not frequently exhibit it.

Exceptions to SOV word order in Papuan languages are concentrated in (1) languages of Bird's Head and the outliers of the North Halmahera family, (2) languages of the Torricelli family on the north central coast of New Guinea, and (3) Papuan languages of the Bismark archipelago (New Ireland, New Britain, and Bougainville islands). With the exception of VSO Kuot (New Ireland) and Bilua (Solomons) with no dominant order, these Papuan languages display SVO order. Austronesian influence has been assumed to be responsible for this word

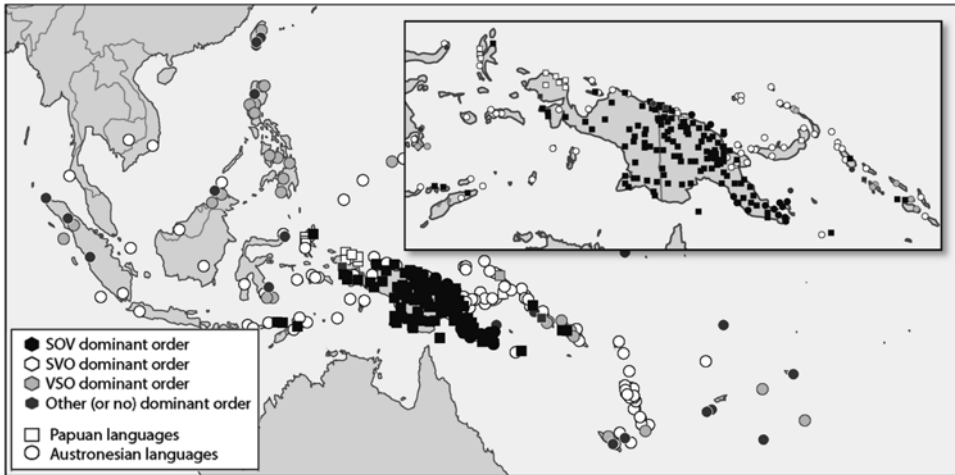


Figure 24.6 Dominant order of subject, verb, and object in Linguistic Melanesia

Source: Adapted from Dryer (2013a) and follows the argumentation of Donohue (2007)

Note: This map presents 325 languages, 171 Austronesian and 154 Papuan. Of these, 14 Austronesian and 130 Papuan languages have SOV in main clauses.

order by many authors (e.g., Voorhoeve, 2004; Reesink, 2005). However, because SVO is itself an innovation in Austronesian languages replacing the conservative VSO Austronesian order, Donohue (2005b, 2007) suggests that SVO characterizes an ancient subarea that extended over the top of New Guinea and through much of today's Indonesia (see also Gil, 2015).

The order of the noun and its possessor (or so-called genitive) is another feature which can be used to define Linguistic Melanesia. As is clear from Figure 24.7, almost all Papuan languages (95% in the sample) have the genitive preceding the noun. An areally very restricted pocket of Papuan languages with noun-genitive order on the north-central coast as well as Kuot on New Ireland are the only exceptions to genitive-noun order in Papuan languages. The typical Austronesian pattern outside of Melanesia has the genitive following the noun. By contrast, within Linguistic Melanesia, Austronesian languages have the order of the genitive 'reversed,' that is, typically preceding the noun. Austronesian languages with the reversed genitive are heavily concentrated in Melanesia (two languages in Micronesia, Ulithian and Puluwat, are the only outliers). The feature is dispersed throughout eastern Indonesia and into New Guinea where it is consistently present in Austronesian languages, but not beyond into the Bismarck Archipelago or further afield into the Pacific.

The order of the numeral and the noun it enumerates also changes in Austronesian languages with proximity to New Guinea (Figure 24.8). In most Austronesian languages outside of Linguistic Melanesia, the numeral typically precedes the noun. By contrast, in Austronesian languages within the area the numeral follows the noun. The vast majority of Papuan languages (over 90% in the sample) have the numeral following the noun. Accordingly, we again see a clear skewing of noun-numeral order in Austronesian languages towards Melanesia. With the exception of three outliers west of Melanesia, the feature extends unbrokenly from Timor to New Guinea, and is almost invariably present in Austronesian languages of the New Guinea mainland and further into Vanuatu. The Austronesian languages of the Bismarck Archipelago, New Caledonia, Micronesia, and Polynesia are in their majority numeral-noun order languages, but each region has a few exponents of the Melanesian noun-numeral order.

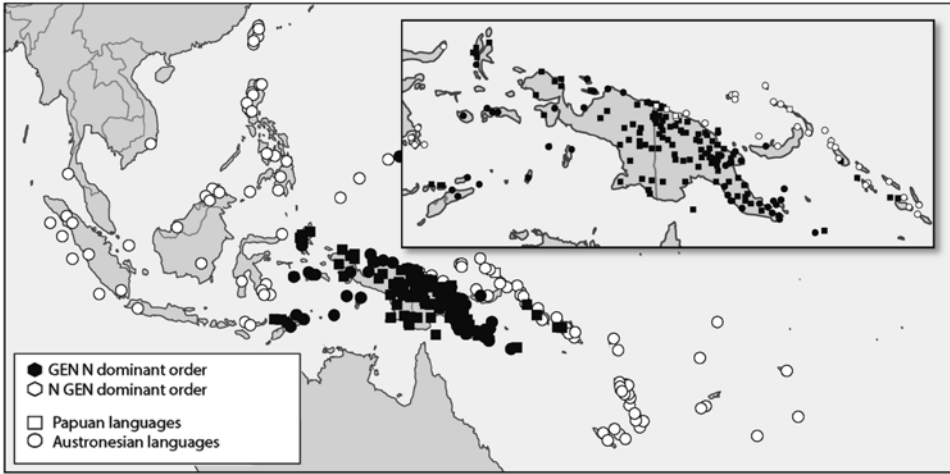


Figure 24.7 Dominant order of noun and genitive in Linguistic Melanesia

Source: Adapted from Dryer (2013b) and follows the argumentation of Donohue (2007)

Note: This map presents 277 languages, 155 Austronesian and 122 Papuan. Of these, 43 Austronesian and 116 Papuan languages have the genitive preceding the noun.

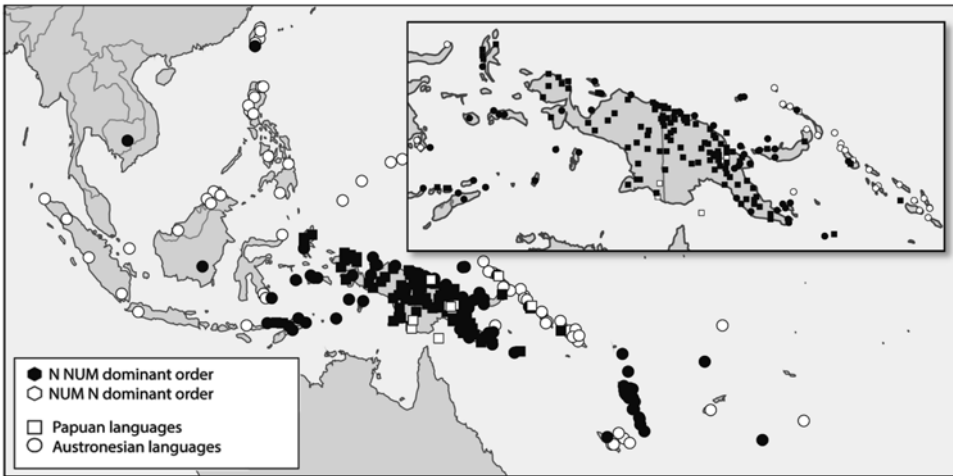


Figure 24.8: Dominant order of noun and numeral in Linguistic Melanesia

Source: The map uses the data from Dryer (2013c) and follows the argumentation of Donohue (2007).

Note: This map presents 266 languages, 145 Austronesian, and 121 Papuan. Of these, 78 Austronesian and 111 Papuan languages have the numeral following the noun.

2.3 Lexical and semantic features

Lexical and semantic features cover a vast domain of linguistic structure, covering anything from the sense ranges of individual lexemes, through the structure of word compounds, lexical collocations, and constructions, to the organization of entire semantic domains and even whole lexicons. Convergence in lexico-semantics results in ready inter-translatability between

languages and can be observed to have occurred in many domains in Linguistic Melanesia. Areal lexico-semantic features that have been described for (parts of) Linguistic Melanesia include evidential categories (San Roque and Loughnane, 2012a, 2012b), patterns of suppletion in kin terms (Baerman, 2014) and lexicalized alienability distinctions in possessive constructions (Klamer, Reesink and van Staden, 2008; Schapper, 2015).

An example of convergence in lexical senses in Linguistic Melanesia is the frequent colexification of ‘fire’ and ‘firewood,’ a polysemy pattern in which a language uses one and the same lexeme to refer to both fire and firewood. Figure 24.9 sets out the conspicuous areal skewing of fire/firewood colexification to Linguistic Melanesia. Whilst only 47% of sampled Papuan languages display the pattern, the appearance of fire/firewood colexification in Australian languages and the languages of other genetically Austro-Melanesian peoples of Southeast Asia indicates that the pattern represents an ancient feature that was once modal in Papuan languages (Schapper, 2017; Schapper, San Roque and Hendersy, 2016). The distribution of fire/firewood colexification in Austronesian confirms the Melanesian status of the pattern: over 90% of Austronesian languages with fire/firewood colexification were found within Linguistic Melanesia. The inherited Austronesian pattern involves distinct lexemes for the two senses (Proto-Malayo-Polynesian *hapuy ‘fire’ and *aliten ‘firewood’). Examination of the etyma for ‘fire’ and ‘firewood’ in the sampled Austronesian languages indicates that fire/firewood colexification was innovated dozens of times independently, indicating erratic diffusion by either contact with or shift from Papuan languages (Schapper, 2017).

A second example of lexico-semantic convergence can be seen in the lexical constructions involved in expressing comparison of inequality (as in an English clause such as *Hildegard is nicer than Ludwig*). The types for this feature also define concentric circles of convergence around the island of New Guinea (Figure 24.10). The core of Linguistic Melanesia is characterized by the exclusive use of the conjoined comparative, a construction in which the standard of comparison and the comparee appear in separate clauses (as in *Hildegard is nice, Ludwig*

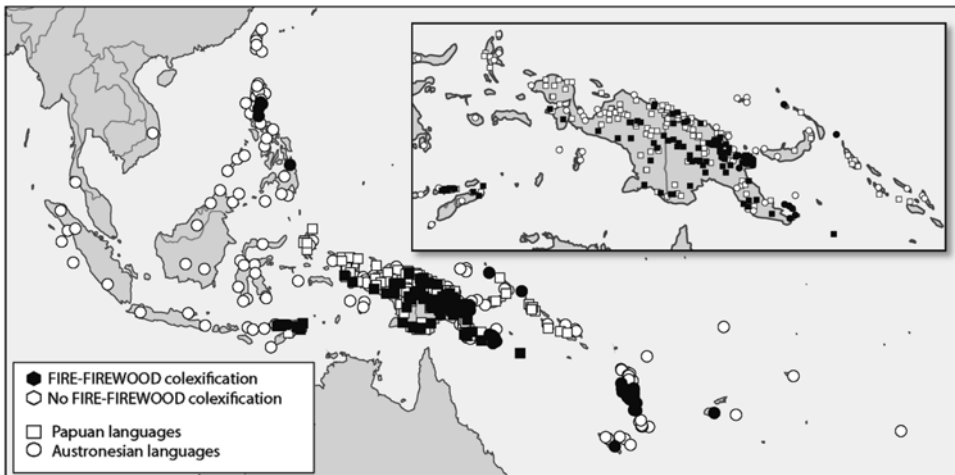


Figure 24.9 Colexification of ‘fire’ and ‘firewood’ in Linguistic Melanesia

Note: This map presents 412 languages, 197 Austronesian, and 215 Papuan. Of these, 51 Austronesian languages and 96 Papuan languages have fire/firewood colexification. Source: Data from Schapper, San Roque and Hendersy (2016) and Schapper (2017)

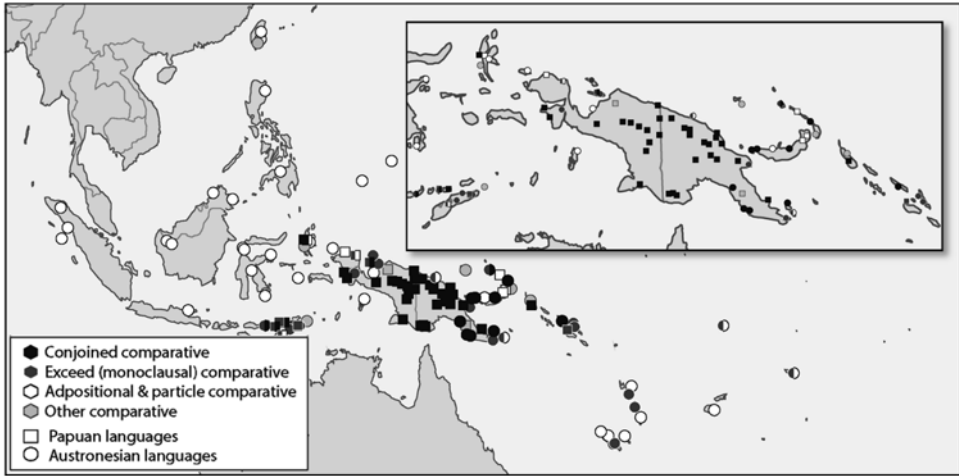


Figure 24.10 Comparative constructions in Linguistic Melanesia

Source: Adapted from Schapper and de Vries (2018)

Note: This map presents 116 languages, 68 Austronesian, and 48 Papuan. Of these, 16 Austronesian languages and 33 Papuan languages have conjoined comparatives.

is mean or *Hildegard is nice, Ludwig is not*). Of the Papuan languages sampled, 68% have conjoined comparatives and on the mainland of New Guinea, only three Papuan languages had no conjoined comparative. The conjoined comparative is found as the exclusive comparative strategy only in Austronesian languages within Melanesia. Outside of Melanesia, only three of the sampled Austronesian languages have conjoined comparatives, and this is always secondary to another non-conjoined strategy. In maritime Melanesia, the coastal and island belt around New Guinea, we find a construction shared between Papuan and Austronesian languages in which a verb with the meaning ‘exceed’ (or similar) is used to introduce the standard of comparison (as in, *Hildegard is nice exceeding Ludwig*). Although a common comparative strategy world-wide, that the exceed comparative is exclusively found in languages of Maritime Melanesia suggests an areal feature. Beyond Linguistic Melanesia, other comparative constructions dominate, with Austronesian languages typically possessing locative and particle comparatives.

A more complex instance of layers of lexico-semantic convergence around New Guinea is found in the structure of numeral and counting systems (Figure 24.11). The chief variable here is the base that a language’s numeral system uses, that is, the value upon which higher numeral expressions are constructed. A language may have more than one base, for instance, using 5 as the base to form numerals 6–9, and 10 for higher numerals. A language can lack a numeral system where it has no recursive base for the formation of higher numerals, meaning counting occurs only to an upper limit. The Melanesian Linguistic Area stands out for its lack of pure decimal (base-10) numeral systems. Proto-Austronesian (PAN) and its major daughters, Proto-Malayo-Polynesian and Proto-Oceanic, have reconstructed decimal systems. This pattern is maintained in the vast majority of Austronesian languages outside Linguistic Melanesia; only four Austronesian outliers in the west have bases other than 10 (Enggano, Ilongot, and Pazih use base 5–10, and Arta uses base 10–20). In Papuan languages, a pure decimal system is exceedingly rare with less than 8% of sampled Papuan languages evincing such systems. All

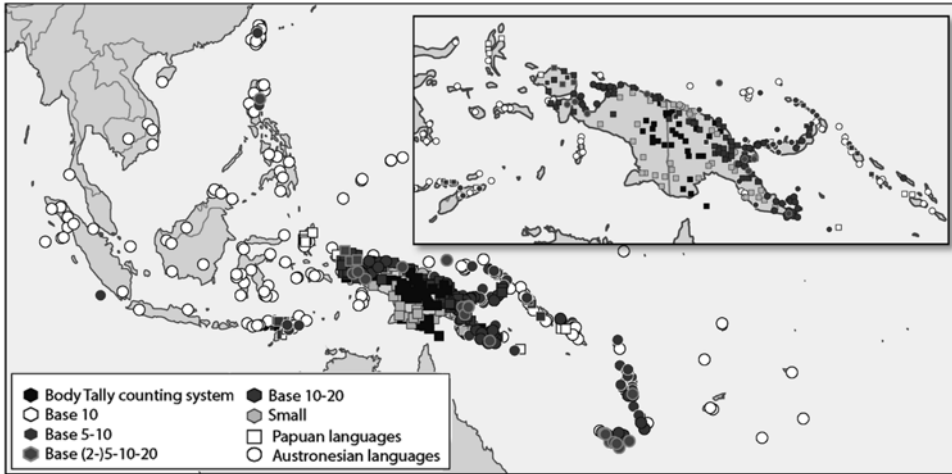


Figure 24.11 Numeral bases and counting systems in Linguistic Melanesia

Source: The data underlying this map is found in the numeral database at <https://mpi-lingweb.shh.mpg.de/numeral/>.

Note: This map presents 466 languages, 308 Austronesian, and 158 Papuan. Of these, the exclusive use of a decimal base is found in 121 Austronesian and 11 Papuan languages.

12 Papuan languages of this type are spoken off the New Guinea mainland and typically display significant Austronesian elements in their lexicons.

Papuan languages exhibit a large variety of counting systems with strong areal patterning across Linguistic Melanesia (Lean, 1992). The centre of New Guinea is dominated by body-tally systems in which counting is done on defined points on the body, extending up one side of the body and then down the other (Hammarström, 2010). Around this central core of body-tally languages we find languages with small counting systems in which there is no base and counting rarely goes higher than 10. Beyond these small systems, whilst there are a few marginal base types including small pockets of the typologically unusual base-4 and base-6 systems on the north and south coasts of New Guinea respectively, languages of the coastal and insular ring around New Guinea display different configurations of numeral systems involving base-5 and base-20. These are so common that over half of the Papuan languages in our sample have either base-5 and/or base-20 in their numeral systems. Although both these bases have good physiological motivations and can therefore be reasonably expected to emerge spontaneously, the concentration of Austronesian languages with base-5/20 in Maritime Melanesia indicates that significant areal pressure has given rise to them. This picture of convergence is reinforced by the observation that base-5 and base-20 have developed independently in even closely related Austronesian languages; over 50 distinct innovations of these bases have been documented for Austronesian languages of the Melanesian area (Blust, 2008; Dunn et al., 2008; Schapper and Hammarström, 2013).

2.4 Summary of Linguistic Melanesia

In this chapter we have looked at a small, but indicative set of features from phonology, word order, and (lexico-)semantics across which the diverse languages of Linguistic Melanesia display structural commonalities. We have seen that the Melanesian Linguistic Area is a complex

one, and as befits its linguistic diversity, it cannot be defined by simple bundles of isoglosses or by straightforwardly contrasting types. Rather Linguistic Melanesia is characterized by concentric circles of isoglosses clustering around New Guinea and into its interior. On the highest level, these circles represent changes in broad typological profiles and define large regions in which Austronesian languages incrementally converge on the linguistic norms of Papuan languages the closer they are to New Guinea. At the same time, Papuan languages in extensive contact with Austronesian languages often lack linguistic features typical of Papuan languages that are not in contact with Austronesian languages. We also find features shared between Papuan and Austronesian languages that are not typical of either group, but nonetheless define a convergence area, such as is the case with the ‘maritime’ Melanesian area, a coastal and insular ring around New Guinea. In numerous other parts of New Guinea, such as the Bird’s Head, central Highlands or southern New Guinea, we also find smaller areas of typological similarity between languages arising out of more localized processes of convergence.

The fact that Melanesian languages do not consistently show the same clusters of linguistic properties led Ross (2017, p. 806) recently to assert that the New Guinea area ‘is neither a geographic nor a typological area.’ It is true that there is no single ‘Melanesian linguistic type’; the cluster analyses of linguistic features by Reesink, Singer and Dunn (2009) and Nichols (1997) illustrate the kinds of divisions that can be made between Melanesian languages on the basis of statistical groupings of typological features. Yet, the absence of a coherent Melanesian type does not, in my view, negate the existence of a Melanesian Linguistic Area. Rather than the absolute adherence to a set of linguistic features, it is the contrast between the skewed presence of shared features in the Melanesian languages and the (near-)absence of those same features in the languages immediately outside the area which is important for defining Linguistic Melanesia. Convergence is dynamic and in such a large geographical region with so many small, unrelated or only distantly related language groups, low-level processes of language change will continuously act on the degree of similarity or dissimilarity between languages. In this context, it would be unexpected to have perfect isomorphism between any two given languages, let alone a large number. Evans (forthcoming) points out for the Morehead region of southern New Guinea that whilst there are certainly features shared amongst the unrelated languages of the region, a surprising amount of phonological and grammaticality diversity exists between languages. In short, Linguistic Melanesia is a complex area forged by competing forces of convergence and divergence.

3. Critical issues

How did Linguistic Melanesia come into being? This is a key issue for understanding the nature of language contact in the New Guinea area. If linguistic areas are, as commonly held, the outcome of diffusion of linguistic patterns across languages, contact between languages is a necessary pre-condition for the emergence of a linguistic area. A vast convergence area taking in hundreds of languages like that of Linguistic Melanesia must, therefore, arise substantially from bi- or multilingualism between speakers of chains of contiguous languages. This ‘chained’ language contact enables the diffusion of linguistic features across geographical space. Against this background, it becomes crucial to understand the nature of language contact in Linguistic Melanesia to appreciate how the area came into existence. In this section, I outline the interplay of convergent and divergent forces which have been described for the Melanesian Linguistic Area.

Linguistic Melanesia is thought to be an area over much of which people have in the past combined language loyalty with multilingualism. Sankoff (1976, p. 10) outlines how these

two were coupled together in traditional Melanesian society: ‘each group was ethnocentric about its own variety, but since groups were all very small, since people knew that other people thought their own was the best, and since within a region there was no consensus that a particular variety was the best, the situation was certainly an egalitarian one.’ ‘Egalitarian multilingualism,’ as it has come to be called, is a language ecology in which small speech communities maintain multilingual repertoires which include their own language alongside the languages of their neighbours in a stable way (Haudricourt, 1961; François, 2012). It is important to mention that egalitarian multilingualism does not necessarily mean that the status of all languages in a region is equal, but simply that language shift does not typically occur despite extensive multilingualism; diglossia and language loyalty are documented in numerous Melanesian settings (e.g., Thurston, 1992; Clifton, 1994). The fact that knowledge of other languages has been noted as a means of gaining prestige in Melanesia (Salisbury, 1962; Sankoff, 1977) would contribute to preventing shift in diglossic situations. Although significantly disrupted in much of Melanesia today (cf. Kulick, 1992), egalitarian multilingualism is assumed to have once been widely present across the area and provides a social mechanism for understanding Melanesia’s diverse linguistic situation: language loyalty would have presented an obstacle to shifting to another language, while entrenched multilingualism provided the mechanism for convergence of linguistic structures.

At the same time, Melanesian egalitarian multilingualism has been seen to drive diversification in languages, i.e., the proliferation of language numbers. Laycock (1982) observes the inverse correlation between group size and language diversity in Melanesia:

It would seem *a priori* plausible to attribute the linguistic diversity of Melanesia to a combination of the factors of isolation, terrain, and time – a result of the languages of small communities being cut off from their neighbours for thousands of years. But such a simple explanation does not account for the fact that the largest languages – of the Papuan groups at least – are found in the most isolated areas [. . .], whereas the greatest diversity is found in areas of easy mobility and extensive trading contacts.

This distribution is argued to be the result of small speech communities in frequent contact with one another co-opting linguistic differences to mark their identity as distinct from their neighbours. Laycock (1982) describes how Uisai, a dialect of the Buin language of Bougainville Island, has flipped gender assignment of its nouns such that it is the reverse of that in other dialects – what is masculine in other dialects is feminine and vice versa. Laycock (1982, p. 35) writes ‘the only plausible hypothesis would appear to be that [at] some stage in the past an influential Uisai speaker innovated a linguistic change to differentiate his linguistic community from the rest of the Buins.’ Thurston (1987, 1994) argued that on the island of New Britain, off the New Guinea coast, complex language structures, such as difficult grammatical rules or irregular word formations, had been built up for the purpose of group-differentiation in small ‘esoterogenic’ language groups. While it is impossible to tell how Laycock’s and Thurston’s explanations correspond to prehistorical reality, we do have modern-day reports of Melanesian groups purposefully changing their language to foster a distinct identity. Kulick (1992, pp. 2–3), for example, describes how people of Indu village met and decided to distinguish themselves from other Selepet-speaking villages by adopting a new word for ‘no,’ which they have used ever since. This is only one small tweak to a language (but perhaps more significant than would appear at first glance given that words for ‘no’ are frequently used as ethnonyms in New Guinea), yet over time the accumulation of such changes might plausibly give rise to the proliferation of languages we see in Melanesia today.

Melanesia has been characterized as an area in which there is convergence in structural features of language but divergence in lexicon. Laycock (1982), for example, writes that because many adjacent languages in Melanesia have very similar morphosyntactic structures, learning another language is largely a matter of learning new lexicon – ‘that is, the two languages involved are very close to being the same language with a different set of labels.’ This situation itself almost certainly arises out of language contact: long-term multilingualism between speech groups over time leads to similar structural features being used in both languages, while the emblematic function of words, as the most salient markers of identity for speech groups, acts as a brake on adoption of foreign word-forms and even prompts lexical replacement through spontaneous innovation of distinctive word-forms (see also Brooks, 2019). The best-known study showing such a situation in Melanesia is Ross’ (1996, 2001) description of structural convergence between two languages, Takia (Austronesian) and Waskia (Papuan). Called ‘metatypy’ by Ross, the process involves strictly re-mapping of the meaning and distribution of inherited material, but with no transfer of word-forms or morphemes (see also François, 2009).

It is uncertain whether a prohibition on borrowing lexemes, which Ross argues to be central to this specific convergence behaviour, is applicable to the wider convergence area of Melanesia. Bradshaw (1978) appears to show that Numbani has both significant quantities of structural calquing and lexical borrowing from Yabem. What is clear is that lexical borrowing is by no means widely prohibited in Melanesia. Comrie (1989, 2000) documents a case of rapid lexical shift in the Haruai language, showing that recent contact with the only distantly related Kobon language of the Madang family has led to a massive influx of Kobon lexical items. The result of this extensive lexical borrowing is the masking of the original genealogical affiliation of Haruai. Only similarities in the morphology and pronoun systems reveal Haruai’s actual relationship to its other neighbour, Aramo of the Piawi family. A deeper-time case of lexical convergence is that of Apali, whose two dialects, Aki and Aci, have regular sound correspondences between apparently cognate vocabulary, but little to no cognates or correspondences in morphology or grammatical items. Wade (1993) observes grammatical correspondences of Aki and Aci with neighbouring languages and argues that the dialects, in fact, started out as separate languages and through rampant lexical borrowing have converged so much as to have obscured their genealogy almost entirely. ‘Unnatural’ levels of vocabulary replacement have widely reported in Melanesia, with a combination of lax attitudes in relation to the norms of language use and systems of word-tabooing thought to underpin this kind of convergence (e.g., Simons, 1982; Holzkecht, 1988).

The case studies outlined here have focused on the different forces of, often radical, convergence processes that are likely to have been at play in forming Linguistic Melanesia. There is evidence for language contact with almost every kind of outcome in Linguistic Melanesia (Foley, 2010). So, for example, whilst language loyalty has seen to be an important feature of the Melanesian Linguistic Area, numerous instances of language shift are also documented. Our picture of contact in Melanesia is still limited to a small number of case studies relative to the very large number of languages in the area, but these provide important context for understanding how the complex isoglosses defining zones of structural similarity may have come into being in Linguistic Melanesia.

4. Current and future directions of research

Given the huge number of languages spoken within Linguistic Melanesia, much of the effort of linguists in the area has been focused on the documentation and description of individual

languages or small groups of languages. Great strides have been made in the last two decades and following on from these there is the potential for advances to be made in understanding Linguistic Melanesia and the prehistorical events that underpinned its formation.

The genealogical affiliations of the Papuan languages are one of the unsolved issues of present-day linguistics. A lack of reliable descriptive materials for the many languages of the region plus frequent scepticism on the part of linguists working in New Guinea about the possibility of establishing language families, at least large ones, has resulted in relatively few attempts to apply historical linguistic methods to Papuan language data to-date. The cases of lexical divergence and rapid lexical replacement in Linguistic Melanesia discussed in the previous section, in particular, gave rise to a trope in the literature that the area is one of parts of the world which present potentially serious problems for the comparative method (Thomason, 1999). Recent successes in demonstrating smaller Papuan language groupings (e.g., Timor-Alor-Pantar family, Schapper, Huber and van Engelenhoven, 2014; Anim family, Usher and Suter, 2015) and in reconstructing the in-depth history of others (e.g., Awyu-Dumut family, Wester, 2014; Binandere family, Smallhorn, 2011; Sogeram family, Daniels, 2015) show, however, that the discovery of relationships between Papuan languages is far from intractable, where sufficient material is collated and carefully analyzed (Greenhill, 2015).

Advances in our understanding of the prehistory of Melanesia using linguistic data require careful sorting of descent from diffusion. Early comparative-historical work (see the papers in Wurm, 1975) particularly suffered from the conflation of typological features and lexical resemblances in defining Papuan language families (cf. Pawley's, 2005 review of these methods). Typological features such as body-tally systems for counting and switch reference marking on medial verbs are still used alongside lexical and pronominal lookalikes as diagnostic of membership of the Trans-New Guinea macro-family (e.g., Pawley and Hammarström, 2017). The increasing availability of high-quality data on languages in the region has revealed neither body-tally systems nor switch reference is limited to presumed TNG languages; in fact, both features appear in languages north and south of the central cordillera of New Guinea where TNG languages are dominant, and neither system appears to be reconstructable across TNG families. Overall very little is known about the different historical states of Papuan language families, and, as a result, it is difficult to ascertain with any level of confidence where language convergence has occurred between Papuan languages. In order to formulate better hypotheses about the contact that underlies the formation of Linguistic Melanesia, we must have a clearer picture of which linguistic features reconstruct to which families.

A puzzle that is increasingly being addressed through interdisciplinary scholarship in Melanesia is that presented by the different distributions and higher-level affiliations of Papuan languages spoken offshore to the west and the east of New Guinea respectively. Papuan languages to the east of New Guinea occupy small isolated pockets over numerous far-flung islands and are from around a dozen families each with a handful of languages (Stebbins, Evans and Terrill, 2017). This contrasts with the west where there are only two Papuan families; these are considerably larger in size (25 languages in the Timor-Alor-Pantar family and 15 languages in the North Halmahera family) and are each contained within relatively compact regions. In addition, the Papuan families to the west of New Guinea both have credible proposals positing that they originate on the New Guinea mainland and have relatives there today: the Timor-Alor-Pantar family is thought to be related to the Papuan languages of West Bomberai (Hull, 2004) and thus part of the wider TNG macro-family (Ross, 2005), while the North Halmahera family has been linked to the West Bird's Head languages (Voorhoeve, 1988) and a possible West Papuan macro-family (Donohue, 2008). To the east of New Guinea no such relationships between the Papuan languages of the islands and those of the New Guinea mainland

have been suggested. These asymmetries point to significant differences in the prehistories of the western and eastern maritime halves of Linguistic Melanesia that are still in need of unravelling. The pre-Austronesian period to the east of New Guinea in Island Melanesia appears to have been characterized by relative stasis, with economically simple people continuing a broadly similar subsistence lifestyle of hunting and foraging for many thousands of years (Spriggs, 1997, p. 43ff). To the west of New Guinea there is evidence of a dynamic, maritime culture with extensive inter-island trade networks established before the Austronesian arrival (Schapper, 2015; Shipton et al., 2019). It seems probable that these networks could have facilitated the spread of Papuan languages from coastal areas of New Guinea to the islands, thus giving us the two Papuan families linked to the mainland that we have today. However, it is crucial that the claims of historical relatedness between the outlier Papuan families of western Melanesia and those on the mainland be demonstrated by means of rigorous historical linguistic methodologies.

Another problem in asymmetry has been to explain the unequal extent of Linguistic Melanesia either side of New Guinea: whilst in the west Linguistic Melanesia correlates roughly with the extent of Papuan languages, in the east the ‘effects’ of Melanesian influence extend far beyond extant Papuan languages, reaching into Vanuatu, New Caledonia, and on occasion also Fiji, well beyond the Solomon Islands where the eastern-most Papuan languages are found. We noted some of these Melanesian features appearing in this area in Section 2. In Vanuatu, linguists combined the observation of Melanesian linguistic features with the fact that the people of Vanuatu have Melanesian phenotypes to infer that, although originally settled by Austronesian speakers of Asian descent, Papuans speakers who had adopted Austronesian seafaring culture arrived in Vanuatu shortly after the initial settlement. These new arrivals were then inferred to have shifted to speaking the Austronesian languages of their predecessors, but brought over many features of their original Papuan languages (Blust, 2005, 2008; Donohue and Denham, 2008). This prehistorical scenario has recently been borne out by studies in ancient DNA which have found precisely this genetic layering in dated human remains from Vanuatu (Posth et al., 2018; Lipson et al., 2018). Similar scenarios are likely to account for the appearance of Melanesian linguistic features in New Caledonia and Fiji, but more work to pinpoint the Papuan features is needed in these places, as in Vanuatu, to clarify the varied nature of the Papuan substrates in the languages.

In the previous section, the reader will have noticed that most studies of contact in Linguistic Melanesia focus on Austronesian languages, either in contact with one another or in contact with particular Papuan languages. This is because we have a reasonable picture of the structural states of the Austronesian family at several points in its history, including outside of Linguistic Melanesia. Significantly, however, there is a major gap in our understanding of Austronesian family when it enters the westernmost part of Linguistic Melanesia. In the time between Proto-Malayo-Polynesian and Proto-Oceanic, the two clearest and most well-understood major nodes of the Austronesian family tree outside of the Formosan homeland, speakers of pre-existing, presumably Papuan, languages were encountered and their influence is thought to be responsible, at least in part, for the significant differences in these two reconstructed languages (Blust, 1993; Kamholz, 2014). Blust (1993) proposes that the Austronesian languages of Linguistic Melanesia all belong to a single subgroup, Central Eastern Malayo-Polynesian (CEMP), which takes in the Central Malayo-Polynesian (CMP), South Halmahera-West New Guinea (SHWNG), and Oceanic subgroups, the latter two grouping together into a proposed Eastern Malayo-Polynesian subgroup. The lack of in-depth historical reconstruction of PCEMP and its daughters, apart from POc, means that it is difficult to make reliable inferences about prehistorical Papuan-Austronesian contact situations. A case in point is the

appearance of bound person-number markers in CEMP languages. Blust (1993, pp. 258–259) tentatively reconstructs a paradigm of bound person number markers for PCEMP, but Donohue and Grimes (2008, pp. 131–132) cast doubt on whether the different forms can convincingly be treated as a single innovation such as would characterize a single ancestral protolanguage. Without a reconstruction of PCEMP, if indeed such a language existed, and its constituent subgroups, we do not know whether there was a single event in which a substrate/contact language introduced the person-number markers to PCEMP, or whether they were innovated on numerous separate occasions independently, potentially on the basis of distinct substrate/contact events. Understanding the subgroups of western Linguistic Melanesia in the future will illuminate what was likely one of the first points of contact between speakers of Austronesian and Papuan languages, when Austronesian speakers moved southwards out of the Philippines and into the area of the Bird’s Head around 3,000BP (Bellwood, 2017).

5. Further reading

Palmer, B., ed. (2017). *The languages and linguistics of the New Guinea area*. Berlin: De Gruyter Mouton.

This edited volume presents an up-to-date picture of our knowledge on the Papuan languages of Linguistic Melanesia. The main focus of these chapters is the synchronic description of the chief structural characteristics of different regions.

Foley, W.A. (2010). Language contact in the New Guinea region. In: R. Hickey, ed., *The handbook of language contact*, 1st ed. Chichester: Wiley-Blackwell, pp. 795–813.

This chapter provides the reader with an overview of described cases of contact in Linguistic Melanesia. It makes clear that examples of diffusion of essentially all aspects of linguistic structure can be found in the area.

Ross, M. (1996). Contact-induced change and the comparative method: Cases from Papua New Guinea.

In: M. Durie and M. Ross, eds., *The comparative method revisited: Irregularity and regularity in language change*, 1st ed. New York: Oxford University Press, pp. 180–217.

This chapter establishes the term ‘metatypy’ in the contact linguistics literature for situations in which radical restructuring of syntax and lexical semantics takes place in one language on the model of another.

Ross, M. (2017). Languages of the New Guinea region. In: R. Hickey, ed., *The Cambridge handbook of areal linguistics*, 1st ed. Cambridge: Cambridge University Press, pp. 758–820.

This chapter offers a detailed look at the distribution of some grammatical features not covered in the present contribution over a portion of Linguistic Melanesia and presents a different view on what constitutes areality.

6. Related topics

Typological factors, social factors

Abbreviations

| | |
|-----------------|--|
| CEMP | Central Eastern Malayo-Polynesian |
| CMP | Central Malayo-Polynesian |
| O | object |
| PA _N | Proto-Austronesian |
| PCEMP | Proto-Central Eastern Malayo-Polynesia |
| POc | Proto-Oceanic |
| S | subject |

SHWNG South Halmahera-West New Guinea
TNG Trans-New Guinea
V verb

References

- Aikhenvald, A. and Stebbins, T.N. (2007). Languages of New Guinea. In: O. Miyaoka, O. Sakiyama and M.E. Krauss, eds., *Vanishing languages of the pacific rim*. Oxford: Oxford University Press, pp. 239–266.
- Anderson, G.D.S. (2013). The velar nasal. In: M.S. Dryer and M. Haspelmath, eds., *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.
- Arnold, L. (2018). A preliminary archaeology of tone in Raja Ampat. *NUSA: Linguistic Studies of Languages in and Around Indonesia*, 64, pp. 7–37.
- Baerman, M. (2014). Suppletive kin term paradigms in the languages of New Guinea. *Linguistic Typology*, 18, pp. 413–448.
- Bellwood, P. (1998). From bird's head to bird's eye view: Long-term structures and trends in Indo-Pacific prehistory. In: J. Miedema, C. Odé and R.A.C. Dam, eds., *Perspectives on the bird's head of Irian Jaya, Indonesia proceedings of the conference*. Leiden: Rodopi, pp. 951–975.
- Bellwood, P. (2017). *First Islanders: Prehistory and human migration in Island Southeast Asia*. Hoboken, NJ: John Wiley and Sons.
- Blust, R. (1993). Central and central-Eastern Malayo-Polynesian. *Oceanic Linguistics*, 32, pp. 241–293.
- Blust, R. (2005). Review of Lynch, Ross, and Crowley, the Oceanic languages. *Oceanic Linguistics*, 44, pp. 44–58.
- Blust, R. (2008). Remote Melanesia: One history or two? An addendum to Donohue and Denham. *Oceanic Linguistics*, 47, pp. 445–459.
- Blust, R. (2009). *The Austronesian languages*. Canberra: Pacific Linguistics.
- Bradshaw, J. (1978). Multilingualism and language mixture among the Numbami. *Kivung*, 11, pp. 26–49.
- Bradshaw, J. (1982). *Word order change in Papua New Guinea Austronesian languages*. PhD. University of Hawaii.
- Brooks, J. (2019). Secret language and resistance to borrowing in Chini. *International Journal of Language and Culture*, 6(1), pp. 83–94.
- Clifton, J. (1994). Stable multilingualism in a small language group: The case of Kaki Ae. *Language and Linguistics in Melanesia*, 25, pp. 107–124.
- Codrington, R.H. (1885). *The Melanesian languages*. Oxford: Clarendon Press.
- Comrie, B. (1989). Genetic classification, contact, and variation. In: *Georgetown University round table on languages and linguistics 1989*. Washington, DC: Georgetown University Press, pp. 81–93.
- Comrie, B. (2000). Language contact, lexical borrowing, and semantic fields. *Studies in Slavic and General Linguistics*, 28, pp. 73–86.
- Daniels, D. (2015). *A reconstruction of proto-sogeram: Phonology, lexicon, and morphosyntax*. PhD. University of California.
- Delpada, B. (2016). *Abui phonology*. MA thesis. Nanyang Technological University.
- Donohue, M. (1997). Tone in New Guinea languages. *Linguistic Typology*, 1, pp. 347–386.
- Donohue, M. (2005a). Tone and the trans New Guinea languages. In: S. Kaji, ed., *Crosslinguistic studies of Tonal phenomena*, 1st ed. Tokyo: ILCAA, pp. 33–53.
- Donohue, M. (2005b). Word order in New Guinea: Dispelling a myth. *Oceanic Linguistics*, 44, pp. 527–536.
- Donohue, M. (2007). Word order in Austronesian: From North to South and West to East. *Linguistic Typology*, 11, pp. 351–393.
- Donohue, M. (2008). Bound pronominals in the West Papuan languages. In: C. Bower, B. Evans and L. Miceli, eds., *Morphology and language history: In honour of Harold Koch*, 1st ed. Amsterdam: John Benjamins, pp. 43–58.

- Donohue, M. and Denham, T. (2008). The language of Lapita: Vanuatu and an early Papuan presence in the Pacific. *Oceanic Linguistics*, 47, pp. 365–376.
- Donohue, M. and Grimes, C.E. (2008). Yet more on the position of the languages of Eastern Indonesia and East Timor. *Oceanic Linguistics*, 47(1), pp. 114–158.
- Donohue, M., Hetherington, R., McElvenny, J. and Dawson, V. (2013). *World phonotactics database*. Department of Linguistics, The Australian National University. Available at: <http://phonotactics.anu.edu.au>.
- Donohue, M. and Whiting, B. (2011). Quantifying areality: A study of prenasalisation in Southeast Asia and New Guinea. *Linguistic Typology*, 15, pp. 101–121.
- Dryer, M.S. (2013a). Order of subject, object and verb. In: M.S. Dryer and M. Haspelmath, eds., *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.
- Dryer, M.S. (2013b). Order of genitive and noun. In: M.S. Dryer and M. Haspelmath, eds., *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.
- Dryer, M.S. (2013c). Order of numeral and noun. In: M.S. Dryer and M. Haspelmath, eds., *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.
- Dumont D'Urville, J.S.C. (1832). *Voyage de la Corvette l'Astrolabe: Exécuté par Ordre du Roi, Pendant les Années 1826-1827-1828-1829*. Paris: J. Tastu.
- Dunn, M., Levinson, S.C., Lindström, E., Reesink, G. and Terrill, A. (2008). Structural phylogeny in historical linguistics: Methodological explorations applied in Island Melanesia. *Language*, 84, pp. 710–759.
- Evans, N. (forthcoming). Linguistic divergence under contact. In: Michela Cennamo and Claudia Fabrizio, eds., *Historical linguistics 2015: Selected papers from the 22nd international conference on historical linguistics*. Amsterdam: John Benjamins, pp. 564–591.
- Foley, W.A. (1986). *The Papuan languages of New Guinea*. Cambridge: Cambridge University Press.
- Foley, W.A. (2000). The languages of New Guinea. *Annual Review of Anthropology*, 29, pp. 357–404.
- Foley, W.A. (2010). Language contact in the New Guinea region. In: R. Hickey, ed., *The handbook of language contact*. Malden, MA: Wiley-Blackwell, pp. 795–813.
- François, A. (2009). The languages of Vanikoro: Three lexicons and one grammar. In: B. Evans, ed., *Discovering history through language: Papers in honour of Malcolm Ross*. Canberra: Pacific Linguistics, pp. 103–126.
- François, A. (2012). The dynamics of linguistic diversity. Egalitarian multilingualism and power imbalance among Northern Vanuatu languages. *International Journal of the Sociology of Language*, 214, pp. 85–110.
- Gabelentz, H.G.C. von der. (1860). *Die melanesischen Sprachen nach ihrem grammatischen Bau und ihrer Verwandtschaft unter sich und mit den Malaisisch-Polynesischen Sprachen untersucht*. Leipzig: S. Hirzel.
- Gil, D. (2015). The Mekong-Mamberamo linguistic area. In: N.J. Enfield and B. Comrie, eds., *Languages of mainland Southeast Asia, the state of the art*. Berlin: Pacific Linguistics, De Gruyter Mouton, pp. 266–355.
- Green, R.C. (1991). Near and Remote Oceania – Disestablishing ‘Melanesia’ in culture history. In: A. Pawley, ed., *Man and a half: Essays in pacific anthropology and ethnobiology in honour of Ralph Bulmer*. Auckland: Polynesian Society, pp. 491–502.
- Greenhill, S.J. (2015). Trans New Guinea.org: An online database of New Guinea languages. *PLoS One*, 10(10), e0141563. Available at: <https://doi.org/10.1371/journal.pone.0141563>.
- Hammarström, H. (2010). Rarities in numeral systems. In: J. Wohlgemuth and M. Cysouw, eds., *Rethinking universals: How rarities affect linguistic theory*. Berlin and New York: De Gruyter Mouton, pp. 11–60.
- Haudricourt, A-G. (1961). Richesse en phonèmes et richesse en locuteurs. *L'Homme*, 1, pp. 5–10.
- Haudricourt, A-G. (1968). La langue de Gomen et la langue de Touho en Nouvelle- Calédonie. *Bulletin de la Société de Linguistique de Paris*, 63(1), pp. 218–235.
- Hill, C., Soares, P., Mormina, M., Macaulay, V., Clarke, D., Blumbach, P.B., et al. (2007). A mitochondrial stratigraphy for Island Southeast Asia. *American Journal of Human Genetics*, 80, pp. 29–43.

- Holzkecht, S. (1988). Word taboo and its implications for language change in the Markham family of languages. *Language and Linguistics in Melanesia*, 18, pp. 43–69.
- Hull, G. (2004). The Papuan languages of Timor. *Estudos de Línguas e Culturas de Timor Leste/Studies in Languages and Cultures of East Timor*, 6, pp. 23–99.
- Kamholz, D. (2014). *Austronesians in Papua: Diversification and change in South Halmahera-West New Guinea*. PhD. University of California, Berkeley.
- Kamholz, D. (2017). Tone and language contact in southern Cenderawasih Bay. *NUSA: Linguistic Studies of Languages in and Around Indonesia*, 62, pp. 7–39.
- Kirby, J. and Brunelle, M. (2017). Southeast Asian tone in areal perspective. In: R. Hickey, ed., *The Cambridge handbook of areal linguistics*. Cambridge: Cambridge University Press, pp. 703–731.
- Klamer, M., Reesink, G. and van Staden, M. (2008). Eastern Indonesia as a linguistic area. In: P. Muysken, ed., *From linguistic areas to areal linguistics*, vol. 95. Amsterdam: John Benjamins, p. 149.
- Kulick, D. (1992). *Language shift and cultural reproduction: Socialization, self, and syncretism in a Papua New Guinea village*. Cambridge: Cambridge University Press.
- Laycock, D. (1982). Linguistic diversity in Melanesia: A tentative explanation. In: R. Carle, M. Heinschke, P. Pink, et al., eds., *Gava': Studies in Austronesian languages and cultures dedicated to Hans Kähler*. Berlin: Reimer, pp. 31–37.
- Lean, G. (1992). *Counting systems of Papua New Guinea and Oceania*. PhD. University of Technology, Papua New Guinea.
- Lipson, M., Skoglund, P., Spriggs, M., et al. (2018). Population turnover in Remote Oceania shortly after initial settlement. *Current Biology*, 28, pp. 1157–1165.
- Macaulay, V. et al. (2005). Single, rapid coastal settlement of Asia revealed by analysis of complete mitochondrial genomes. *Science*, 308, pp. 1034–1036.
- Maddieson, I. (2013). Tone. In: M.S. Dryer and M. Haspelmath, eds., *The world atlas of language structures online*. Leipzig: Max Planck Institute for Evolutionary Anthropology.
- Nichols, J. (1997). Sprung from two common sources: Sahul as a linguistic area. In: P. McConvell and N. Evans, eds., *Archaeology and linguistics: Aboriginal Australia in global perspective*. Melbourne: Oxford University Press, pp. 135–168.
- O'Connell, J. (2012). The restaurant at the end of the universe: Modelling the colonization of Sahul. *Australian Archaeology*, 74, pp. 5–17.
- Pawley, A. (2005). Trans New Guinea hypothesis: Recent research and its implications. In: A. Pawley, R. Attenborough, J. Golson and R. Hide, eds., *Papuan pasts: Cultural, linguistic and biological histories of Papuan-speaking peoples*. Canberra: Pacific Linguistics, pp. 67–107.
- Pawley, A. and Hammarström, H. (2017). The Trans New Guinea family. In: B. Palmer, ed., *The languages and linguistics of the New Guinea area: A comprehensive guide*. Berlin: De Gruyter Mouton, pp. 21–195.
- Posth, C., Nägele, K., Collieran, H., et al. (2018). Language continuity despite population replacement in Remote Oceania. *Nature Ecology & Evolution*, 2, pp. 731–740.
- Ray, S.H. (1907). The Melanesian languages of British New Guinea. In: S.H. Ray, ed., *Linguistics*, 1st ed. Cambridge: Cambridge University Press, pp. 413–503.
- Reesink, G. (2002). Clause-final negation: Structure and interpretation. *Functions of Language*, 9, pp. 239–268.
- Reesink, G. (2005). Sulka of East New Britain: A mixture of Oceanic and Papuan traits. *Oceanic Linguistics*, 44, pp. 145–193.
- Reesink, G., Singer, R. and Dunn, M. (2009). Explaining the linguistic diversity of Sahul using population models. *PLoS Biol*, 7(11), e1000241.
- Rivierre, J-C. (1993). Tonogenesis in New Caledonia. In: J.A. Edmondson and K.J. Gregerson, eds., *Tonality in Austronesian languages*. Honolulu: University of Hawaii Press, pp. 155–173.
- Ross, M. (1993). Tonogenesis in the North Huon Gulf Chain. In: J.A. Edmondson and K.J. Gregerson, eds., *Tonality in Austronesian languages*. Honolulu: University of Hawaii Press, pp. 133–153.
- Ross, M. (1996). Contact-induced change and the comparative method: Cases from Papua New Guinea. In: M. Durie and M. Ross, eds., *The comparative method revisited: Irregularity and regularity in language change*. New York: Oxford University Press, pp. 180–217.

- Ross, M. (2001). Contact-induced change in Oceanic languages in North-West Melanesia. In: A.Y. Aikenvald and R.M.W. Dixon, eds., *Areal diffusion and genetic inheritance: Problems in comparative linguistics*. Oxford: Oxford University Press, pp. 134–166.
- Ross, M. (2005). Pronouns as preliminary diagnostic for grouping Papuan languages. In: A.K. Pawley, R. Attenborough, J. Golson, and R. Hide, eds., *Papuan pasts: Cultural, linguistic and biological histories of Papuan speaking peoples*. Canberra: Pacific Linguistics, pp. 15–65.
- Ross, M. (2017). Languages of the New Guinea region. In: R. Hickey, ed., *The Cambridge handbook of areal linguistics*. Cambridge: Cambridge University Press, pp. 758–820.
- Salisbury, R. (1962). Notes on bilingualism and language change in New Guinea. *Anthropological Linguistics*, 4, pp. 77–109.
- Sankoff, G. (1976). Political power and linguistic inequality in Papua New Guinea. In: W.M. O’Barr and J.F. O’Barr, eds., *Language and politics*, 1st ed. The Hague: Mouton, pp. 283–310.
- Sankoff, G. (1977). Multilingualism in Papua New Guinea. In: S.A. Wurm, ed., *New Guinea area languages and language study: Language, culture, society and the modern world*. Canberra: Pacific Linguistics, pp. 265–307.
- San Roque, L. and Loughnane, R. (2012a). The New Guinea highlands evidentiality area. *Linguistic Typology*, 16, pp. 111–167.
- San Roque, L. and Loughnane, R. (2012b). Inheritance, contact and change in the New Guinea Highlands evidentiality area. *Language and Linguistics in Melanesia: Special Issue Part II*, pp. 397–427.
- Schapper, A. (2015). Wallacea, a linguistic area. *Archipel*, 90, pp. 99–151.
- Schapper, A. (2017). Substratum inference through lexical patterns: A case study from Sunda-Sahul. In: *Workshop on language shift and substratum interference in (pre)history, MPI-SHH*. Jena.
- Schapper, A. and Hammarström, H. (2013). Innovative complex numerals and numeral bases in Austronesian languages of Eastern Indonesia and East Timor. *Oceanic Linguistics*, 52, pp. 425–456.
- Schapper, A., Huber, J. and van Engelenhoven, A. (2014). The relatedness of Timor- Kisar and Alor-Pantar languages: A preliminary demonstration. In: M. Klamer, ed., *Alor-Pantar languages: History and typology*, 1st ed. Berlin: Language Science Press, pp. 99–154.
- Schapper, A., San Roque, L. and Hendery, R. (2016). Tree, firewood and fire in the languages of Sahul. In: M. Koptjevskaja-Tamm and P. Juvonen, eds., *Lexico- typological approaches to semantic shifts and motivation patterns in the Lexicon*, 1st ed. Berlin: De Gruyter Mouton, pp. 355–422.
- Schapper, A. and de Vries, L. (2018). Comparatives in Melanesia: Concentric circles of convergence. *Linguistic Typology*, 22, pp. 437–494.
- Shipton, C., O’Connor, S., Reepmeyer, C. and Kealy, S. (2019). ShellAdzes, exotic obsidian, and inter-Island voyaging in the terminal Pleistocene and Holocene of Wallacea. *Journal of Island and Coastal Archaeology*. doi:10.1080/15564894.2019.1581306.
- Simons, G. (1982). Word taboo and comparative Austronesian linguistics. In: A. Halim, L. Carrington and S.A. Wurm, eds., *Papers from the third international conference on Austronesian linguistics, Vol. 3: Accent on variety*. Canberra: Pacific Linguistics.
- Smallhorn, J. (2011). *The Binanderean languages of Papua New Guinea: Reconstruction and subgrouping*. Canberra: Pacific Linguistics.
- Spriggs, M. (1997). *The Island Melanesians*. Oxford: Blackwell.
- Stebbins, T., Evans, B. and Terrill, A. (2017). The Papuan languages of Island Melanesia. In: B. Palmer, ed., *The languages and linguistics of the New Guinea area: A comprehensive guide*. Berlin: De Gruyter Mouton, pp. 775–894.
- Tabbada, K.A., Trejaut, J., Loo, J.H., Chen, Y.M., Lin, M., Mirazón-Lahr, M., et al. (2010). Philippine mitochondrial DNA diversity: A populated viaduct between Taiwan and Indonesia? *Molecular Biology and Evolution*, 27, pp. 21–31.
- Thomason, S.G. (1999). Speakers’ choices in language change. *Studies in the Linguistic Sciences*, 29(2), pp. 19–43.
- Thurston, W.R. (1987). *Processes of change in the languages of North-Western New Britain*. Canberra: Pacific Linguistics.

- Thurston, W.R. (1992). Sociolinguistic typology and other factors effecting change in northwestern New Britain, Papua New Guinea. In: T. Dutton, ed., *Culture change, language change: Case studies from Melanesia*. Pacific Linguistics Canberra: Pacific Linguistics, pp. 123–139.
- Thurston, W.R. (1994). Renovation and innovation in the languages of north-western New Britain. In: T. Dutton and D. Tryon, eds., *Language contact and change in the Austronesian world*, 1st ed. Berlin: De Gruyter Mouton, pp. 573–609.
- Tumonggor, M.K., Karafet, T.M., Hallmark, B., Lansing, J.S., Sudoyo, H., Hammer, M.F., et al. (2013). The Indonesian archipelago: An ancient genetic highway linking Asia and the Pacific. *Journal of Human Genetics*, 58, pp. 165–173.
- Usher, T. and Suter, E. (2015). The Anim languages of Southern New Guinea. *Oceanic Linguistics*, 54(1), pp. 110–142.
- Voorhoeve, C.L. (1988). The languages of the North Halmahera Stock. In: *Papers in New Guinea linguistics A-76*. Canberra: Research School of Pacific and Asian Studies, Australian National University, pp. 181–209.
- Voorhoeve, C.L. (2004). Contact-induced change in the non-Austronesian languages in the north Moluccas, Indonesia. In: T. Dutton and D.T. Tryon, eds., *Language contact and change in the Austronesian world*, 1st ed. Berlin: De Gruyter Mouton, pp. 649–674.
- Wade, Ma. (1993). Language convergence or divergence: The case of the Apalī (Emerum) language. *Language and Linguistics in Melanesia*, 24, pp. 73–93.
- Wester, R. (2014). *A linguistic history of Awyu-Dumut: Morphological study and reconstruction of a Papuan language family*. PhD. Vrije Universiteit Amsterdam.
- Wurm, S.A., ed. (1975). *New Guinea area languages and language study. Vol 1: Papuan languages and the New Guinea linguistic scene*. Canberra: Research School of Pacific and Asian Studies, Australian National University.

Acknowledgements

Many thanks to the editors for inviting me to be part of the volume and their thoughtful feedback on this chapter. Rachel Hendery also gave me useful comments that improved the clarity of the writing – thank you! Research funding is gratefully acknowledged from the Netherlands Organisation for Scientific Research VENI project ‘The evolution of the lexicon. Explorations in lexical stability, semantic shift and borrowing in a Papuan language family,’ the Volkswagen Stiftung DoBeS project ‘Aru languages documentation,’ and the Australian Research Council project (ARC, DP180100893) ‘Waves of words.’ All errors are my own.

Language contact in North America

Marianne Mithun

1. Introduction

The languages indigenous to North America north of Mexico have much to tell us about the potential effects of language contact. They are numerous: at least 275 are known to have been spoken at the time of contact with Europeans, and there were undoubtedly more. They show considerable genealogical diversity, constituting 55–60 distinct families and isolates. They are typologically diverse, but certain characteristics are pervasive, in particular often complex morphological structures. Many have been shaped by contact with their neighbours. Some similarities are shared across pairs of adjacent languages, and many across larger groups.

A number of contact effects in North America differ from those in some other parts of the world in an interesting way. While it is often assumed that vocabulary is the first feature to be affected in contact situations, many neighbouring languages show surprisingly few shared words, but extensive structural parallelisms. In these areas speakers often put special efforts into keeping their languages distinct, working to avoid transferring what they are most conscious of: words. At the same time, they may transfer other less obvious features, such as propensities to specify certain distinctions or patterns of expression. Over time, the most frequent can become routinized and ultimately crystallized in grammar. Of course, the kinds of features copied in contact situations reflect complex social, cultural, and historical circumstances. One of the most important, described in detail by Thomason and Kaufman (1988), is whether features are adopted by speakers *into* their first language (with language maintenance), or carried into another language *from* their first language (with language shift). Another is attitudes toward code-switching. Still others are how long the languages have been in contact and how intense the contact has been. And with long-standing contact, language structures may gradually converge, further facilitating transfer.

North America is also home to some contact languages. Some developed entirely among indigenous groups, and some were used between indigenous groups and Europeans. Some remained mainly trade jargons, some became elaborate pidgins, and some originated as mixed languages in communities of skilled bilingual speakers.

The investigation of language contact in North America has the potential to deepen our understanding of the powerful ways in which contact can shape the development of languages, but it also presents certain challenges. We lack deep historical and philological records comparable to those for some European and Asian peoples, so it is not possible to trace developments through time in the same ways. And, sadly, many of the languages are no longer spoken, so we are often working with closed corpora.

2. Historical overview

In part because of the wealth of languages indigenous to North America, particularly during the first half of the twentieth century efforts were directed at bringing order to apparent chaos by finding deeper genealogical relationships. Edward Sapir (1921a, 1929) proposed an ambitious classification in which all of the families and isolates of North America were grouped into just six superstocks. The proposals raised important questions about what kinds of resemblances between languages might constitute evidence for genealogical relationship, and what might be the result of language contact.

Early on Franz Boas maintained that purely grammatical similarities, particularly morphological structures, could not be spread by contact, so they must constitute evidence of common inheritance. Discussing the relationship between Tlingit and Haida in the Northwest, he maintained that though lexical similarities are few, structural similarities are sufficient to link them.

The structural resemblance of the two languages and their contrast with the neighboring languages can be explained only by the assumption of a common origin. The number of words which may be connected by etymology is small, and the similarities are doubtful. Nevertheless, the structural resemblance must be considered final proof of a historical connection between the two.

(Boas, [1894] 1974, p. 162)

Sapir put forth the same argument when he assembled his six superstocks.

So long as such direct historical testimony as we have gives us no really convincing examples of profound morphological influence by diffusion, we shall do well not to put too much reliance in diffusion theories.

(Sapir, 1921b, p. 206)

But the possibility of contact effects on structure was soon recognized. Boas described his change of heart.

As early as 1893 I pointed out that the study of the grammar of American languages has demonstrated the occurrence of a number of striking morphological similarities between neighboring stocks which, however, are not accompanied by appreciable similarities in vocabulary. At that time I was inclined to consider these similarities as proof of relationship of the same order as that of languages belonging, for instance, to the Indo-European family. While further studies, particularly in California, have shown that we may generalize the observations which I made based on the languages of the North Pacific coast, I doubt whether the interpretation given at that time is tenable.

(Boas, 1920, p. 211)

Sapir similarly acknowledged the possibility of contact effects on structure.

It is well known to students of language that striking phonetic and morphologic similarities are not infrequently found between neighboring languages that, so far as can be ascertained, are in no way genetically related. Such resemblances, insofar as they are not merely fortuitous, must be due to the assimilatory influence asserted by one language over another.

(Sapir, 1916, cited in Darnell and Sherzer, 1971, p. 24)

Since that time, it has become increasingly clear that structure, even morphological structure, can indeed be transferred via contact. Tlingit and Haida are no longer considered genealogically related, and most of Sapir's superstocks, some covering wide swaths of California and the West, are no longer considered established. The classification generally accepted today is not far from that consisting of 55 families established by the late nineteenth century (Powell, 1891).

Some histories of work on contact within North America can be found in Darnell and Sherzer, 1971; Sherzer, 1973, 1976; Bright and Sherzer, 1976; Campbell and Mithun, 1979; Mithun, 1999, 2010, 2017a, to appear; Silverstein, 1996; Campbell, 1997; and Thomason, 2016.

3. Critical topics

A fundamental question is what aspects of language can be transferred in contact situations. It is often assumed that the first aspect of language to be copied is vocabulary. In North America, as elsewhere, there is indeed evidence that speakers have adopted words from neighbouring languages. Kari (1977) provides examples from two Athabaskan languages of Alaska, Tanaina and Ahtna. Tanaina consists of four varieties: three Lower Cook Inlet dialects, which are quite similar, and the more divergent Upper Inlet dialect. Upper Inlet Tanaina shares a number of traits with the neighbouring Ahtna. Kari concludes from patterns of trade, bilingualism, intermarriage, and various mutual cultural influences, that the two groups have been in intimate contact for centuries. The clan system of the Upper Inlet Tanaina more closely resembles that of the Ahtna than it does those of the other Tanaina groups.

Most of the Upper Inlet Tanaina have kin ties with the Ahtna. The pattern of intermarriage favored Ahtna resettlement in or adjacent to Upper Inlet Tanaina territory. Interestingly, Ahtna influence is also present in the kinship terminology of Upper Inlet Tanaina.

(Kari, 1977, p. 279)

The languages are not mutually intelligible, but apparently many Tanaina speakers could use Ahtna.

When the languages in question are related, spotting copied vocabulary can be tricky. Sometimes it is clear simply from the distribution of words. There are Upper Tanaina words that do not resemble their counterparts in the other Tanaina dialects, but match those in Ahtna, as in (1).

1. Some Athabaskan loanwords (Kari, 1977, p. 286)

| Lower Inlet Tanaina | Upper Inlet Tanaina | Ahtna |
|---------------------|-----------------------|-------------------|
| 'robin' | <i>kaɫnay</i> | <i>six</i> |
| 'magpie' | <i>tʃk'naqaɫs'ɣya</i> | <i>taqylbi</i> |
| | | <i>staqaɫbeey</i> |

| | | | |
|-----------------|-------------------|----------------|----------------|
| ‘buteo species’ | <i>q’uluq’yya</i> | <i>k’tsu</i> | <i>k’etsaa</i> |
| ‘goshawk’ | <i>gižakəy</i> | <i>k’əmbəg</i> | <i>k’enbeg</i> |

An understanding of the sound changes which have occurred in the histories of the two languages is useful. The original Proto-Athabaskan vowels have undergone some changes in Tanaina, but remained essentially unchanged in Ahtna. In Tanaina, *a > u, for example. In some Upper Tanaina words, the vowels match those of Ahtna rather than their Lower Tanaina cognates.

2. Athabaskan loanwords (Kari, 1977, p. 283)

| | | | | |
|-------------|---------------|--------------|---------------|-------|
| Lower Inlet | Tanaina | Upper Inlet | Tanaina | Ahtna |
| ‘beach’ | <i>tubbyy</i> | <i>tabay</i> | <i>tabaay</i> | |
| ‘also, too’ | <i>k’u</i> | <i>k’a</i> | <i>k’ah</i> | |

Such differences suggest that Upper Inlet Tanaina speakers borrowed these words from Ahtna. Other words show similar patterns involving other sound changes and stress patterns.

Sometimes borrowed words can shed light on interactions between the groups in contact. Kari notes (1977, p. 285), ‘Patterns of shared lexical innovations in specific semantic domains such as flora and fauna, trade goods, potlatching, or material culture will provide clues to the nature and frequency of intra-Athapaskan intercourse.’ He reports that of the 66 Tanaina and Ahtna names for species of birds he assembled, 25 are shared between Upper Inlet Tanaina and Ahtna but are not found elsewhere in Tanaina, like those in (1). Of 41 animal terms, 11 are common to just Upper Inlet Tanaina and Ahtna. Adoption of terms for flora and fauna can indicate that speakers of the borrowing language moved into territory occupied by speakers of the source language. There is also a probable Ahtna place name in Upper Inlet Tanaina territory: Mt. Kliskon from *tsələs qan* ‘arctic ground squirrel lodge.’ The Ahtna term for ‘arctic ground squirrel’ is *tseles*, while the Tanaina term is *qunša* (1977, p. 278). Kari points to a group of 34 verb stems used in Upper Inlet Tanaina and Ahtna but not elsewhere in Tanaina, including ‘make potlatch speech.’ Such terms can suggest that speakers of the borrowing language adopted cultural practices from their neighbours.

Similar patterns can be seen elsewhere in North America. Nater (1974, 1977, 1984, 1994) describes vocabulary borrowed into Nuxalk (Bella Coola, a Salish language on the British Columbia coast) from neighbouring Northern Wakashan and Athabaskan languages in the Northwest, much of it in earlier times.

Loether (1998) provides evidence of a central California language area. Mono (also called Monache), is a language of the Numic branch of the Uto-Aztecan family, with extensive contacts.

As was common for all the Indians of California, the Western Mono maintained a complex set of relations with neighbouring and more distant tribes, which included political, social, religious, economic, and kinship ties.

(Loether, 1998, p. 102)

Neighbours included speakers of Yokuts (Yokuts family) and Miwok (Utian family). Multilingualism was the norm, and loanwords were transferred in all directions. Many are terms for animals, birds, insects, and plants, but there are others as well, such as ‘boat’ and, significantly, ‘co-parents-in-law.’

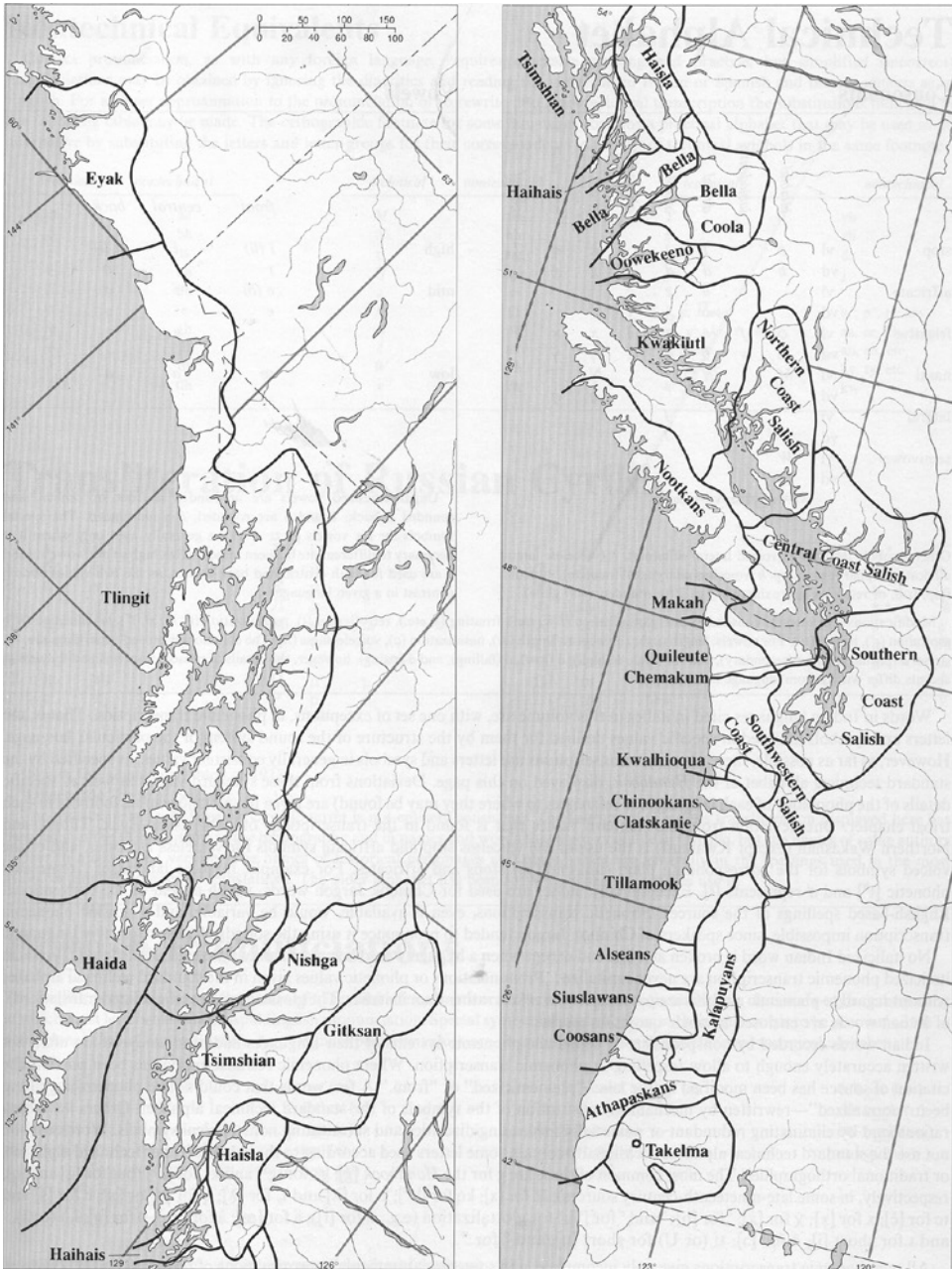


Figure 25.1 Northwest Coast

Source: Suttles, Wayne, ed. (1990). *Handbook of North American Indians: Northwest Coast*. Washington: Smithsonian Institution, 7: ix

Lexical copying can have further effects on a language. If competent bilinguals adopt enough words from another language, sounds in those words may be added to the phonological inventory as well. Callaghan (1964) describes the sounds of Lake Miwok, a language of the Utian family, spoken in the Clear Lake area of California. It has three series of stops not found in any other Utian languages (aspirated, ejective, and voiced), as well as /r/, /h/, and affricates /č, c', č', and λ'/. The Lake Miwok words containing these sounds do not have cognates in other Utian languages. Lake Miwok is spoken, however, in a recognized linguistic area, with Pomoan, Wappo, and Patwin neighbours, groups also not related to each other. The Lake Miwok consonant inventory closely matches those of the neighbours. Of 260 Lake Miwok words containing the unusual sounds, Callaghan identified 78 with identical or nearly identical forms in a nearby language but no other Miwok language. Lake Miwok *búkʰal* 'fish trap,' for example, resembles Southwestern Pomo *buhqʰal*'. Lake Miwok *c'a:yakami* 'salamander' resembles Wappo *č'ay:akamin*. Lake Miwok *c'akattu* 'blackbird' resembles Patwin *c'akatu*. Many terms denote birds ('kingfisher,' 'blackbird,' 'small blackbird,' 'yellowhammer,' 'mudhen,' 'buzzard,' 'snow bird,' 'crested bluejay,' 'small bluebird,' 'hummingbird'), animals ('mink,' 'salamander,' 'snake,' 'rattler's rattles,' 'snail,' 'small eels,' 'angleworm,' 'cricket,' 'large black ants'), or plants ('tan oak,' 'manzanita mushroom'), not surprising given that the Clear Lake area was already inhabited when the Lake Miwok moved into it. There are also place names for Dry Lake, an island in Clear Lake, and Mt. St. Helena. Some loans are more cultural, like 'fish trap,' 'sweet black acorn bread,' 'large storage basket,' 'large bowl-like basket,' 'cocoon rattle,' 'abalone shell ornament,' and 'spouse's sibling.' And some are more general, indicating more intimate contact ('heart,' 'throat,' 'breast,' 'finger,' 'to club,' 'to be bald,' 'to bump one's head,' 'to kiss,' 'to slap,' 'to puff, like smoke from a chimney').

Morphemes themselves have been transferred only in very rare instances. Kroskrity (1985) describes possessive constructions in Arizona Tewa, a Kiowa-Tanoan language. This dialect is spoken in a community within the Hopi reservation, which is in turn surrounded by the Navajo reservation. Arizona Tewa contains a possessive suffix *-bi* that is absent not only from all other Tewa dialects, but also all other Kiowa-Tanoan languages. Some examples are *sen-bí kʰaw* 'a man's song' (man-POSS song) and *na'in-bí cé* 'our dog' (we-POSS dog). The suffix matches a third person possessive prefix *bi-* in the unrelated Apachean languages of the Tlingit-Eyak-Athabaskan family, as in Navajo *hstiin bi-ye* 'the man's son' (man 3POSS-son) and *dzil bi-tsiin* 'the mountain's base' (mountain 3POSS-base) (Kroskrity, 1985, p. 488, citing Young and Morgan, 1980, pp. 73, 77). Kroskrity explains the shift from prefix to suffix in terms of the full construction; in both Arizona Tewa and Navajo the possessive affix occurs between the possessed and the possessor.

But contact effects can proceed along a different path as well. In a number of areas of North America, grammatical structures of languages have been shaped by contact without substantial lexical borrowing.

4. Current contributions and research

It can be tempting to view contact situations as static, to assume that features transferred via contact were passed between languages in their current states among speakers in their current locations. Recognizing the element of time in contact situations can greatly enhance our understanding of the ways that contact can shape not only vocabulary but also language structure on all levels.

4.1 Phonology

The importance of the temporal dimension in shaping sound systems is illustrated by work on some Southern California languages by Hinton (1991). The languages are from five different genealogical groups, stretching from the Central California coast south to the Mexican border. Individual languages are in bold in (3). (Each family contains additional branches and languages.)

3. Southern California
 - Esselen** isolate
 - Salinan** isolate
 - Chumash family
 - Uto-Aztecan family
 - Takic branch
 - Cupan subgroup
 - Cupeno**
 - Cahuilla**
 - Luišeño**
 - Yuman family
 - California-Delta branch
 - Diegueño
 - Ipai, Kumeyaay, Tipai**
 - Cocopa**
 - River branch
 - Mojave**
 - Yuma**
 - Maricopa**

All of these groups participated in networks of contact within this area and beyond.

Hinton's focus is on the Cupan languages Cupeño, Cahuilla, and Luišeño, spoken in an area adjacent to the Yuman languages. They have not always been spoken in their current locations, however. It is estimated that Takic speakers separated from other Uto-Aztecan 4,500–3,000 years ago (Hale, 1958, cited in Hinton, 1991, p. 135). At that time they were in the northeastern portion of their current territory and have since been spreading southward and westward (Moratto, 1984, p. 559 cited in Hinton, 1991, p. 135). The other Southern California families were already in place at that point.

Hinton examined phonological developments within the Takic languages after their separation. She found three new phonological traits shared with Esselen, Salinan, Chumash, and Yuman. One is a distinction between velar and uvular stops (*k/q*). She suggests that both sounds may have been present in Proto-Takic but simply as allophones, with *q* before back vowels. Contact may have pushed them to become distinctive, as vocabulary was copied and conditioning environments disappeared. A second addition is the alveo-palatal affricate *č*. A third is the velar fricative *x*, which may have been a variant of *k* between vowels (except before *a*, where it was *q*).

Two additional traits then developed in Takic that exist only in Yuman: *s/š* and *r/l* distinctions. Within just the Cupan languages Cahuilla, Cupeño, and Luišeño, the sound *x^w* was added, also present in Yuman. In just Cahuilla and Cupeño, palatal *ñ* and *l^v* were added, also present in Yuman. While Proto-Uto-Aztecan had a five-vowel system, Proto-Yuman had a three-vowel system. Cahuilla and Cupeño merged **e* and **i* to *i*, perhaps under the influence of Yuman. A *k^w/q^w* distinction exists in all Yuman languages except Kiliwa, Paipai, and Diegueno, and

appears allophonically in Cahuilla and Cupeño. Interestingly, most Yuman influences did not come from the Yuman languages currently spoken in the region: Diegueño, Yuma, or Mojave. But their sources can be reconstructed for Proto-Yuman, the ancestor of the modern Yuman languages, estimated to have been spoken 4,800–2,300 years ago (Ochoa, 1982; Laylander, 1985, cited in Hinton, 1991, p. 138). Hinton concludes that Takic speakers replaced speakers of other Yuman languages when they moved into the region.

A few phonological traits are now entering Cupan languages from their modern Diegueño neighbours: voiceless laterals *l* and *lʰ*, and a *t/tʰ* distinction. These are not distinctive in any Uto-Aztecan languages of Southern California, but both Luiseño and Cupeño, those in closest contact with Digüeño, have them as allophones.

We can thus see layers of contact shaping the Cupan phonological systems over time. Hinton describes the social context in which they took place.

Language-exogamous marriages were so common among the Cupan peoples that women in the leadership families were raised multilingually with the express purpose of marrying into other groups. Early mission records show a great many intertribal marriages, and certainly today that is still the case . . . Cupan peoples spread their languages southward into Yuman territory through force and marriage, which resulted no doubt in a generation or more of bilingualism in Yuman–Cupan communities, wherein both languages were widely influenced. For each village, the Yuman language eventually disappeared, leaving only the changed Cupan language.

(Hinton, 1991, p. 154)

The evidence suggests that the mechanisms involved were not necessarily instigated by transfer of vocabulary, but rather strong bilingualism and perhaps language shift.

4.2 Morphology

North America provides numerous examples of contact effects on grammatical structures without substance. In many cases neighbouring languages show parallel morphological structures, but the morphemes themselves are nothing alike. It is difficult to imagine that speakers would simply copy an abstract morphological structure. Apparently, the structures were not copied in their current states. It was rather their precursors that were copied: propensities for specifying certain distinctions and using certain patterns of expression, with word-by-word translations in the target language.

Such effects can be seen in languages of California and beyond. Large numbers of languages, belonging to different genetic groups, show sets of means/manner verb prefixes, variously translatable with noun-like, verb-like or adverbial-like meanings. Aoki (1970, pp. 84–86) lists 167 such prefixes in Nez Perce, a Sahaptian language spoken in Idaho and adjacent Washington and Oregon, among them *wextú:-* ‘with seat, by sitting,’ *we:p-* ‘with hand or paw,’ *nim-* ‘with eyes, visually,’ *cú:-* ‘with pointed object,’ *him-* ‘with mouth,’ *té:-* ‘by speech,’ *çimí-* ‘by lying on,’ and *we:-* ‘flying.’

4. Nez Perce means/manner prefix (Aoki, 1970, p. 82)

Pí:waqalpsix.

pí:-weqe-lpis-i-k

RECIP-**in.arms**-seize-PL-INDIC

‘We are hugging each other.’

Large numbers of languages across the area also show locative/directional verb suffixes. Among the Nez Perce suffixes are *-éhse* ‘up,’ *-éhne* ‘down,’ *-éhyek* ‘upstream,’ *-ewí:* ‘downstream,’ *-eylé:k* ‘into,’ and *-éht* ‘out.’

5. Nez Perce locative/directional suffix (Aoki, 1970, p. 83)

watawí:ksa
 wet-**wéi:ks**-e
 wade-**upstream**-SG
 ‘I wade upstream.’

The inventories of such prefixes and suffixes can be quite large, but they vary across the languages in size and content. Within language families they are often cognate, suggesting that they are ancient, but across genetic lines there is little similarity in form. Apparently, it was not the forms themselves that spread, but rather propensities on the part of multilingual speakers to specify certain distinctions. Over time, frequently occurring expressions became routinized and processed as chunks, with means/manner and location/direction words fusing to verb roots and eventually eroding to affixes. A full description is in Mithun (2007a).

Another morphological pattern with no shared substance can be seen in the Northwest, a strong linguistic area, in languages of the Wakashan, Salishan, and Chimakuan families and the isolate Kutenai. All have what are termed ‘lexical suffixes,’ suffixes with generally more concrete meanings than other affixes. They can be numerous: Kinkade (1991) identified 298 in the Salish language Upper Chehalis. Their forms are often cognate within families, indicating that they are old, but there are no similarities across genealogical boundaries, and inventories vary in size and meanings. Apparently, what was transferred was a propensity to form certain kinds of compounds. Over time, perhaps under continuing contact, frequently recurring second constituents of such compounds eroded in form and became more general in meaning. The developments are described in further detail in Mithun (2017b).

4.3 Alignment

While a number of North American languages show nominative/accusative patterns, distinguishing subjects and objects, there are also geographical areas in which languages show other, less common patterns.

The Muskogean languages are indigenous to a wide area over the Southeast. All show essentially the same patterns and forms for indicating core arguments with pronominal affixes on their verbs. Examples in (6) are from Mikasuki of Florida.

6. Mikasuki (Boynton, 1982)

| | | |
|----|--------------------------|-----------------------|
| a. | <i>atayi-li</i> | ‘I’m going’ |
| | <i>impom-li</i> | ‘I’m eating’ |
| | <i>takalkom-li</i> | ‘I was working’ |
| | <i>hi:com-li</i> | ‘I saw (it)’ |
| | <i>yataplom-li</i> | ‘I hit (it)’ |
| | <i>cima:saho:cá:m-li</i> | ‘I will help you all’ |

- b. *ca-hicom* '(it) sees **me**'
ca-ko:slomicka 'you are cutting **me**'
is-ca-nonli:cickal 'you trick **me**'

The system looks straightforward, with suffix *-li* for 'I' and prefix *ca-* for 'me.' The suffix *-li* occurs with both intransitive verbs ('go,' 'eat,' 'work') and transitives ('see,' 'hit,' 'help'), and with all tenses. It appears to mark subjects. The prefix *ca-* appears to mark objects. But a closer look reveals a different picture. *Ca-* sometimes corresponds to English subjects.

7. Mikasuki (Boynton, 1982)
ca-to:lohayhkas 'I coughed'
ca-lipa:mis 'I'm going to die'
ca-ba:n-om 'I want (it)'
ca-wasi:lo:tom 'I'm ticklish'
ca-wasi:lom 'I itch'
ca-no:cipom 'I'm sleepy'

This is an agent/patient system. Essentially, participants in control are classified as grammatical agents, while those not in control but affected are grammatical patients: *-li* marks first person agents, and *-ca-* first person patients. Pronominal inflection is lexicalized with each verb, so that speakers are not making online decisions as they speak.

Such systems have sometimes been referred to as active/stative. The distinction does not reflect aspect, however. 'Go' and 'eat' which appear with *-li*, are actions/events, but so are 'cough' and 'die,' which appear with *ca-*. 'Being ticklish' is a state and it appears with *ca-*, but so is 'living somewhere,' which appears with *-li*. I have no control over being ticklish, but I do control where I live.

8. Mikasuki (Boynton, 1982)
coko:lom-li 'I live there'
no:ci:pa-li 'I'm going to sleep'

Note that 'being sleepy,' over which I have no control, appears with the patient prefix *ca-*, but 'going to bed' appears with the agent suffix *-li*. The agent/patient pattern and pronominal affix forms are old and can be reconstructed for Proto-Muskogean. (There is also a third Muskogean for participants indirectly affected.)

The Southeast is home to a number of other, unrelated language families and isolates as well, but it is a strong linguistic area. The other languages also show agent/patient systems, though the forms of the pronominal affixes are completely different. One of the isolates is Atakapa. Here first person pronominal agents are identified with a suffix *-ō*, and patients with a prefix *hi-*.

9. Atakapa (Swanton, 1929)
a. *tsak'c-ō* 'I chew'
wan puxkint-ō 'I (will) go dancing'
na iwēuckint-ō 'I am mocking you'
cakhu-ō 'I see them'
wi cakits-ō 'I wake somebody up'
wi hattsickic-ō 'I shall be chief'

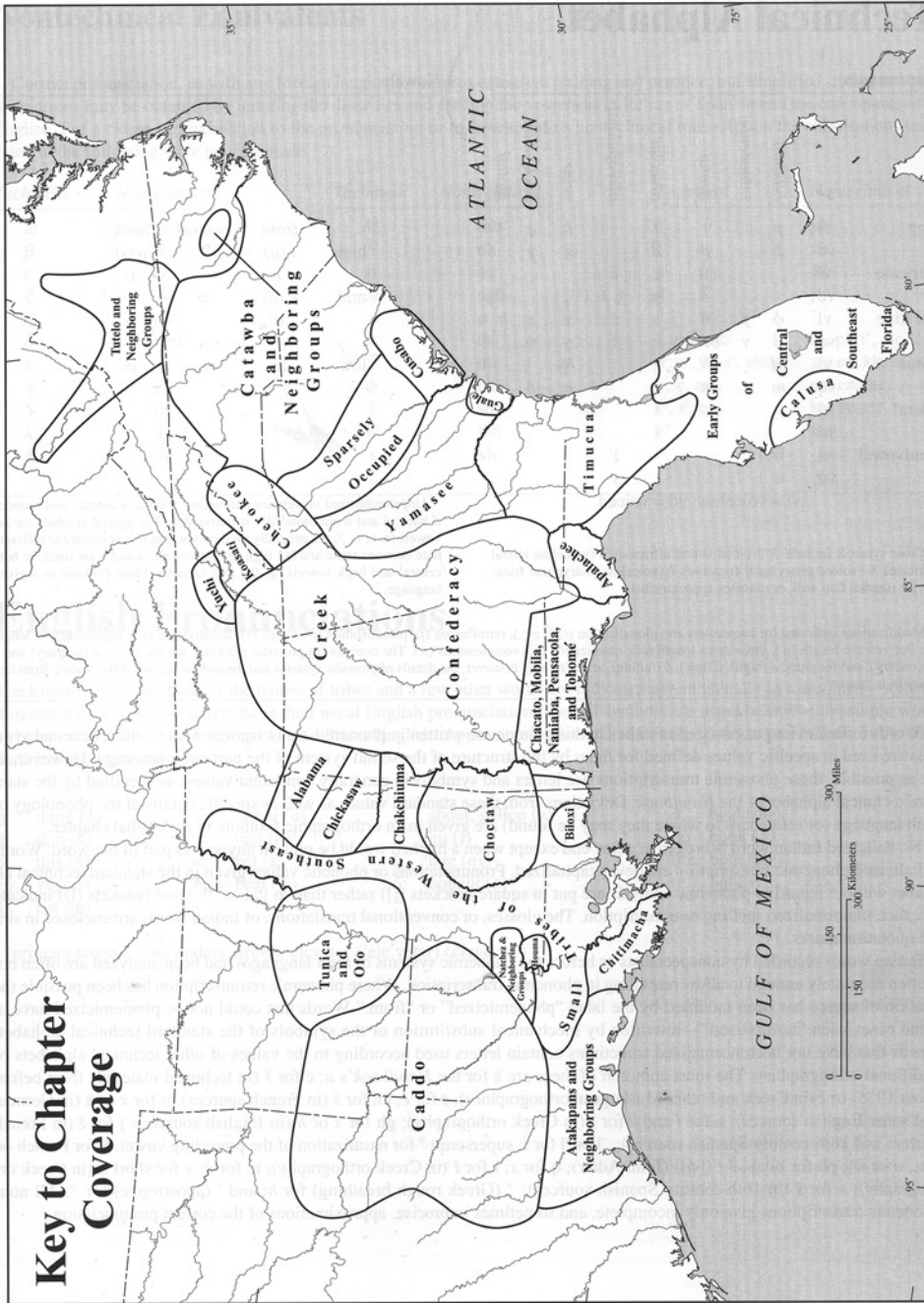


Figure 25.3 Southeast
Source: Fogelson, Raymond, ed. (2004). *Handbook of North American Indians: Southeast*. Washington: Smithsonian Institution, 14.ix

- b. *hi-nimahaxc* ‘Don’t kill **me!**’
kō-hi-pamulet ‘they seized and beat **me**’
hi-tsañct ‘**I** get pushed’
wi-hi-makawetl ‘**I** fell’
hi-cōkec ‘**I** am sick’
icak hi-imilc ‘**I** hate this man’

The system is not based on transitivity: the agent suffix *-ō* occurs with intransitive verbs (‘chew,’ ‘go dancing’) and transitives (‘see,’ ‘mock’). It is not based on aspect: the agent suffix occurs with both events (‘wake someone up’) and states (‘be chief’); as does the patient prefix *hi-* (‘kill,’ ‘beat,’ ‘push,’ ‘fall’) and states (‘be sick,’ ‘hate’).

Similar patterns can be seen across the Southeast, though the actual forms of the pronouns have nothing in common. (The systems are described in further detail in Mithun, 2008). It is unlikely that this is a coincidence. In her survey of alignment systems, Nichols (1992, p. 187) found that agent/patient systems occurred in just 14% of the 172 languages she examined. But how could such a system spread? In most of the languages described here, basic third persons are not represented in verbs. It would be easy for speakers to reanalyze transitive clauses as intransitive (or the reverse), and objects as patients, in accordance with systems in their first language.

10. Reanalysis

- ‘(it) sickens me.OBJECT’ < > ‘I.PATIENT am sick’
‘(it) hurts me.OBJECT’ < > ‘I.PATIENT am in pain’

Agent/patient systems appear in several other areas in North America, in neighbouring but unrelated languages. Yuki, a Yukian language of the Clear Lake area of California, shows an agent/patient system unlike its relative Wappo, but like its Pomoan neighbours. Tlingit, a Tlingit-Eyak-Athabaskan language of the Northwest, does as well, unlike all 36 other languages in the family, but like its unrelated neighbour Haida.

Several other less common alignment systems also show areal distributions. Hierarchical systems are found in the Northwest Coast and Northern California areas (Mithun, 2007b, 2012). Ergativity has developed in three languages on the Oregon coast (Mithun, 2005).

4.4 Contact languages

North America is also home to a number of languages whose vocabulary and grammar are drawn from multiple sources: jargons, pidgins, and mixed languages. The first two are used among speakers who share no other language. Jargons are, typically used for narrow purposes such as trade, with relatively little conventionalization. Pidgins are more conventionalized, must be learned, and are generally not understood by monolingual speakers of any of the contributing languages. The distinction between the two is not sharp, nor can it always be determined, particularly when the only available documentation is in historical accounts. Mixed languages, by contrast, emerge in communities of skilled bilinguals. They, too, are not understood by monolingual speakers of the source languages, but they often draw on the most complex aspects of these languages.

4.4.1 Jargons and pidgins

Jargons and pidgins developed in a number of areas of North America, some before the arrival of Europeans, and some after. Some were rudimentary, but others developed considerable complexity and were used over wide areas.

Pidgin Delaware was used between Delaware speakers of Unami and Munsee (Algonquian languages) and Dutch traders on the Delaware River in New Jersey beginning in the 1620's. It was eventually used by Swedish and English colonists as well. The last documentation of it is from 1785. Goddard (1997) provides a detailed history, description, and analysis.

The phonologies of Eastern Algonquian languages are not complex, and the sounds of Pidgin Delaware essentially match that of Unami, on which it is based. Initial unstressed short-vowel syllables are often lost, and consonant clusters simplified. Goddard cites, for example Pidgin *gōs* 'nail' compared to Southern Unami *mako-s*, and *pack* 'weep' compared to *lapákw* 'he weeps' (1997, p. 64, from Campanius, 1696, pp. 143, 141).

The morphological structures of Algonquian languages can be complex, however. The Pidgin shows no active morphology, though some forms contain frozen affixes. Distinctions indicated by prefixes and suffixes in Algonquian languages, such as pronominal arguments, are expressed instead with independent words. The same pronouns are used for subjects, objects, indirect objects, and possessors; for singulars and plurals; and for animates and inanimates.

11. Delaware Pidgin (Goddard, 1997, p. 59, citing Thomas, 1698)

Keco kee hatah kee weekin?
 what 2 have 2 house
 'What do you have in your house?'

The same verb forms are used for all persons. The Pidgin verb *entaami* 'rise up' probably comes from a verb with a first person subject prefix *n-*, but it is used with all subjects. Algonquian verbs take different forms according to the animacy of their subjects (if intransitive) or objects (if transitive). The same Pidgin verb forms are used with both animate and inanimate participants. The transitive verb *tahóttamen* 'love, like' appears to be adopted from a Unami verb with animate suffix *-amən*, as in *ntahó:t:amən* 'I love it,' but it is used no matter what is liked (Goddard, 1997, p. 67, citing Campanius, 1696, pp. 33, 36, 2).

Vocabulary is built up by compounding. A term for 'mountain lion' is *Manúnckus mochijrick Sínwæs*, literally 'angry big wildcat' (Goddard, 1997, p. 72, citing Campanius, 1696, p. 145). Some function words have developed from content words. The word recorded as *taan*, *thaan* 'to, of, from, until' matches *taan* 'go.' The word *ana* 'by' matches *aana* 'road, way, route,' exemplified in the phrase *ana mochijrick bij* 'by sea,' literally 'road big water' (Goddard, 1997, p. 73, citing Campanius, 1696, p. 4).

Like most pidgins, Delaware Pidgin shows considerable variation, in part because it was used by speakers of different dialects and languages.

On the other side of the continent is Chinook Jargon. It was once spoken from the southern border of Oregon north to the Alaska panhandle, and from the Pacific coast east to Montana. It is estimated that there were perhaps 100,000 speakers, with over 100 mother tongues, languages from the Chinookan, Salishan, Wakashan, Tsimshianic, Athabaskan, Sahaptian, and Kalapuyan families in the area, and also eventually French, English, Hawaiian, Chinese, and others. There is an extensive literature on it. Good overviews are in Jacobs, 1932; Boas, 1933; Thomason, 1983; and Thomason and Kaufman, 1988, pp. 256–263. The earliest documentation

is from the 1830's by Europeans not trained in linguistics, but later records are more sophisticated. As with most pidgins, particularly those spoken over a wide area, there was variation. Thomason argues convincingly that it did not begin with the arrival of Europeans, but rather was already fully formed well before. It shows many marked features that are unusual cross-linguistically, but common in languages indigenous to the Northwest.

The phonological inventory contains numerous sounds that are rare cross-linguistically but common in the region, such as distinctions between velars and uvulars ($k, k^w, k', k^{w'}$, x, x^w versus $q, q^w, q', q^{w'}$, $\text{ɕ}, \text{ɕ}^w$), plain and labialized velars and uvulars ($k, k', x, q, q', \text{ɕ}$ versus $k^w, k^{w'}, x^w, q^w, q^{w'}, \text{ɕ}^w$), plain and ejective stops and affricates ($p, t, ts, t\text{ɕ}, k, k^w, q, q^w$ versus $p', t', t', ts', k', k^{w'}, q', q'$), and voiced and voiceless laterals (l versus l').

As is common among pidgins, there are not complex morphological structures, though such structures are common in the area. Jacobs notes, however, the beginning of such developments.

The most interesting structural trait in Jargon is this compounding, clustering, or tying of words. To generalize: compounds and clusters are of several degrees of firmness of knit; the more frequent component elements of compounds are, in order from firmer to looser: the shorter personal pronouns, the demonstrative article, the auxiliary verbs, the adverbs.

(Jacobs, 1932, p. 34)

In certain traits, the language seems to be moving towards polysynthesis (Jacobs, 1932, p. 38).

As in many languages indigenous to the region, pronominal enclitics appear even when coreferential lexical nominals are also present in the clause. The same proclitics are used for subjects and possessors. Enclitics are used for objects.

12. Chinook Jargon pronominal clitics (Jacobs, 1932, p. 44)

- a. *Ya-ládwá.*
3SG-go
'He went.'
- b. *Ya-wáwa-yáxga.*
3SG-tell-3SG
'He told me.'
- c. *Ya-húšmən* *ya-sa:likš.*
3SG-woman 3SG-be.angry
'The woman was angry.'
- d. *uk-húšmən* *ya-háus*
the-woman 3SG-house
'the woman's house'

Tense and aspect are not indicated by verbal morphology, but time can be indicated with separate adverbials. As in many pidgins, negation is marked with an initial particle, here *wik* 'no, not.' Sometimes Jacobs writes it as a separate word, sometimes attached to a following verb.

13. Chinook Jargon negation (Jacobs, 1932, p. 40)

- Wik háyu nánitš ayáq ayáq!*
not very look quick quick
'Don't be looking so often!'

14. Chinook Jargon negation (Jacobs, 1932, p. 47)

Wík-ya-tq'í mákmaq lí:li.
 NEG-3SG-want eat-eat long.time
 'For a long time she would not eat.'

Vocabulary is largely from Lower Chinook (Chinookan family), Nuuchahnulth (Nootka, Wakashan family), Chehalis (Salishan family), with later loans from English and French.

Some of these traits can be seen in the following excerpt from a text recorded by Jacobs in 1930 from Mrs. Victoria Howard, of Oregon City, whose first language was Clackamas (Upper Chinook). Word boundaries and hyphens are as in the original.

15. Chinook Jargon text (Jacobs, 1932, pp. 45–50)

Las-mí:lait íxt-lamyái bi-yá-kwi?im.
 3PL-live one-old.woman and-3SG-grandchild
 they were living one old lady and her grandchild
 'One old lady and her grandchild were dwelling there.'

Uk-dənəs-lútšmən gwa:nisim ya-ládwa múŋk-lagámas.
 that-little-female always 3SG-go do-camas
 that little girl always she went to dig camas
 'The little girl always used to go dig camas (cat-ear roots).'

Gwá:nisim ya-q'ú? k'í:labai kába-las-háus.
 always 3SG-come home there-3PL-house
 always she come home to their house
 'She would always return home.'

Alda-ya-p'inəs úk-lagamás.
 then-3SG-bake that-camas
 then she bake the camas
 'Then she would bake the camas under ashes.'

Álda-ya-ú:ma ya-tsə'tš.
 then-3SG-give 3SG-grandmother
 then she give to her grandmother.
 'Then she would give them to her grandmother.'

Jargons and pidgins in other areas of North America have also been described. Bakker (1989, 1991, 1996a, 1996b) discusses possible trade jargons and pidgins in eastern Canada between the Gulf of St Lawrence and the entry to Hudson Bay. The area was in use by Inuit, Innu (Montagnais), Saint Lawrence Iroquoians, Beothuk, and Micmac over centuries. Their languages are mutually unintelligible, but archaeological and historical evidence indicates they engaged in trade. There were also early Norse settlements, and from about 1500, Portuguese, English, Norman, Breton, and Basque fishermen and traders. Bakker discusses a possible Micmac pidgin Portuguese from the sixteenth century, a Micmac pidgin Basque, and Montagnais pidgin Basque from the sixteenth and seventeenth centuries, an Inuktitut French jargon from the seventeenth and eighteenth centuries, an Inuktitut English Pidgin from the

seventeenth century, a Micmac Pidgin English from the nineteenth century, and pidginized Inuktitut used in the twentieth century. Van der Voort (1996) describes a more or less stable contact variant of Kalaallisuut (Greenlandic) used between the Greenlandic Inuit and European whalers and traders from the seventeenth to the nineteenth centuries. Bakker points out that most of these pidgins contain vocabulary from multiple sources and identifies an Inuit Pidgin French-Inuktitut-Montagnais-Basque, and an Inuit Pidgin Inuktitut-Cree, with terms passed along by middlemen through trade networks. Discovering the true nature of these early contact languages is a challenge, because documentation is sparse and orthographic conventions not established.

In the western Canadian Arctic there were at least two jargons. Broken Slavey, also known as Slavey Jargon, was used in the Mackenzie River district and Jargon Loucheux along the Yukon River. Bakker (1996b) surveys descriptions of the languages and cites the missionary Petitot (1889), who reported that the first was made up of French, Cree, and Dene Slave, and the second of French, English, Chipewyan, Slave, and Gwich'in (Loucheux). Cree is an Algonquian language, and the others are Athabaskan, known for their complex phonologies and verb morphologies. They were apparently used by speakers of Slave and Chipewyan (Athabaskan), Cree, French and other European languages, and Inuit.

In the Southwest was Trader Navajo, a simplified version of Navajo used by Anglo traders at isolated trading posts (Werner, 1963).

In the Southeast was Mobilian Jargon (Crawford, 1978; Drechsel, 1997). It apparently developed in the lower Mississippi Valley in what is now French Louisiana, and eventually spread north into present Illinois, east into Florida, and west into Texas. It was eventually used by speakers of Choctaw, Chickasaw, Apalachee, Alabama, and Koasati (Muskogean family); Atakapa, Chitimacha, Natchez, and Tunica (Gulf isolates); Ofo and Biloxi (Siouan family); Caddo (Caddoan family); perhaps Kiowa Apache (Athabaskan); some Algonquian languages; and by European speakers of French, English, Spanish, and German. Most of the lexicon is from Choctaw and Alabama.

4.4.2 *Mixed languages*

Mixed languages arise from a combination of two identifiable source languages, usually in communities of fluent bilinguals. Some important ones are indigenous to North America.

Michif is spoken by the Métis, descendants of Cree-speaking mothers and French-speaking fur trader fathers. It developed in Saskatchewan and Manitoba in Canada, and North Dakota and Montana in the United States. It is estimated that there were 2,000–3,000 speakers at the turn of the last century (Bakker, 1997, p. 3). Though it was originally formed by skilled Cree-French bilinguals, it is not mutually intelligible with either Cree or French.

Cree, an Algic language, has elaborate verb morphology, but relatively simple nouns. Michif combines the complexity of Cree verb structure with the complexity of French noun phrase structure. Essentially, verbs show Cree phonology, while noun phrases show French phonology. Both languages have obligatory grammatical gender distinctions, but Cree distinguishes animates and inanimates, and French masculines and feminines. Referents are accordingly distinguished as animate/inanimate within the verb, but masculine/feminine within the noun phrase. In (16), the verb 'he pulled it' is essentially Cree. The suffix *-am* indicates that it is transitive inanimate (with grammatically inanimate object, here the head), and has a third person subject and fourth person object. The rest of the sentence is essentially a Canadian

French noun phrase: ‘the turtle’s head.’ Here the head is classified as grammatically feminine, as we can see from the form *sa* of ‘his.’ The turtle is also classified as grammatically feminine, indicated by the article *la* ‘the.’

16. Michif (Bakker, 1997, p. 87)

| | | | | |
|------------------------------------|-----------|------------|-----------|---------------|
| <i>Ki:-učipit-am</i> | <i>sa</i> | <i>tet</i> | <i>la</i> | <i>tərčy.</i> |
| PAST-pull-TR.INAN.3>4 | his.F | head | F.SG.DEF | turtle |
| he pulled it | his | head | the | turtle |
| ‘The turtle pulled back his head.’ | | | | |

The constituent order is essentially Cree, governed by the information each constituent contributes at that point in the discourse.

Another North American mixed language is Mednyj Aleut or Copper Island Aleut (CIA). It apparently emerged in the nineteenth century on Mednyj (Copper) Island, part of a western extension of the Aleutian chain off the southwest coast of Alaska. Russian fur trappers and traders established settlements on the islands in 1819, then brought in Aleuts from various islands beginning in 1826. Children of Aleut mothers and Russian fathers, known as Creoles, were fluent Aleut-Russian bilinguals, but they were not fully accepted by either group. Thomason (1997, p. 452) reports that in 1860, before the departure of the fur company, there were 4656 Aleuts, 595 Russians, and 1896 Creoles. The Mednyj Aleut language, a combination of Aleut and Russian, emerged in this last group. In the late 1960s all speakers were moved to Bering Island.

English descriptions of the language are in Golovko (1994, 1996, 2003), Golovko and Vakhtin (1990), Sekerina (1994), and Thomason (1997). It combines structural features of both languages, but most vocabulary is from Aleut. Working from a vocabulary of about 500 words, Sekerina (1994) found that 94% of verb roots were from Aleut and 6% from Russian; 61.5% of nouns were from Aleut and 38.5% from Russian (particularly terms for introduced items), and 31.5% of function words were from Aleut and 68.5% from Russian.

The phonology is nearly the same as that of Aleut, though the Aleut velar/uvular opposition is no longer obligatory, and Aleut phonemes /δ/, /hŋ/, and /hw/ are absent from Mednyj Aleut. Russian words that entered the language early were assimilated to Aleut phonology, but later ones were not. A *p/b* distinction not present in Aleut appears in Russian loans, and Russian words maintain voicing assimilation and final devoicing of obstruents. Phrasal intonation is Russian.

Eskimo-Aleut morphology is complex, and most of the complexity appears in Mednyj Aleut as well. Aleut itself distinguishes ergative and absolutive noun case, with the ergative matching the genitive. Mednyj Aleut has both forms, but the absolutive is now used for all subjects, and the ergative now serves only as a genitive. Possessive pronouns in Eskimo-Aleut languages are transitive (POSSESSOR>POSSESSED), and Mednyj Aleut concurs. Nominal and verbal derivational suffixes match Aleut forms as well. Noun suffixes include an augmentative; diminutive; agentive, instrumental, and locative nominalizers; and suffixes ‘good’ and ‘bad.’ Verb suffixes include a rich set of causatives, transitivity and detransitivizers, a verbalizer, desiderative, continuative, durative, inceptive, completive, iterative, distributive, resultant state, suffixes ‘have’ and ‘do well.’

Finite verb inflection, however, is essentially Russian. Verbs are inflected for person, number, tense, and optionally gender. (The symbols \hat{x} and \hat{g} represent uvular fricatives. Apostrophes indicate palatalization.)

17. Mednyj Aleut verb inflection ‘stand’ (Golovko, 1994, p. 70)

| | Mednyj Aleut | Russian |
|-----|--------------------|----------------|
| | <i>anqaġta-</i> | <i>sto-</i> |
| 1SG | <i>anqaġta-ju</i> | <i>sto-ju</i> |
| 2SG | <i>anqaġta-iš</i> | <i>sto-iš’</i> |
| 3SG | <i>anqaġta-it</i> | <i>sto-it</i> |
| 1PL | <i>anqaġta-im</i> | <i>sto-im</i> |
| 2PL | <i>anqaġta-iti</i> | <i>sto-ite</i> |
| 3PL | <i>anqaġta-jut</i> | <i>sto-jat</i> |

Past tense sentences are formed with the Russian verb suffix *-l* and Russian independent pronouns. Future tense is formed periphrastically on the Russian model, with the auxiliary *bud-* inflected with Russian-based person endings followed by a verb with the Russian infinitive suffix *-t’*. Non-finite verb forms, however, are Aleut.

In all Eskimo-Aleut languages, arguments are identified by pronominal suffixes on verbs. An example from the Aleut spoken on the same island is in (18a). Mednyj Aleut has largely replaced these with independent Russian subject and object pronouns, as can be seen by comparing (18b) and (18c).

18. Mednyj Aleut pronouns (Golovko, 1994, p. 71)

a. Bering Island Aleut

Ilaġta-ku-u.

love-REAL-3SG>3SG

‘S/he loves him/her/it.’

b. Mednyj Aleut

Ona hiġta-it ċto ona ego ilaġta-it.
 she.NOM say-3SG.PRES that she.NOM he.ACC love-3SG.PRES
 ‘She says that she loves him.’

c. Russian

Ona govor-it ċto ona ego ljub-it.
 she.NOM say-3SG.PRES that she.NOM he.ACC love-3SG.PRES
 ‘She says that she loves him.’

Aleut contains two negative suffixes, a realis *-laka(ġ)-* and an irrealis *-(g)ula-(x)-*. Both are replaced in Mednyj Aleut with the Russian prefix *ne-* [ni], as can be seen in (19). (Aleut itself has no prefixes.)

19. Mednyj Aleut negation (Golovko, 1994, p. 72)

a. Bering Island Aleut

Iglu-ng tuta-qaġi-laka-ġ.

grandson-1SG listen-DETRANS-NEG-3SG.PRES

‘My grandson does not listen.’ (= ‘My grandson does not obey.’)

b. Mednyj Aleut

Iglu-ng n’i tuta-qaġi-it.

grandson-1SG NEG hear-DETRANS-3SG.PRES

Other words from Russian include some modals (*nado* ‘ought to,’ *dolžen* ‘must,’ *moč’* ‘can,’ *xotet* ‘want’), the predicate negator *net(u)*, and various conjunctions and adverbs of time, degree, and quality.

While Eskimo-Aleut languages generally show clear SOV order, Mednyj Aleut order follows the more pragmatic Russian patterns. In complex sentences, Russian conjunctions and complementizers are used, as in (18b).

A more recent mixed language is Bilingual Navajo, described by Schaengold (2004). It emerged when Navajo-speaking children were sent away to boarding school during the first half of the twentieth century and forbidden to speak their language. When they returned to their Navajo communities, they brought it with them. Neither Navajo nor English monolinguals understand it, and fluent Navajo-English bilinguals cannot automatically speak it.

Bilingual Navajo draws most of its grammatical and discourse structure from Navajo, but some vocabulary from English, primarily nouns, some adjectives and adverbs, and a few verbs. Standard Navajo often contains nouns borrowed from English. They are easily integrated into Navajo structures, where nouns have little morphological complexity in Navajo, apart from possessive prefixes and locative suffixes. Loanwords for ‘clinic’ and ‘check-up’ in Standard Navajo can be seen in (20a). Some adverbs and adverbial phrases are integrated for similar reasons. In both Standard and Bilingual Navajo the pronunciation of the copied words is adjusted to Navajo phonology, like that of ‘clinic.’ The Bilingual Navajo sentence in (20b) contains a borrowed adverbial phrase *all day* and lacks the subordinate clause structure of the Standard equivalent, marked morphologically with the subordinator *-go*.

20. Bilingual Navajo (Schaengold, 2004, p. 50)

a. Standard Navajo

Tłinic-di check-up *bi-niyé sédá:-go*

clinic-at 3POSS-reason 1SG.PFV.wait-SUB

‘I went to the clinic for a check-up and’

shei’í’ah.

1SG.OBJ.PFV.went.down

‘waited until it [the sun] went down with me.’

b. Bilingual Navajo

Tłinic-di check-up *bi-niyé all day sédá-*

clinic-at 3POSS-reason 1SG.PFV-wait

‘I went to the clinic for a check-up and waited [there] all day.’

Navajo verb morphology can be complex, like that of many North American languages, with distinctions of person (first, second, third, fourth), number (singular, dual, plural), tense, aspect, mood, and transitivity, often with considerable fusion. It is not easy to incorporate English verbs into the structure. When verbs are brought in, an auxiliiation strategy is exploited involving a verb *-lééh* ‘make/prepare.’

21. Bilingual Navajo (Schaengold, 2004, p. 54)

a. *Bi-’ée’ change íi-lééh.*

3POSS-clothes 1DU-make

‘We are changing her clothes.’

b. Standard Navajo

Bi-’ée’ lahgo át’éhí b-ii ndeezhteeh.

3POSS-clothes altered differently 3-into 1DU.move.animate.being.PROG

Furthermore, sets of what are termed ‘handling verbs’ in Navajo distinguish the shape, consistency, number, or animacy of items moving, handled, or located somewhere. The Standard Navajo verb in (21b) is based on the progressive form *-teeh* of the stem *-tj* ‘handle single animate being.’ The Bilingual Navajo counterpart does not require such classification. The event is portrayed slightly differently in the Bilingual and Standard Navajo versions: in (21a), it is the clothes that are manipulated. In (21b) it is the person who is manipulated, propelled into clothes.

5. Future directions

Over the last century, purism has often been a value in the documentation of languages indigenous to North America. Speakers working to leave a record of their traditional languages have often put special effort into avoiding obvious recent loanwords from English, French, or Spanish, or code-switching, as they were being recorded and certainly as they produced curriculum materials for heritage language learners. Such concern is not surprising in these situations where European languages are menacing the traditional ones to the point where so many are now spoken only by the oldest generation, and many others by no one. As a result, there may be some underdocumentation of the dynamics of language contact situations involving these languages. In some communities, however, attitudes are beginning to change, as revitalization efforts gain momentum and knowledge of more than one language is recognized as a value. Because the vast majority of speakers of languages indigenous to North America are bilingual, opportunities are now greater than ever for learning about effects of language contact under both situations of language maintenance (in elder first-language speakers) and language shift (in younger, second-language learners).

At the same time, modern speakers are generally less aware of older contact effects among indigenous languages, so there is little effort to avoid evidence of them. For the most part, the lingua franca for indigenous peoples of North America communicating across language lines is now English, or in some areas French or Spanish. Fewer speakers now recognize loanwords from other indigenous languages, and they are certainly not conscious of the kinds of factors that can lead to the long-term structural convergence discussed in earlier sections, such as replication of frequencies of certain semantic distinctions or particular syntactic constructions.

At present, as more detailed documentation of North American languages becomes available, we continue to learn about the kinds of features that can be transferred in contact situations. Particularly important is the documentation of unscripted connected speech, especially conversation, which holds the key to understanding the kinds of features that can be transferred in contact situations, some of the cultural circumstances which may promote or restrain such transfer, and the pathways via which grammatical patterns may converge.

6. Further reading

Bakker, P. (1997). *A language of our own*. Oxford: Oxford University Press.

This volume describes the genesis of the mixed Cree-French language Michif.

Mithun, M. (2008). The emergence of agentive systems. In: M. Donohue and S. Wichmann, eds., *The typology of semantic alignment systems*. Oxford: Oxford University Press, pp. 297–333.

This chapter traces the areal concentrations of agent/patient systems in North America and describes how they can spread through contact.

Mithun, M. (2012). Core argument patterns and deep genetic relations: Hierarchical systems in Northern California. In: P. Suihkonen, B. Comrie and V. Solovyev, eds., *Typology of argument structure and grammatical relations*. Amsterdam: John Benjamins, pp. 257–294.

This chapter shows how alignment systems in unrelated neighboring languages have converged via different pathways of development.

Thomason, S. (1983). Chinook Jargon in areal and historical context. *Language*, 59, pp. 820–870.

This article addresses two controversies about the nature and origin of Chinook Jargon: whether it is a true pidgin rather than a jargon, and whether it existed before the arrival of Europeans.

7. Related topics

Social factors, pragmatic factors, cognitive factors, typological factors, convergence, creoles and pidgins, mixed languages

Abbreviations

| | |
|---------|---------------------|
| 1 | first person |
| 2 | second person |
| 3 | third person |
| ACC | accusative |
| CIA | Copper Island Aleut |
| DEF | definite |
| DETRANS | detransitivizer |
| DU | dual |
| F | feminine |
| INAN | inanimate |
| NEG | negative |
| NOM | nominative |
| OBJ | object |
| PFV | perfective |
| PL | plural |
| POSS | possessive |
| PROG | progressive |
| PRES | present |
| SG | singular |
| SOV | subject-object-verb |
| SUB | subordinate |
| TR | transitive |

References

- Aoki, H. (1970). Nez Perce Grammar. *University of California Publications in Linguistics* 62. Berkeley: University of California.
- Bakker, P. (1989). The language of the coast tribes is half Basque. A Basque-Amerindian pidgin in use between Europeans and Native Americans in North America, ca. 1540 – ca. 1640. *Anthropological Linguistics*, 31, pp. 117–147.
- Bakker, P. (1991). Trade languages in the strait of Belle Isle. *Journal of the Atlantic Provinces Linguistic Association*, 13, pp. 1–19.

- Bakker, P. (1996a). Language contact and pidginization in Davis Strait, Hudson, Strait, and the Gulf of Saint Lawrence (northeast Canada). In: E.H. Jahr and I. Broch, eds., *Language contact in the Arctic. Northern Pidgins and contact languages*, 1st ed. Berlin: De Gruyter Mouton, pp. 261–310.
- Bakker, P. (1996b). Broken Slavey and Jargon Loucheux: A first exploration. In: E.H. Jahr and I. Broch, eds., *Language contact in the Arctic. Northern Pidgins and Contact Languages*, 1st ed. Berlin: De Gruyter Mouton, pp. 317–320.
- Bakker, P. (1997). *A language of our own*. Oxford: Oxford University Press.
- Boas, F. ([1894] 1974). Classification of the languages of the North Pacific Coast. In: *Memoirs of the International Congress of Anthropology*. Chicago: Schulte, pp. 339–346.
- Boas, F. (1920). The classification of American languages. *American Anthropologist*, 22, pp. 367–376.
- Boas, F. (1933). Note on the Chinook Jargon. *Language*, 9, pp. 208–213.
- Boynton, S. (1982). *Mikasuki grammar in outline*. PhD. University of Florida.
- Bright, W. and Sherzer, J. (1976). Areal features in North American Indian languages. In: W. Bright, ed., *Variation and change in language*, 1st ed. Stanford: Stanford University, pp. 228–268.
- Callaghan, C. (1964). Phonemic borrowing in Lake Miwok. In: W. Bright, ed., *Studies in California Linguistics*. University of California Publications in Linguistics 34. Berkeley, CA: University of California Press, pp. 46–53.
- Campanius, J. (1696). *Lutheri Catechismus Öfwersatt på American-Virginiske Språket*. Stockholm: Burchardi. (Facsimile edition issued with the title: *Martin Luther's little catechism translated into Algonquian Indian*, bound with Collijn 1937. Stockholm: Ivar Hæggström and Uppsala: Almqvist and Wiksell, 1937).
- Campbell, L. (1997). *American Indian languages: The historical linguistics of Native America*. Oxford: Oxford University Press.
- Campbell, L. and Mithun, M. (1979). Introduction: North American Indian historical linguistics in current perspective. In: L. Campbell and M. Mithun, eds., *The languages of Native America: Historical and comparative assessment*, 1st ed. Austin, TX: University of Texas, pp. 3–69.
- Crawford, J. (1978). *The Mobilian Trade Language*. Knoxville: University of Tennessee.
- Darnell, R. and Sherzer, J. (1971). Areal linguistic studies in North America: A historical perspective. *International Journal of American Linguistics*, 37(1), pp. 20–28.
- Drechsel, E. (1997). *Mobilian Jargon*. Oxford: Oxford University Press.
- Goddard, I. (1997). Pidgin Delaware. In: S. Thomason, ed., *Contact languages: A wider perspective*, 1st ed. Amsterdam: John Benjamins, pp. 43–98.
- Golovko, E.V. (1994). Mednyj Aleut or Copper Island Aleut: An Aleut-Russian mixed language. In: P. Bakker and M. Mous, eds., *Mixed languages: 15 case studies in language intertwining*, 1st ed. Amsterdam: Uitgever IFOTT, pp. 113–121.
- Golovko, E.V. (1996). A case of non-genetic development in the Arctic area: The contribution of Aleut and Russian to the formation of Copper Island Aleut. In: E.H. Jahr and I. Broch, eds., *Language contact in the Arctic. Northern Pidgins and Contact Languages*, 1st ed. Berlin: De Gruyter Mouton, pp. 63–77.
- Golovko, E.V. (2003). Language contact and group identity: The role of 'folk' linguistic engineering. In: Y. Matras and P. Bakker, eds., *The mixed language debate: Theoretical and empirical advances*, 1st ed. Berlin: De Gruyter Mouton, pp. 177–208.
- Golovko, E.V. and Vakhtin, N. (1990). Aleut in contact: The CIA enigma. *Acta Linguistica Hafniensia*, 72, pp. 97–125.
- Hale, K. (1958). Internal diversity in Uto-Aztecan: I. *International Journal of American Linguistics*, 14, pp. 101–107.
- Hinton, L. (1991). Takic and Yuman: A study in phonological convergence. *International Journal of American Linguistics*, 57, pp. 133–157.
- Jacobs, M. (1932). Notes on the structure of Chinook Jargon. *Language*, 8(1), pp. 27–50.
- Kari, J. (1977). Linguistic diffusion between Tanaina and Ahtna. *International Journal of American Linguistics*, 43(4), pp. 274–288.
- Kinkade, M.D. (1991). Upper Chehalis dictionary. In: *University of Montana Occasional Papers in Linguistics No. 7*. Montana: Linguistics Laboratory, University of Montana.

- Kroskirty, P. (1985). Areal-historical influences on Tewa possession. *International Journal of American Linguistics*, 51, pp. 46–49.
- Laylander, D. (1985). Some linguistic approaches to Southern California's prehistory. *Casual papers, cultural resource management*, 2.1. San Diego: Cultural Resource Management Center, Department of Anthropology, San Diego State University.
- Loether, C. (1998). Yokuts and Miwok loanwords in Western Mono. In: J. Hill, P.J. Mistry, and Lyle Campbell, eds., *The life of language*, 1st ed. Berlin: De Gruyter Mouton, pp. 101–121.
- Mithun, M. (1999). *The languages of Native North America*. Cambridge: Cambridge University Press.
- Mithun, M. (2005). Ergativity and language contact on the Oregon Coast: Alsea, Siuslaw, and Coos. *Proceedings of the Berkeley Linguistics Society*, 26, pp. 77–95.
- Mithun, M. (2007a). Grammar, contact, and time. *Journal of language Contact – THEMA*, 1, pp. 144–167.
- Mithun, M. (2007b). Integrating approaches to diversity: Argument structure on the Northwest Coast. In: Y. Matsumoto, D. Oshima, O. Robinson and P. Sells, eds., *Diversity in language*, 1st ed. Stanford: Center for the Study of Language and Information, pp. 9–36.
- Mithun, M. (2008). The emergence of agentive systems. In: M. Donohue and S. Wichmann, eds., *The typology of semantic alignment systems*. Oxford: Oxford University Press, pp. 297–333.
- Mithun, M. (2010). Contact and North American languages. In: R. Hickey, ed., *Handbook of language contact*, 1st ed. Oxford: Blackwell, pp. 673–694.
- Mithun, M. (2012). Core argument patterns and deep genetic relations: Hierarchical systems in Northern California. In: P. Suihkonen, B. Comrie and V. Solovyev, eds., *Typology of argument structure and grammatical relations*. Amsterdam: John Benjamins, pp. 257–294.
- Mithun, M. (2017a). Native North American languages. In: R. Hickey, ed., *The Cambridge handbook of areal linguistics*, 1st ed. Cambridge: Cambridge University Press, pp. 878–933.
- Mithun, M. (2017b). Polysynthesis in North America. In: N. Evans, M. Fortescue and M. Mithun, eds., *Handbook of polysynthesis*, 1st ed. Oxford: Oxford University Press, pp. 235–259.
- Mithun, M. (to appear). Contact and North American languages. In: R. Hickey, ed., *Handbook of language contact*, 2nd ed. Oxford: Wiley-Blackwell.
- Moratto, M. (1984). *California archaeology*. New York: Academic.
- Nater, H. (1974). Lexical comparisons between Bella Coola and neighboring languages. In: *Ninth International Conference on Salish languages*. Leiden: University of Leiden.
- Nater, H. (1977). Stem list of the Bella Coola language. In: *PdR Publications on Salish Languages*, No. 4. Lisse: Peter de Ridder.
- Nater, H. (1984). *The Bella Coola language*. Ottawa: National Museum of Man Mercury Series 92.
- Nater, H. (1994). The Athapaskan component of Nuxalk. *International Journal of American Linguistics*, 60(2), pp. 177–190.
- Nichols, J. (1992). *Linguistic diversity in space and time*. Chicago: University of Chicago.
- Ochoa, Z. and Angel, J. (1982). *Sociolingüística de Baja California*. Los Mochis: Universidad de Occidente.
- Petitot, E. (1889). *Quinze ans sous le Cercle Polaire: Mackenzie, Anderson, Youkon*. Paris: E. Dentu.
- Powell, J.W. (1891). Indian linguistic families of America north of Mexico. *Bureau of American ethnology-annual report to the secretary of the Smithsonian Institution*, 7. (Reprinted in P. Holder, ed. (1996)) Lincoln: University of Nebraska.
- Sapir, E. (1916). Time perspective in aboriginal American culture, a study in method. In: *Geological Survey Memoir 90: 13*. Ottawa: Government Printing Bureau.
- Sapir, E. (1921a). A bird's eye view of American languages north of Mexico. *Science* n.s. 54: 408. (Reprinted in W. Bright, ed. (1990). *The collected works of Edward Sapir 5: American Indian languages*. Berlin: De Gruyter Mouton, pp. 93–94).
- Sapir, E. (1921b). *Language: An introduction to the study of speech*. New York: Harcourt, Brace.
- Sapir, E. (1929). Central and North American Indian languages. *Encyclopaedia Britannica*, 14th ed., vol. 5, pp. 138–141. (Reprinted in D. Mandelbaum, ed. (1949, 1963). *Selected writings of Edward Sapir in language, culture, and personality*. Berkeley: University of California, pp. 169–178, and in W. Bright,

- ed. (1990). *The collected works of Edward Sapir 5: American Indian languages*. Berlin: De Gruyter Mouton, pp. 95–104.
- Schaengold, C. (2004). *Bilingual Navajo: Mixed codes, bilingualism and language maintenance*. PhD. Ohio State University.
- Sekerina, I. (1994). Copper Island (Mednyj) Aleut: A mixed language. *Languages of the World*, 8, pp. 14–31.
- Sherzer, J. (1973). Areal linguistics in North America. In: T. Sebeok, ed. *Current trends in linguistics 10: Linguistics in North America*, pp. 749–795. (Reprinted in T. Sebeok, ed., *Native languages of the Americas*. New York: Plenum, pp. 121–173).
- Sherzer, J. (1976). *An areal-typological study of American Indian languages North of Mexico*. Amsterdam: North-Holland.
- Silverstein, M. (1996). Dynamics of recent linguistic contact. In: I. Goddard, ed., *Handbook of North American Indians 17: Languages*, 1st ed. Washington, DC: Smithsonian Institution, pp. 117–136.
- Suttles, W. ed. (1990). *Handbook of North American Indians: Northwest Coast*. Washington: Smithsonian Institution.
- Swanton, J.R. (1929). A sketch of the Atakapa language. *International Journal of American Linguistics*, 5(2/4), pp. 129–141.
- Thomas, G. (1698). *An historical and geographical account of the province and country of Pennsylvania; and of West-New-Jersey in America*. London: A. Baldwin. (Facsimile reprint: New York: For Henry Austin Brady by Francis Michelin, 1848).
- Thomason, S. (1983). Chinook Jargon in areal and historical context. *Language*, 59, pp. 820–870.
- Thomason, S. (1997). Mednyj Aleut. In: S. Thomason, ed., *Contact languages: A wider perspective*. Amsterdam: John Benjamins, pp. 449–468.
- Thomason, S. (2016). Language contact and change in the Americas: The state of the art. In: A. Berez-Kroeker, D. Hintz and C. Jany, eds., *Language contact and change in the America: Studies in honor of Marianne Mithun*, 1st ed. Amsterdam: John Benjamins, pp. 1–13.
- Thomason, S. and Kaufman, T. (1988). *Language contact, creolization, and genetic linguistics*. Berkeley: University of California Press.
- Voort, H. van der. (1996). Eskimo pidgin in West Greenland. In: E.H. Jahr and I. Broch, eds., *Language contact in the Arctic. Northern Pidgins and contact languages*, 1st ed. Berlin: De Gruyter Mouton, pp. 157–258.
- Werner, O. (1963). *A typological comparison of four Trader Navajo Speakers*. PhD. Indiana University.
- Young, R. and Morgan, W. (1980). *The Navajo language: A grammar and colloquial dictionary*. Albuquerque: University of New Mexico.

Language contact in West Africa

Friederike Lüpke and Rachel Watson

1. Introduction

Language contact is the impact left by heteroglossic and multilingual language use on a conventionalized language system. Languages are socio-political constructs with varying relations to language use, because of its inherent variation. We conceptualize variation as heteroglossia, here understood as heterogeneity within what is construed as a single language, and as multilingualism, describing language use perceived as involving more than one language. Multilingualism is ubiquitous in (West) Africa, to the extent that it can be considered the unmarked sociocultural condition of communication (Fardon and Furniss, 1994; Gafaranga and Torras, 2016; Lüpke and Chambers, 2010; Myers-Scotton, 1993). Hence, contact is omnipresent and has particularly salient manifestations in explicit contact languages such as the Portuguese- and English-based creole languages spoken on its coasts and mixed varieties such as Urban Wolof. Yet, no linguistic setting in this area is exempt from the influences of long-standing mobility, intermingling, trade networks and colonization. How these factors have shaped languages and language use in this particular geographical area is the focus of this chapter.

The geographical term West Africa is employed as a term of mere convenience, loosely interpreted as corresponding to the Western African portions of long-observed fragmentation and convergence areas (see Section 2). This area spans the Sahel region in the north and is delimited by the Atlantic Ocean in its west and south. There are no externally motivated clear-cut territorial boundaries other than the borders of colonially created nation states. Within the region and beyond it, several meaningful socio-political spaces have existed through history (for a discussion of the Mandé and Atlantic spaces, see Lüpke (forthcoming b), often coinciding with empires, states and larger polities. None of these had clear demarcations, but all tapered out and overlapped with the peripheries of other powers (Canós-Donnay, 2019). The area hosts great linguistic diversity, with hundreds of local and regional languages co-existing with languages of European provenance serving today as the official languages of West African states which have been appropriated in various ways by its inhabitants. Urban multilingualism and language contact originating in the (pre)colonial trade posts and in modern postcolonial cities is relatively well-studied (Auzeanneau, Bento and Leclère, 2016; Barry, 1990; Biagui, Nunez and Quint, forthcoming; Calvet, 1993; Cruise O'Brien, 1998; Dreyfus and Juillard,

2004; Intumbo, Inverno and Holm, 2013; Jacobs, 2010; Juillard, 1995; Mc Laughlin, 2001, 2008; Ndecky, 2011; Nunez, 2015; Swigart, 1994); much less known is the widespread rural multilingualism in the region (Lüpke, forthcoming b; Lüpke and Chambers, 2010) and in Africa as a whole (Di Carlo, 2018; Good, Di Carlo and Ojong, 2019), and the interaction between rural and urban sites, resulting in different but connected planes of language use. Sites and their connections, historical developments, language ideologies and territorialization models and linguistic identity formations are surveyed in Section 2.

Section 3 addresses critical issues arising from the study of multilingualism and language contact in a geographical area in which standardized language varieties play only a minor, mainly symbolic, role in communication and in which language history remains largely unknown. Standard languages commonly serve as the reference points to measure multilingualism (in terms of repertoires captured in terms of named languages) and to describe multilingual speech through allowing the identification of contributing codes. Where standard languages are not much used in linguistic interaction, as is the case in many West African settings (Canut, 2010; Donaldson, 2018; Lüpke, forthcoming c), language use is fluid in many contexts and thus defies easy categorization.

Because of the concentration of much historical linguistic research on genealogical relatedness, to the exclusion of language contact, the diachronic development of West African languages is an area of great debate. The morpho-phonological characteristics of many West African languages, in particular Atlantic languages, many of which feature stem-initial consonant mutation, make the historical method inapplicable (Pozdniakov, 2007). Additionally, monoparental models of relatedness such as the genealogical model can only account for a limited proportion of language change in an area where speakers of closely related languages remain multilingual and/or intermingle continuously over long time-spans. It is often impossible to distinguish features inherited from an ancestor from features diffused through language contact, in particular because features arising from diffusion are also passed down genealogically once established (François, 2017) and vice versa. Genealogical groupings beyond the level of family remain contested (Childs, forthcoming; Dimmendaal and Storch, 2016; Pozdniakov and Segerer, forthcoming). Newer network methods modelling both vertical and horizontal relatedness have been used (Pozdniakov and Segerer, 2010), but not found reflections in publications yet.

Despite the availability of languages of wider communication for at least the last millennium, linguistic diversity has been upheld in West Africa. Language evolution and change must therefore be conceptualized along the axes of convergence and divergence. Convergence results in contact languages and in contact features touching all areas of language, while divergence is manifest in the maintenance of separate languages and of small emblematic differences. While convergences testify to the homogenizing forces of language processing demands and of communicative needs, divergences counteract them because they enregister features that express socially meaningful facets of identity (Silverstein, 2003; Agha, 2005; Eckert, 2008). We therefore present a new epistemological focus in Section 4 that moves interactional features and social-indexical motivations to centre stage (Matras, 2009, 2012). Such recentring recognizes that language use (which we also call *linguaging*, see Li Wei, 2018 for a history of the term and of the related notion of *translinguaging*) is dynamic and variable while at the same time containing a solid core that is reified in the mind of users and analysts. The social and cognitive motivations underlying this reification that operate according to different parameters from standardization, and the features indexing social difference, move to centre stage in this approach. The framework we suggest relies on examining language use and meta-linguistic judgements to determine the prototypicality of different features in both individual and collective conceptualizations of language. We formalize these different viewpoints on

linguistic matter by introducing the notions of scale and perspective (Gal, 2016; Irvine, 2016), and discuss how the perspectives of different actors – speakers/signers/writers, hearers/readers, and analysts – can contrast or coalesce.

In Section 5, we address methodological advances in the study of West African language contact, focusing on the combination of multiple perspectives through working in multidisciplinary teams, complementing corpus-based research with sociolinguistic methods and linguistic anthropology, the adoption of genre-differentiated analyses and the inclusion of the spoken, signed, and written modality, which all constitute promising avenues for future research.

2. Historical overview

2.1 West Africa as an area of linguistic fragmentation and diffusion

The simultaneous existence of a high number of small languages and of far-reaching multilingualism and intense language contact can be captured in the area being classified as a zone of fragmentation and spread at the same time. On the one hand, West Africa is part of the ‘Sub-Saharan fragmentation belt’ (Dalby, 1970), a zone of great linguistic diversity, and on the other hand it is characterized by significant yet uneven spreads of linguistic features formalized in the Macro-Sudan belt (Güldemann, 2008, 2010; Güldemann and Hammarström, forthcoming). The Macro-Sudan belt is based on the (partial) presence of logophoric pronouns, labiovelar consonants and labial flaps, ATR-based vowel harmony, and the word orders S-(AUX)-O-V-X and V-O-NEG. Phonological features single out a largely overlapping spread zone called the Sudanic belt, based on the common occurrence of labial flaps, labial-velar stops, implosives, phonemic nasal vowels, ATR-based vowel harmony, complex tone systems and ‘lax’ prosodic marking of questions (Clements and Rialland, 2008). These divergence and convergence areas span grass and tree savannah and dry forests, with some areas of rain forest on the coastal and southern fringes. While north-south spread of features is attested and mainly associated with long-distance trade, the major direction of feature spread is latitudinal, as are most migratory movements, in line with the universal observation that features have a greater potential to spread laterally and are more constrained in their longitudinal diffusion (Diamond, 1999; Güldemann and Hammarström, forthcoming). Similar climatic and ecological environments aligned along latitudinal bands facilitate migration, cohabitation and exchange, and hence language contact is produced by social interaction in particular environments (Bentz et al., 2018; Mufwene, 2001), reflected in the orientation of the convergence areas.

2.2 Different West African language ecologies

Since particular environmental and social factors give rise to site-specific patterns of multilingualism and language contact, these settings are becoming a focus of research in their own right. This endeavour is particularly relevant in the light of the call for globalizing sociolinguistics (Childs, Good and Mitchell, 2014; Smakman and Heinrich, 2015; Di Carlo, Agwara, and Ojong, forthcoming) and language documentation and description (Di Carlo, 2016; Di Carlo and Good, 2017; Lüpke, 2017; Mufwene, 2017) so as to shed light on the small-scale societies in which most of language contact takes place (Trudgill, 2011; Lüpke, 2016b). Understanding how verifiable migration and settlement patterns, social exchange and mobility shape language practice and language ideologies is an indispensable prerequisite for the interpretation of divergence and convergence patterns with greater time-depths and larger geographical distribution, for which no detailed historical knowledge and no linguistic records are available.

Throughout West Africa, small-scale rural language ecologies in Frontier societies (Kopytoff, 1987), widespread at the fringes of larger states, are noteworthy in this regard. In these societies, small family-based groups (often patrilineal) have moved and settled over long periods of time, regrouping and intermingling with near-by clans and constantly engaged in social exchanges including intermarriage. At the same time, they maintain(ed) discrete social identities that can also be indexed by language and allow the forging of maximal, yet flexible, alliances. These ecologies are characterized by very high levels of multilingualism, with three to ten languages comprising typical individual repertoires, whose named languages in turn comprise multiple lects. Only very few of these small-scale ecologies have been studied in depth. Among the case studies are a number of geographically close, but sociolinguistically distinct settings in the Lower Casamance in Southern Senegal (Cobbinah, 2010; Lüpke, 2010; Cobbinah et al., 2016; Goodchild, 2016; Goodchild and Weidl, 2018; Watson, 2018; Weidl, 2019; Goodchild, 2019, forthcoming); Cobbinah, forthcoming), and the Lower Fungom area in Northwestern Cameroon (Good, 2013; Di Carlo and Good, 2014; Di Carlo, Agwara and Ojong, forthcoming). All these areas are characterized by topographical features that favour the existence of small, village and ward-based groups, such as marshlands and river deltas (in the Casamance case) or mountainous terrain (in Cameroon).

In larger polities, less linguistic heterogeneity is built into social organization, since endogamous marriage and polity-internal exchange becomes more widespread. One such polity in the Lower Casamance, where its location on a peninsula has resulted in one of the larger political formations in the region, the small kingdom of Mof Avvi, has been described in detail (Goodchild, 2019; forthcoming (2019); Sagna, forthcoming). Larger socio-political spaces often associated with state formations (and situated in regions that climatically and geographically facilitate easy movement of larger groups and armies, such as the West African savannah) also exhibit linguistic diversity on a smaller scale, as discussed for the contrasting Atlantic and Mande spaces of West Africa (Lüpke, forthcoming b; Lüpke et al., under revision).

Yet, all these settings were or are linked to larger exchange patterns through seasonal or permanent migration, pilgrimage, trade, and external and internal slavery. The focal points of much, though not all, migration, has been to urban centres. Precolonial cities have existed throughout the known history of the area (Coquery-Vidrovitch, 2005). For some, we have evidence for the patterns of multilingualism practiced in them through the written language contact exhibited in manuscripts in the Arabic script produced in urban centres of scholarship such as Djenné, Timbuktu, and many others (Ngom, 2017). The co-existence of Arabic-script writing in Arabic and in languages other than Arabic (so-called Ajami writing) in these manuscripts is subject of a growing body of research (Tamari and Bondarev, 2013). Other urban centres originated as trading and slaving posts of Iberian and later other European settlers on the West African coast. In the wake of their foundation and the formation of a broker class of intermediaries, intermarrying with the settlers and participating in the slave trade, a Portuguese lexifier Creole (Kriolu) emerged on the Cape Verde archipelago from where it was transported to the Upper Guinea Coast mainland in the sixteenth century (Jacobs, 2010), where it now has become the major lingua franca of Guinea Bissau and an important vernacular and vehicular language of the adjacent Casamance region of Senegal (Biagui, Nunez and Quint, forthcoming; Intumbo, Inverno and Holm, 2013). Other, English-lexified, contact languages are Nigerian Pidgin English (Faraclas, 2013) or *Naija* and *Krio* or *Aku*, spoken in Sierra Leone (Finney, 2013). A number of originally urban varieties such as ‘Urban Wolof’ in Senegal which are sometimes described as creoles (Swigart, 1994) can also be attributed to the existence of local intermediaries between colonists and indigenous populations. Their speech exhibited many lexical influences from French or English which now have become

a widespread and unmarked way of speaking in many settings (cf. Nouchi in Côte d'Ivoire (Boutin and N'Guessan, 2015), Camfranglais in Cameroon (Chia, 1990; Kießling, 2005, etc.). The excolonial European languages of West Africa are distinct from their metropolitan varieties and exhibit great fluidity (Manessy, 1994; Chaudenson and Calvet, 2001; Gut, 2015). Many speakers are multilingual in the mixed registers and their component codes, which are all highly heteroglossic, giving rise to forms that can be seen as monolingual, code-switched, fused or ambivalent, depending on perspective. Rural and urban settings were always connected, but rapidly increasing urbanization and the availability of new media to facilitate oral and performative written communication over large distances intensify and transform multilingualism patterns, and their influences on language contact remain to be researched in more detail (Lexander, 2010; McLaughlin, 2014; Lexander and Alcón, forthcoming).

3. Critical issues and topics

3.1 *West African ideas of language and of language territorialization*

Strong ideological connections between language(s) and place exist in West Africa. However, contrary to the European model of the ethnic nation state, with language as the central marker of ethnicity (Fishman, 1989), most West African settlements are not based on the idea of an ideally homogeneous language community.

Widespread firstcomer-latecomer or landlord-stranger settlement patterns (Brooks, 1993; Lentz, 2013) accord a special status to the founders of a place and their (mostly male) descendants, and it is the founder's or ruler's language that becomes associated with the place, a deictic value that is likely mirrored in the Founder principle formulated for creole studies (Mufwene, 1996) that ascribes far-reaching influences on its linguistic organization to the founders of a settlement. Founder languages have been called patrimonial languages by Lüpke (2018) to draw attention to the fact that these languages only correspond to a fraction of the linguistic repertoire of the inhabitants of a place; that they are not expected to be shared by all of them, and the contexts in which they are exclusively used are often limited to particular ceremonial contexts related to ancestors, land and land rights – the patrimony. Such an imagination of language contrasts with an ancestral territorialization model (Blommaert, 2010; Woodbury, 2005), more akin to the language ideologies of the nation state, in which only one language has a right of residence, and linguistic assimilation of all who are deviant from it is desired. Ancestral language ideologies are linked to a categorical view of linguistic identity, where one and only one language occupies a special position in a given space and is deemed to express its inhabitants' identity in an absolute fashion.

Ethnicity is a concept introduced to (West) Africa in colonial times (Ranger, 1983; Amselle and M'Bokolo, 1985; Amselle, 1990; Lentz, 1995). Ethnic labels, encapsulating outsiders' views, are applied over internally heterogeneous linguistic spaces in which local differences signal important local identities. Ethnic identities have gained social reality in contemporary societies, but without yet finding a direct linguistic equivalent. However, ethnic affinity among speakers of related languages invites the categorization of these languages, which can be named based on more locally confined features, as being instantiations of one superordinate ethnic language at a larger scale and may over time result in convergence (see Bazin, 1985 for a Malian example and Lüpke, 2016a for a case study in Senegal).

Relational, rather than categorical, identities are widespread in West African contexts (Good, Di Carlo and Ojong, 2019). Both in terms of ethnicity and in terms of language use, they enable an individual to claim and perform multiple and variable identities dependent on their trajectories, repertoires and situational needs. It is difficult to formalize West Africans'

adaptive repertoires, especially since categories such as mother tongue, first and second language, even when assumed as plural categories, in many cases do not capture language acquisition and socialization (Cissé, forthcoming) of the region's mobile inhabitants (see the snapshots of GS's repertoire in Figures 26.1 and 26.2 for an illustration). Patrimonial languages are often mistaken to correspond to mother tongues or L1s, but their function is to link individuals to ancestral places of origin where they have land rights according to historically anchored mobility patterns (Brooks, 1993; Lentz, 2013), and individuals may not speak them at all. In particular, we find that a triglossic picture, with excolonial and now official languages at the top, languages of wider communication in the middle, and locally confined languages at the bottom of a hierarchically organized pyramid, does not fully explain situated and relational language use patterns. We consider the terms 'vehicular' and 'vernacular' useful, in addition to the notions of patrimonial language, not to categorically delimit a language's role in a society, but to discuss the different and changeable roles and differential affective values of named languages in an individual's repertoire. The role of a language is multidimensional, dependent on factors such as patrimonial, other ancestral and relational identity links, frequency of use, attitudes, and access to conventionalized norms, such that the same named language may play quite a different role in different people's repertoires; moreover that role is dynamic and may change over time and even in different situations. Repertoire studies make use of language portraits (Busch, 2006) and other, individual-specific profiles resulting from in-depth ethnographic fieldwork. While these allow gauging the complexity of repertoires and the speaker-individual circumstances shaping them, they do not lend themselves to generalizations or comparisons. Therefore, we here suggest a frame built on the factors mentioned previously that contribute to the role that a language plays in an individual's repertoire.

We visualize these factors as contributing to a cumulative whole which represents the role of the language in an individual's repertoire at a given point in time. The relative value of each factor is represented by the size of the corresponding section in the bar; the relative size of the entire bar indicates whether the language plays a more vernacular (longer bar) or vehicular (shorter bar) role in the repertoire of the individual. The visualizations underscore the fact that the different factors are not necessarily congruent. A language may have a high value for, say, patrimonial identity, but not for frequency of use or access to conventionalized norms. The values can be adjusted – both increased and decreased – over an individual's lifetime, or even in different spatial or situational contexts. Figures 26.1 and 26.2 illustrate how these functions can cluster differently at different times (in 1995 and in 2015) in the life of GS, a man in his 30ies from Casamance, Senegal. The scores are based on ethnographic research and sociolinguistic interviews, and are assigned on a heuristic and relational, rather than an absolute, basis.

A more complete vignette on GS can be found in Lüpke et al. (under revision). In 1995, he was living in Essyl, a village whose patrimonial language is Joola Banjal. The patrimonial language of GS' father's village is Baïnouk Gubëeher, which he only learned when he moved to live with his father in Djibonker. Wolof is the de facto national language of Senegal, while French is its official language. Bayot is a macro-language associated with Niassia, the village where he spent his early childhood.

Three languages have been added by 2015, at which time GS lives in his father's village Djibonker and has had occasion to learn and speak its patrimonial language Baïnouk Gubëeher. The village next door is associated with Joola Kujireray, closely related to Joola Banjal, which he has also learned. Additionally, he has picked up Kriolu by interacting with palm wine tappers from neighbouring Guinea Bissau. The snapshots on these figures are intended to serve as a heuristic representation only, to complement and not to replace detailed descriptions of repertoires (Goodchild, 2019; Weidl, 2019). They illustrate the great size (both in terms of

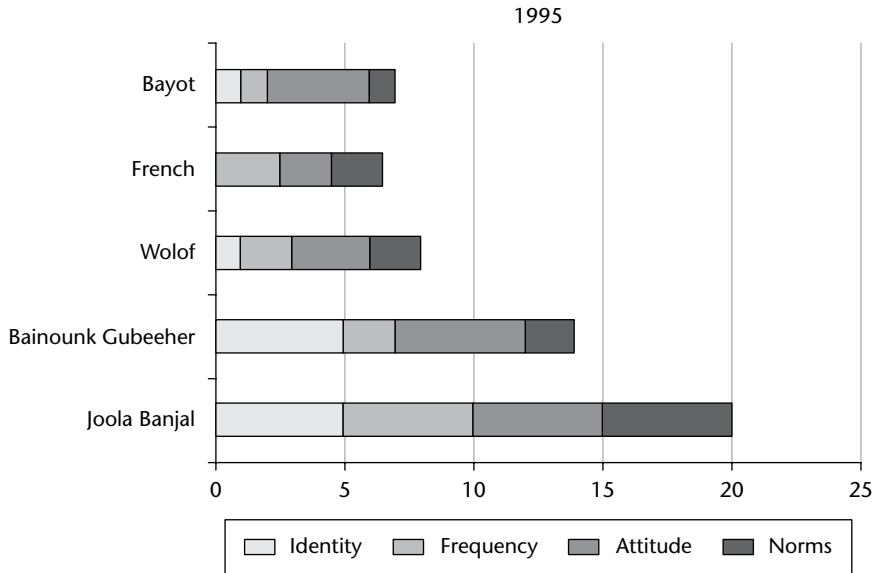


Figure 26.1 Roles of named languages in GS's repertoires in 1995

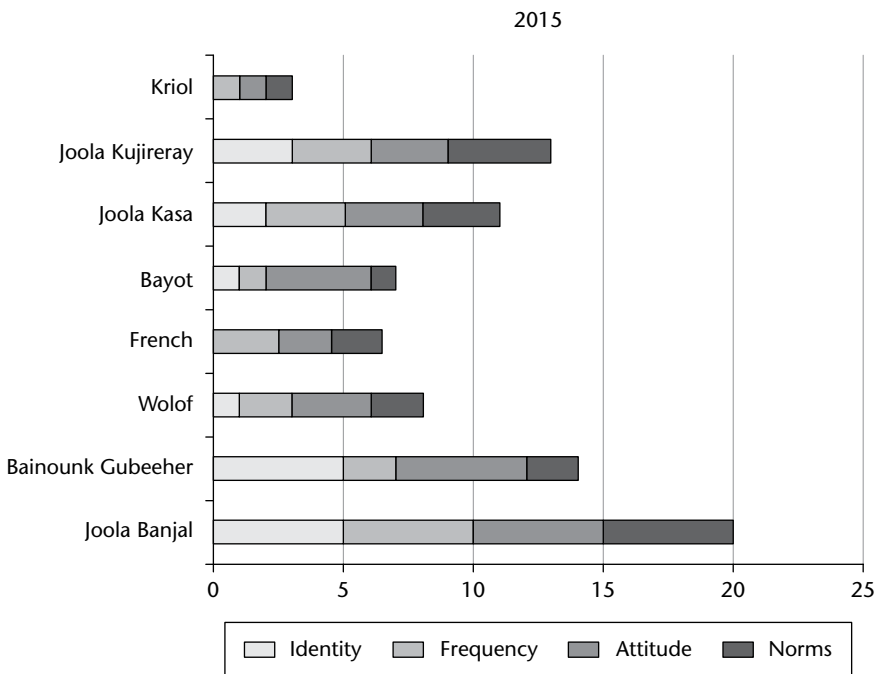


Figure 26.2 Roles of named languages in GS's repertoires in 2015

un- and remotely related languages and closely related lects) and the enormous variability and adaptability of repertoires, induced by small-scale mobility (the two places of residence are ca. 10 km apart from each other) and changing social networks throughout an individual's life. Further research is required to refine and formalize such visualizations, which would ideally include information on adherence to linguistic norms and frequency mined from corpus data – although see Section 4.1 for discussion of obstacles to be overcome in this respect.

3.2 Emblematicity, scale, and perspective

Languages are bound to exhibit similarities at a structural level, through close genetic inheritance and/or intense and ongoing contact. They can also converge through indexing overarching 'macro-identities' such as those created by the introduction of ethnic categories. In the case of closely related languages, small differences become significant in distinguishing differential social meaning. Social identities are multifaceted, and considerations of scale and perspective come into play as different identities crosscut or nest within each other. These factors lead to a structurally and socially multi-layered contact situation, which results in complex patterns of convergence and divergence.

Even a single utterance can contain social meanings on different levels. This is illustrated concisely in formulaic greetings in Joola (see also Hantgan, 2017; Lüpke et al., under revision). The Joola languages belong to the Bak branch of Atlantic, spoken from the Gambia to Guinea Bissau. Their genetic relatedness is relatively uncontroversial, although internal classification is confounded by long and extended contact. In Joola languages, greetings include the formulaic couplet translatable as '[Is there] peace?', 'Peace, only.' Such a formula is widespread, though not ubiquitous, in the region. Many other languages use equivalent, although not phonologically related, expressions in their greetings, which can be regarded as instances of pattern convergence. Table 26.1 shows the typical response form – 'peace only' according to a broad sub-regional distribution. The form *KASUMAY keb* is associated with Joola Fogy, the dominant Joola variety and thus has rather universal usage throughout Casamance. *KASUMAY* is shown in caps to indicate that this is a schematic representation of a form that has many different phonetic realizations across the region.

The convergent part of the formula facilitates communication, as the form is universally understood throughout the region. It also allows the indexation of a broad, pan-Joola identity. The second part, 'only,' also allows social indexation on a more fine-grained level, showing a connection with a particular sub-region. Zooming in even further, we also find levels of distinction maintained within these sub-areas, which can be indexed through differences in pronunciation of *KASUMAY*. For example, it has been shown that a contrast in word initial [g] and [k], and other small differences in pronunciation in the form *KASUMAY*, can be used to index differential linguistic identities at the village level even when the pan-Joola (i.e., *KASUMAY*), and sub-regional (i.e., 'only') identity indexed are identical (Sapir, 1975; Hantgan, 2017;

Table 26.1 Joola greetings: 'peace only'

| Form: 'peace only' | Region |
|----------------------|--|
| <i>KASUMAY keb</i> | North of Casamance river (most widespread) |
| <i>KASUMAY bare</i> | South of Casamance river |
| <i>KASUMAY lamba</i> | North-western part of Casamance |

Watson, 2019). Some features may take on special indexical power and become emblematic of a particular identity. A form can be emblematic at different levels of specificity, with an inverse relation between the reach of a form and the specificity of the identity it can index. The indexical power of a linguistic feature is also subject to effects of perspective, as it is dependent on contrast. It is only a feature that is salient from a number of different viewpoints that has the potential to be emblematic.

The following section discusses the issues of scale, perspective and emblematicity through identifying convergent and divergent features in a text that can be differently categorized as heteroglossic and/or multilingual.

4. Current contributions and research

4.1 Rethinking reification

Reification of language is a universal practice driven by the human propensity to categorize. Languages may be reified in different, sometimes conflicting ways by actors, according to contrasting perspectives and motivations. Speakers of languages reify them to express social-indexical requirements, reinforcing identity and group membership, facilitating maintenance of many small languages, and perpetuating prescriptive ideals and myths of ‘correctness.’ Descriptive linguists reify languages in word lists and grammars in a process of artefactualization which may also carry ideological baggage prioritizing monolingual language situations, standardization and nostalgic notions of an ancestral code. Reifications result in erasure of variation, and often background multilingualism, obscuring facts that may allow us to investigate language contact. In addition, effects of scale and perspective can mean that numerous reifications may exist that are seemingly contradictory. It is increasingly recognized that descriptive linguists must find a more dynamic way to describe languages that can take into account variation, multilingualism, and different levels of scale and perspective (Childs, Good and Mitchell, 2014; Marten and Petzell, 2016; Watson, 2019; Lüpke, forthcoming b), with reifications treated as cognitive phenomena rather than empirical facts.

Nevertheless, some form of description is a prerequisite if we wish to analyze multilingual discourse, mine multilingual corpora, and examine contact effects. At present such analysis is hampered by the fact that it can be impossible to identify the languages involved in such data in a principled way, particularly, but not only, if the languages are closely related. If languages are un- or underdescribed, spoken by highly multilingual speakers in dynamic contact situations, we have no way of knowing which features come from where. Speakers’ categorizations are often based on non-linguistic features such as the known identity of discourse participants or the setting, in addition to the shape of their repertoires. Dependence on descriptions which present a single reification of a language can mean that we risk misinterpreting the data; for example, mistaking heteroglossia or borrowing for code-switching. Watson (2019) proposes conceptualizing, and thus describing, languages based on principles of prototype theory (Rosch, 1975; Rosch and Mervis, 1975) in order to accommodate and account for variation, and to provide flexible, multiperspectival, reference points for the analysis of multilingual discourse and corpora. A language can be modelled as a collection of features that are more or less prototypical of that language, and prototypicality is determined empirically from a wide variety of data from different speakers, according to factors including consensus, frequency and stability. Languages can then be compared to each other to identify commonalities, with salient differences thus foregrounding the relationships between languages in contact and to allow an examination of contact effects from a fine-grained, diachronic perspective, as well

as providing a more reliable foundation for historical studies. The model is compatible with sociolinguistic and lexico-grammatical views of language as collections of features associated with sociocultural constructions that are named languages (Mufwene, 2001; Cheshire et al., 2011; Jørgensen et al., 2011; Blommaert and Rampton, 2012; François, 2014). It is commensurate with work on social networks (Milroy, 1980) which predicts that a stable core is maintained by central conservative speakers, with innovations introduced by social outliers also described for West African settings (Beyer, 2010; Beyer and Schreiber, 2013).

Describing languages in this way requires a systematic methodology based on a holistic research paradigm combining methods from theoretical linguistics and sociolinguistics in novel and properly defined ways. A linguistic description constitutes a model that necessarily must be built empirically, based on large volumes of linguistic data from numerous different speakers, genres, and contexts, correlated with rich sociolinguistic data on all speakers and contexts. Tolerance for variation means the only condition on the data used to create this model is that it is a speaker's ascertainable pragmatic intention to speak the language in question. The creation of a prototypicality model therefore begins with elicitation and staged communicative events, since in more naturalistic data it is often not possible to ascertain which language a speaker intends to speak. In addition, token frequency even for high-frequent lexeme types in highly multilingual corpora is not sufficient to systematically establish variational ranges.

The prototypical form of any given feature (from phonetic to syntactic) can be found through examination and aggregation of numerous exemplars from this data, which are extracted and collated in a data base, which also gives us detailed information about the utterance. This is illustrated in Table 26.2 for the form for 'water' as just discussed (the table is illustrative only – in practice, many more tokens would be required for a reliable description).

An aggregation of these results shows that there is a general consensus for the form [muhem]. This is the prototypical realization of this form in Joola Kujireray. We also come across alternative forms. These are not discarded but rather can be quantitatively and qualitatively analyzed, with the help of the rich sociolinguistic data which is an integral part of the descriptive methodology under this approach. Any form can be cross-referenced with data about the speaker's linguistic repertoire or the circumstances of the conversation. We may find that, for example, JT's deviation from the prototype is due to the prominence of Joola Banjäl in his repertoire, where [mal] is prototypical. We can discover whether GS's deviation from the prototype in token (n) is to do with influence from his own repertoire, or that of his interlocutors, while in the more controlled context of elicitation, he successfully accesses the prototypical form. This approach enables the detailed observation of variation to investigate the dichotomy between languages as imaginary reifications and actual language use.

Table 26.2 Forms for 'water' in Joola Kujireray

| <i>Token No.</i> | <i>Form</i> | <i>Speaker</i> | <i>Context</i> | <i>Interlocutors</i> |
|------------------|-------------|----------------|---------------------|----------------------|
| 1 | [muhem] | LM | elicitation | researcher |
| 2 | [muhem] | UB | staged conversation | LM |
| 3 | [muhem] | GS | staged conversation | LM, UB |
| 4 | [mal] | JT | elicitation | researcher |
| 5 | [muhem] | GS | elicitation | researcher |
| ... | | | | |
| n | [mal] | GS | staged conversation | JT, UB |

Furthermore, variation can be motivated according to repertoires and contexts, providing a more nuanced view of contact effects. Once comparable databases are built up for languages across a contact area, this allows researchers to interpret thickly multilingual data of the type in example (1) more meaningfully. They can also be used to carry out frequency searches to discover in which circumstances different variants occur, where borrowings and code-switching are likely to occur, and investigate emerging contact effects (see Adamou, 2016; Torres Cacoullos and Travis, 2018). Additionally, these analyses can then feed back into the prototype description in a virtuous research cycle.

4.2 Identifying and analyzing convergence and divergence in multilingual discourse

We have seen that complex patterns of convergence and divergence are found in multilingual language ecologies, that languages can be conceived as collections of features and that social difference can be indexed through large and small linguistic differences. In the following we present a short piece of discourse that illustrates how speakers use their linguistic repertoires in a fluid way, and discuss the implications that data of this sort has for an understanding of contact in such settings.

The extract is taken from a conversation between five men, in the village of Brin, Lower Casamance. The following conventions are used in the notation: Joola Kasa is in bold; material common to more than one Joola language, including Kasa, is in plain italics; material common to more than one Joola language, excluding Kasa, is in underlined italics.

1. Conversation at the seminary

PB *nu-has-om*

2S-know-1S

‘Do you know me?’

LM *ah inti huken ni-juk-i kajak-uya ni-jok-ut*

ah CONJ yesterday 1S-saw-2S name-2S.POSS 1S-grab-NEG

‘Ah yesterday I saw you, it’s your name that escapes me’

LM *kajak-oli buu?*

name-2S.POSS how

‘What’s your name?’

PB: *kajaw-om pier pier betraj*

Name-1S.POSS Pierre Pierre Bertrand

‘My name’s Pierre, Pierre Bertrand’

The extract was transcribed by LM, a local, multilingual resident, who also happens to be one of the participants in the conversation. He segmented the discourse and tagged each segment for language. Nearly every segment was tagged as Joola Kasa, a Joola language patrimonial to a group of villages in the southern zone, and with relatively extensive reach as a language of wider communication, with some French, and some mixed French-Joola Kasa segments. Kasa itself is an intermediate-level category integrating more locally confined sub-varieties, and in this case, the categorization is based on an outsider’s view and an overarching linguistic identity commonly presented to outsiders. In subsequent discussion, LM confirmed that the conversation had taken place in Joola Kasa, and that the reason for this choice of Kasa among several shared languages was that the majority of the men present counted Joola Kasa as their most prominent vernacular language, including an important shrine holder with high

social status. LM and PB were the only participants for whom Joola Kasa does not play a highly vernacular role, but for reasons of communicative expedience and respect, they aligned their language to their interlocutors.

However, from the perspective of the descriptive linguist, a different picture appears with respect to the prototypicality of the features in the discourse. First, much of the material is common to more than one Joola language, including Kasa (plain italics). There are indeed some features that can be identified as prototypically Joola Kasa (bold). Notably, however, given the transcriber's judgement, there are several features that are *not* prototypical of Joola Kasa, but of other Joola languages (underlined italics). These latter features are ambiguous if looked at from a purely structural perspective, as they are found in several different Joola languages, including Joola Kujireray, Joola Banjal and Joola Fogy.

In order to interpret the data meaningfully we need to examine the intersection between the speakers' and analysts' perspectives. It was LM and PB's pragmatic intention to align with their interlocutors and speak Kasa, and they achieved this by using features they knew to be prototypically Kasa. However, since they do not have full access to the conventionalized forms of this language (it plays a more vehicular role in their repertoires) there is necessarily a process of bricolage, making do and filling in the gaps with material from other languages in their repertoire – such a process is recognized by speakers. It also follows, therefore, that an examination of LM and PB's repertoires may help to disambiguate the source of the 'rogue' features in the discourse. While the data in the preceding examples are based on many years of fieldwork, participant observation, continued discussion with numerous speakers and language experts, the presentation of features as prototypical is still heuristic. Furthermore, the features discussed are some of the more prominent and agreed upon in the respective languages; a meaningful examination of features that are the subject of variation and disagreement requires a more thorough methodology, as described in 4.1. Scholarly works consulted in the analysis include Sapir (1965), Sambou (1979), Bassène (2007) and Sagna (2008).

This exchange is typical in its complexity, but unique, as is every interaction, in its particular constellation of speakers, repertoires and social and communicative requirements. Such intense mixing often makes it difficult to establish which language(s) are psycholinguistically dominant in discourse (van Coetsem, 2000; Winford, 2007). We therefore need to ask to what extent the heterogeneity of multilingual interaction precludes and/or shapes the sedimentation of contact effects (see Watson, 2018).

Where contact effects are observed, they are rarely uniform or unidirectional, reflecting the multiplex, constantly shifting relationships between populations (McMahon et al., 2005). Borrowing from one language to another can be affected by factors including geographical proximity, political dominance relations between communities, and traditions of marriage exchange or agricultural cooperation, all of which can have opposed effects simultaneously as well as changing over time. This results in different strata of convergence and divergence, which can be analyzed in a fine-grained, qualitative fashion with attention paid to semantic domains as well as socio-historical context.

For example, in Casamance it is overall far more common for Joola forms to be borrowed into Bainouk languages than vice versa, due to the recent dominance of Joola populations in the region (Pozdniakov and Segerer, forthcoming). However, Watson (2018) shows that in one particular situation the reverse pattern is observed. The lexicon of Joola Kujireray contains a surprisingly high proportion of items apparently of Bainouk origin, suggesting an asymmetric dominance relation in the favour of Bainouk, contrary to the prevailing pattern in Casamance. This corroborates oral history accounts of Brin as a new settlement, possibly even a captive settlement which eventually integrated with the surrounding Bainouk populations.

Conversely, it is shown that Joola Kujireray and Joola Banjal which are closely related and spoken in close geographical proximity, maintain quite different forms for basic and common vocabulary items such as ‘water,’ ‘sleep,’ ‘cook,’ ‘name,’ and ‘fan palm.’ It is posited that the maintenance of small but prominent differences, such as these frequent lexical items, as well as the phonetic differences described previously, allows speakers to index their respective identities. Lüpke (forthcoming b) discusses a number of morphosyntatic patterns in Casamance languages and their areal spread and shows that their diffusion cannot be interpreted without a knowledge of the micro-history of settlement and migration. McLaughlin (1997) reveals different historical layers of borrowing and how they have reshaped the noun class system of Wolof, whose linguistic history is relatively well known. Yet, while it is straightforward to identify convergences, it is often impossible to determine their exact origin, except in a deep etymological sense, as also evident in Vydrine’s (1999) overview of lexical similarities between Mandinka and Joola Fogany. Additionally, it is not always clear whether a construction that etymologically belongs to a particular language reached another language via direct transfer or was diffused areally. ‘Wanderwörter’ such as *caabi* ‘key,’ ultimately from Portuguese or Kriol, are attested over wide areas, as are semantic calques such as nightly avoidance terms for ‘salt,’ ‘charcoal,’ ‘needle’ and ‘snake’ (Luffin, 2002; Minkailou, 2016; Biagui, Nunez and Quint, forthcoming).

Newer approaches to language contact and multilingualism are therefore agnostic to motivation (genealogical relatedness vs. language contact) and directionality of convergences but flag forms that share features of several named codes as exhibiting features of all of them (Nunez and Léglise, 2016). However, the perspectives of users and analysts in categorizing constructions as ‘belonging’ to one or several reified codes are not always identical as we have seen in example (1), since speakers,’ hearers’ and analysts’ repertoires and the codes associated within them constrain their base of comparison. Goodchild and Weidl (2018), Goodchild (2019), Weidl (2019) and Lüpke (forthcoming a) therefore build perspective into the analytical process by instrumentalizing rather than resolving difference in categorisation. Figure 26.3 illustrates different perspectives on the categorization of the lexical item *grawul* ‘it’s not a problem’, often analyzed as a mixed form combining a French stem and a Wolof suffix, based on the speaker’s repertoire:

The inclusion of perspective is not only crucial for an understanding of speakers’ possibilities and intentions and hearers’ possible interpretations; it is also indispensable for an analysis of multilingual speech and traces of language contact in it, particularly in the case of closely related lects reified as different socially meaningful entities but characterized by huge areas of feature overlap or ambiguity in lexico-grammatical terms.

5. Main research methods and future directions

Long driven by individual research, newer multilingualism and contact studies in West Africa are taking a collaborative and multidisciplinary turn. Such larger initiatives are required because of the complexity of settings, both in terms of size of repertoires and number of named languages, and in terms of the state of linguistic description, which is dire for most West African languages. Despite persistent and systemic challenges in terms of access to funding and intercontinental mobility and exchange for researchers based in Africa, there is a tendency for greater collaboration with West African researchers and a greater involvement of research participants in the research design, transcription, and analysis. A number of disciplinary perspectives that have not been joined in the past are combined, most notably in the complementation of descriptive and documentary approaches with sociolinguistic and anthropological linguistic

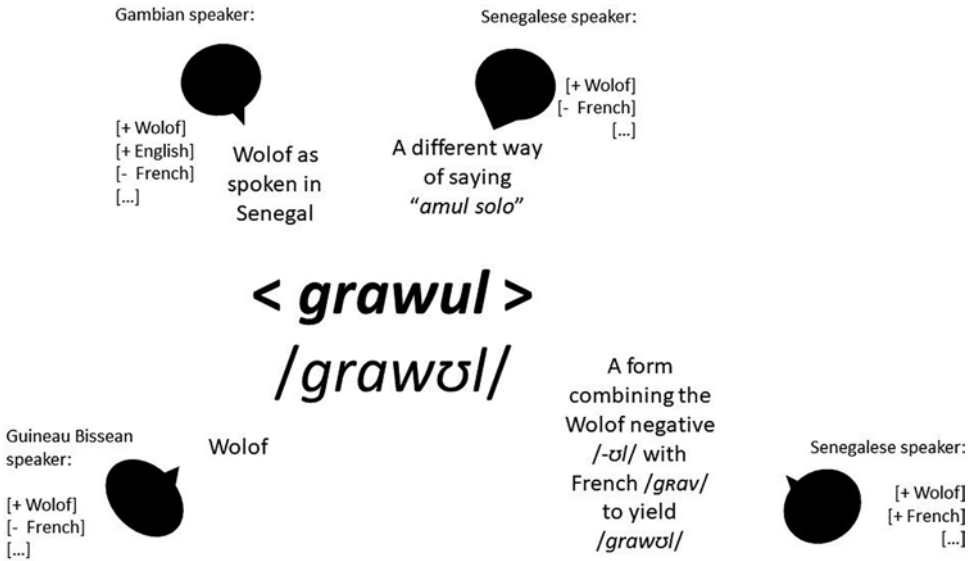


Figure 26.3 Influence of perspective on linguistic categorization

Source: Lüpke forthcoming a

viewpoints (Childs, Good and Mitchell, 2014; Meyerhoff and Stanford, 2015). The focus of many sociolinguistic investigations has been on urban settings, while language description and documentation, adopting a language-based outlook, tended to concentrate on rural settings. Language contact studies concentrated on language pairs exhibiting great genetic and typological contrasts, thereby sidestepping issues of variation and contact on a smaller scale and an investigation of settings in their full complexity. That focus is now shifting, and rural configurations and more multifaceted aspects of multilingualism are receiving more attention. New methods are being developed to create optimal workflows enabling international and interdisciplinary teams to transcribe, annotate, and share data (Watson, in preparation, 2017), to analyze multilingual corpora (Léglise and Alby, 2016), and to visualize socio-spatial patterns of multilingualism (in an ongoing project at the University at Buffalo).

Studies on language socialization and acquisition in West Africa’s multilingual settings remain rare but receive growing attention (Moore, 2008; Cissé, 2013, forthcoming). Psycholinguistic studies with bi- or multilingual West African speakers are non-existent, apart from a pioneering study of gesture- and speech alignment in multilingual speech in the Casamance region of Senegal (Krajcik, 2019). Findings in both areas challenge frameworks and findings of mainstream studies and thus have great potential to advance the understanding of language interaction in linguistic settings that are more representative for the evolution of language overall than the W.E.I.R.D – Western, educated, industrialized, rich, democratic – settings (Henrich, Heine and Norenzayan, 2010) that dominate language acquisition research and psycholinguistics.

Multimodality – the co-existence and interaction of speech with other linguistic modalities including gesture, sign language, and writing – emerges as another domain of great relevance to better capture the multidimensional nature of language use in West Africa and firmly establishes it as an area of investigation that can advance fields such as language planning and sign

language research. Regarding the former, multilingualism management in West Africa, through the ways in which communities have structured written communication in heteroglossic and multilingual communities and have created ways for lifelong informal language learning, is an important area for contact research that also has much inspiration to offer to multilingualism management and education (Souag, 2010; Lüpke et al., 2018; Lüpke, forthcoming c). Concerning sign language research, the existence of village-sign language communities consisting of Deaf signers and a large proportion of their hearing co-inhabitants (Nyst, 2010) is of similar relevance for Deaf studies and Deaf education.

Finally, how West African and European multilingualism and contact patterns are transformed through migration at the national and international scale (Juffermans and Tavares, 2017) and what its impact is on the maintenance and adaptation of multilingual repertoires, including their locally confined parts, is an important research area that can provide urgent insights for research on language endangerment and language vitality. West Africa's resilient multilingualism, despite the availability of languages of wider communication since the onset of early globalization, is remarkable and illustrates that language contact, change and shift do not operate uniformly but based on social functions of language in locally meaningful settings (Di Carlo and Good, 2017; Lüpke, 2017; Mufwene, 2017). Investigations of these settings are timely because insight into different patterns and planes of language contact and multilingualism in (West) Africa can transform epistemological frames and methods to free empirical research from a priori theorizations inapplicable to its ecologies.

6. Further reading

Beyer, K. and Schreiber, H. (2017). Social network approach in African sociolinguistics. *Oxford Research Enclopedias*. Oxford: Oxford University Press. doi:10.1093/acrefore/9780199384655.013.236.

Focusing on West Africa, where most of this research (by the authors) has been conducted, this article gives insight in methods and analyses of social network studies, linking the spread of linguistic features in multilingual populations to social interaction.

Good, J., Di Carlo, P. and Ojong, R. (2019). Multilingualism in rural Africa. In *Oxford Research Enclopedias*. Oxford: Oxford University Press. doi:10.1093/acrefore/9780199384655.013.227.

This article provides the first systematic overview of under-researched multilingualism in rural areas of Africa.

Juffermans, K. and Abdelhay, A. (2016). Literacy and multilingualism in Africa. In: B.V. Street and S. May, eds., *Literacies and language education. Encyclopaedia of language and education*, 1st ed. Berlin and Bonn: Springer.

This work presents a synthesis of research on written multilingualism and language contact on the African continent.

Lüpke, F. (forthcoming). Language contact in West Africa. *Oxford research enclopedias*.

This article offers an extended overview on the Atlantic and Mande spaces of West Africa, including a detailed historical account, as well as on epistemological and methodological issues.

Lüpke, F. and Chambers, M., eds. (2010). Multilingualism and language contact in West Africa: Towards a holistic perspective. *Journal of Language Contact THEMA*, 3.

This special issue offers a number of case studies on multilingualism and language contact in West Africa.

7. Related topics

Pidgins and creoles, social factors, processing multilingual data, usage-based approaches

Abbreviations

| | |
|-----------|---|
| 1 | first person |
| 2 | second person |
| AUX | auxiliary |
| CONJ | conjunction |
| NEG | negative |
| O | object |
| POSS | possessive |
| S | subject |
| V | verb |
| W.E.I.R.D | Western, educated, industrialized, rich, democratic |

References

- Adamou, E. (2016). *A corpus-driven approach to language contact: Endangered languages in a comparative perspective*. Boston and Berlin: De Gruyter Mouton.
- Agha, A. (2005). Voice, footing, enregisterment. *Journal of Linguistic Anthropology*, 15, pp. 38–59.
- Amselle, J.L. (1990). *Logiques métisses. Anthropologie de l'identité en Afrique et ailleurs*. Paris: Payot.
- Amselle, J.L. and M'Bokolo, E., eds. (1985). *Au cœur de l'ethnie: ethnies, tribalisme et état en Afrique*. Paris: La Découverte and Poche.
- Auzanneau, M., Bento, M. and Leclère, M. (2016). Introduction: Étude des pratiques plurilingues et défis contemporains. Hommage à Caroline Juillard, Professeure. In: M. Auzanneau, M. Bento and M. Leclère, eds., *Pratiques plurilingues, mobilités et éducation: éclairages d'Afrique ou d'ailleurs*, 1st ed. Paris: Édition des Archives Contemporaines, pp. i–viii.
- Barry, A. (1990). Étude du plurilinguisme au Mali: Le cas de Djenné. In: J. Kawada, ed., *Boucle du Niger: Approches multidisciplinaires*, 2 vols, 1st ed. Tokyo: Institut de Recherches sur les Langues et Cultures d'Asie et d'Afrique, pp. 183–207.
- Bassène, A.C. (2007). *Morphosyntaxe du joola banjal. Langue atlantique du Sénégal*. Köln: Rüdiger Köppe.
- Bazin, J. (1985). À chacun son bambara. In: J.L. Amselle and E. M'Bokolo, eds., *Au cœur de l'ethnie: ethnies, tribalisme et état en Afrique*, 1st ed. Paris: La Découverte and Poche, pp. 87–125.
- Bentz, C., Jager, G., Dediu, D. and Verkerk, A. (2018). The evolution of language families is shaped by the environment beyond neutral drift. *Nature Human Behaviour*, 2(11), pp. 816–821.
- Beyer, K. (2010). Language contact and change: A look at social factors in an African rural environment. In: F. Lüpke and M. Chambers, eds., *Multilingualism and language contact in West Africa: Towards a holistic perspective*, *Journal of Language Contact THEMA*, 3, pp. 131–152.
- Beyer, K. and Schreiber, H. (2013). Intermingling speech groups: Morphosyntactic outcomes of language contact in a linguistic area in Burkina Faso (West Africa). In: I. Leglise and C. Chamoreau, eds., *The interplay of variation and change in contact settings – Morphosyntactic studies*, 1st ed. Amsterdam and Philadelphia: John Benjamins.
- Biagui, N.B., Nunez, J.J.F. and Quint, N. (forthcoming). Casamance Creole. In: F. Lüpke, ed., *The Oxford guide to the Atlantic languages of West Africa*, 1st ed. Oxford: Oxford University Press.
- Blommaert, J. (2010). *The sociolinguistics of globalization*. Cambridge: Cambridge University Press.
- Blommaert, J. and Rampton, B. (2012). Language and superdiversity. *MMG Working Paper*, 12(9).
- Boutin, B.A.B. and N'Guessan, J.K. (2015). Le nouchi c'est notre créole en quelque sorte, qui est parlé par presque toute la Côte d'Ivoire. In: P. Blumenthal, ed., *Peter Blumenthal. Dynamique des français africains : entre le culturel et le linguistique*, 1st ed. Frankfurt: Peter Lang.
- Brooks, G.E. (1993). *Landlords and strangers: Ecology, society, and trade in Western Africa, 1000–1630*. Boulder, CO and Oxford: Westview Press.

- Busch, B. (2006). Language biographies: Approaches to multilingualism in education and linguistic research. *PRAESA Occasional Papers*, pp. 5–18.
- Calvet, L.J. (1993). *Les langues des marchés en Afrique*. Aix-en-Provence: Institut d'Études Créoles et Francophones.
- Canós-Donnay, S. (2019). The empire of Mali. *Oxford research encyclopedias*. doi:10.1093/acrefore/9780190277734.013.266.
- Canut, C. (2010). De l'Afrique des langues à l'Afrique des discours. Les voix du langagier. *Cahiers d'Études Africaines*, 198, 199–20(2–3–4), pp. 1163–1189.
- Chaudenson, R. and Calvet, L.J. (2001). Les langues dans l'espace francophone: de la coexistence au partenariat. *Langues et Développement*. Paris: Institut de la Francophonie, L'Harmattan.
- Cheshire, J., Kerswill, P., Fox, S. and Torgersen, E. (2011). Contact, the feature pool and the speech community: The emergence of multicultural London English. *Journal of Sociolinguistics*, 15(2), pp. 151–196.
- Chia, E. (1990). The new speech forms of rapidly growing city: Pidgin French and Camfranglais in Yaounde. *Annales de la Faculté des Lettres et Sciences Humaines, Université de Yaoundé*, 6, pp. 102–127.
- Childs, G.T. (forthcoming). Genetic clusters in Atlantic. In: F. Lüpke, ed., *The Oxford guide to the Atlantic languages of West Africa*, 1st ed. Oxford: Oxford University Press.
- Childs, G.T., Good, J. and Mitchell, A. (2014). Beyond the ancestral code: Towards a model for sociolinguistic language documentation. *Language Documentation and Conservation*, 8, pp. 168–191.
- Cissé, I. (2013). *Emergence of linguistic systems in interaction: Development of phonological and phonological skills in bilingual children (Bambara-Fulfulde), compared with monolingual children Fulfulde/Bambara*. PhD. Leiden University.
- Cissé, I. (forthcoming). Multilingual children's language socialization in central Mali. In: F. Lüpke, ed., *The Oxford guide to the Atlantic languages of West Africa*, 1st ed. Oxford: Oxford University Press.
- Clements, G.N. and Rialland, A. (2008). Africa as a phonological area. In: B. Heine and D. Nurse, eds., *A linguistic geography of Africa*, 1st ed. Cambridge: Cambridge University Press, pp. 36–85.
- Cobbinah, A.Y. (2010). The Casamance as an area of intense language contact: The case of Bainouk Gubaher. In: F. Lüpke and M. Chambers, eds., *Multilingualism and language contact in West Africa: Towards a holistic perspective: Journal of Language Contact THEMA*, 3, pp. 175–201.
- Cobbinah, A.Y. (forthcoming). An ecological approach to ethnic identity and language dynamics in a multilingual area (Lower Casamance, Senegal). In: J. Good and P. Di Carlo, eds., *African multilingualisms. Selected proceedings from the international conference, Yaounde 12 August 2017*. Lanham, MD: Lexington Books.
- Cobbinah, A.Y., Lüpke, F., Hantgan, A. and Watson, R. (2016). Carrefour des langues, carrefour des paradigmes. In: M. Auzeanneau, M. Bento, and M. Leclère, eds., *Pratiques plurilingues, mobilités et éducation: éclairages d'Afrique ou d'ailleurs*, 1st ed. Paris: Édition des Archives Contemporaines, pp. 79–96.
- Coquery-Vidrovitch, C. (2005). *The history of African cities South of the Sahara*. Princeton: Wiener.
- Cruise O'Brien, D. (1998). The shadow-politics of wolofisation. *The Journal of Modern African Studies*, 36(1), pp. 25–46.
- Dalby, D. (1970). Reflections on the classification of African languages: With special reference to the work of Sigismund Wilhelm Koelle and Malcolm Guthrie. *African Language Studies*, 11, pp. 147–171.
- Diamond, J.M. (1999). *Guns, germs, and steel: The fates of human societies*. New York and London: W. W. Norton.
- Di Carlo, P. (2016). Multilingualism, affiliation, and spiritual insecurity: From phenomena to processes in language documentation. *Language Documentation and Conservation*, pp. 71–104.
- Di Carlo, P. (2018). Towards an understanding of African endogenous multilingualism: Ethnography, language ideologies, and the supernatural. *International Journal of the Sociology of Language*, 254, pp. 139–163.
- Di Carlo, P., Agwara, A.D.E. and Ojong, R.A. (forthcoming). Multilingualism and the heteroglossia of ideologies in Lower Fungom (Cameroon). *Sociolinguistic Studies*.

- Di Carlo, P. and Good, J. (2014). What are we trying to preserve? Diversity, change and ideology at the edge of the Cameroonian Grassfields. In: P.K. Austin and J. Sallabank, eds., *Endangered languages: Beliefs and ideologies in language documentation and revitalization*, 1st ed. Oxford: Oxford University Press and The British Academy, pp. 231–262.
- Di Carlo, P. and Good, J. (2017). The vitality and diversity of multilingual repertoires: Commentary on Mufwene. *Language*, 93(4), pp. e254–e262.
- Dimmendaal, G.J. and Storch, A. (2016). Niger-Congo: A brief state of the art. *Oxford handbooks online*. Oxford University Press. [online] Available at: www.oxfordhandbooks.com/view/10.1093/oxfordhb/9780199935345.001.0001/oxfordhb-9780199935345-e-3.
- Donaldson, C. (2018). Orthography, standardization and register: The case of Manding. In: P. Lane, J. Costa and H. de Korne, eds., *Standardizing minority languages*, 1st ed. New York and London: Routledge, pp. 175–199.
- Dreyfus, M. and Juillard, C. (2004). *Le plurilinguisme au Sénégal : langues et identités en devenir*. Paris: Karthala.
- Eckert, P. (2008). Variation and the indexical field. *Journal of Sociolinguistics*, 12(4), pp. 453–476.
- Faraclas, N. (2013). Nigerian Pidgin. In: S.M. Michaelis, P. Maurer, M. Haspelmath, M. Huber and M. Revis, eds., *The survey of pidgin and creole languages*, 1st ed. Oxford: Oxford University Press.
- Fardon, R. and Furniss, G., eds. (1994). *African languages, development and the state*. London: Routledge.
- Finney, M.A. (2013). Krio. In: S.M. Michaelis, P. Maurer, M. Haspelmath, M. Huber and M. Revis, eds., *The survey of pidgin and creole languages*. Oxford: Oxford University Press.
- Fishman, J.A. (1989). *Language and ethnicity in minority sociolinguistic perspective*. Bristol: Multilingual Matters.
- François, A. (2014). Trees, waves and linkages: Models of language diversification. In: C. Bower, ed., *The Routledge handbook of historical linguistics*, 1st ed. London: Routledge, pp. 161–189.
- François, A. (2017). Méthode comparative et chaînages linguistiques: Pour un modèle diffusionniste en généalogie des langues. In: J.-L. Léonard, ed., *Diffusion: Implantation, affinités, convergence. Mémoires de la Société de Linguistique de Paris*. Louvain-la-Neuve: Peeters, pp. 43–82.
- Gafaranga, J. and Torras, M.-C. (2016). Interactional otherness: Towards a redefinition of codeswitching. *International Journal of Bilingualism*, 6(1), pp. 1–22.
- Gal, S. (2016). Scale-making: Comparison and perspective as ideological projects. In: E.S. Carr and M. Lempert, eds., *Scale: Discourse and dimensions of social life*, 1st ed. Oakland, CA: University of California Press, pp. 91–111.
- Good, J. (2013). A (micro-)accretion zone in a remnant zone? Lower Fungom in areal-historical perspective. In: B. Bickel, L.A. Grenoble, D.A. Peterson, A. Timberlake and J. Nichols, eds., *Language typology and historical contingency: In honor of Johanna Nichols*, 1st ed. Amsterdam and Philadelphia: John Benjamins, pp. 265–282.
- Good, J., Di Carlo, P. and Ojong, R. (2019). Multilingualism in rural Africa. *Oxford research encyclopedias*. doi:10.1093/acrefore/9780199384655.013.227.
- Goodchild, S. (2016). “Which language(s) are you for?” “I am for all the languages.”: Reflections on breaking through the ancestral code: Trials of sociolinguistic documentation. *SOAS Working Papers in Linguistics*, 18, pp. 75–91.
- Goodchild, S. (2019). *Sociolinguistic spaces and multilingual repertoires: Practices and perceptions in Essyl, Senegal*. PhD. School of Oriental and African Studies, University of London.
- Goodchild, S. (forthcoming). Multilingual people with monolingual perceptions: Patterns of multilingualism in Essyl, Basse Casamance, Senegal. In: F. Lüpke, ed., *The Oxford guide to the Atlantic languages of West Africa*, 1st ed. Oxford: Oxford University Press.
- Goodchild, S. and Weidl, M. (2018). Translanguaging practices in the Casamance: Similar but different – Two case studies. In: A. Sherris and E. Adami, eds., *Making signs, translanguaging ethnographies: Exploring urban, rural and educational spaces*, 1st ed. Bristol: Multilingual Matters, pp. 133–151.
- Güldemann, T. (2008). The Macro-Sudan belt: Towards identifying a linguistic area in northern sub-Saharan Africa. In: B. Heine and D. Nurse, eds., *A linguistic geography of Africa*. Cambridge: Cambridge University Press, pp. 151–185.

- Güldemann, T. (2010). Sprachraum and geography: Linguistic macro-areas in Africa. In: A. Lameli, R. Kehrein and S. Rabanus, eds., *Language and space: Volume 2, language mapping*, 1st ed. Berlin: De Gruyter Mouton, pp. 561–593.
- Güldemann, T. and Hammarström, H. (forthcoming). Geographical axis effects in large-scale linguistic distributions. In: P. Muysken, M. Crevels, and J.-M. Hombert, eds., *Language dispersal, diversification, and contact. A global perspective*, 1st ed. Oxford: Oxford University Press.
- Gut, U. (2015). English in West Africa. In: M. Filppula, J. Klemola, and D. Sharma, eds., *The Oxford handbook of world Englishes*, 1st ed. Oxford: Oxford University Press.
- Hantgan, A. (2017). Choices in language accommodation at the Crossroads: Convergence, divergence, and mixing. *Journal of the Anthropological Society of Oxford Online*, 9(1), pp. 102–118.
- Henrich, J., Heine, S.J. and Norenzayan, A. (2010). The weirdest people in the world? *Behavioral and Brain Sciences*, 33(2–3), pp. 61–83; discussion pp. 83–135.
- Intumbo, I., Inverno, L. and Holm, J. (2013). Guinea-Bissau Kriyol. In: S.M. Michaelis, P. Maurer, M. Haspelmath, M. Huber and M. Revis, eds., *The survey of pidgin and creole languages*. Oxford: Oxford University Press.
- Irvine, J.T. (2016). Going upscale: Scales and scale-climbing as ideological projects. In: E.S. Carr and M. Lempert, eds., *Scale: Discourse and dimensions of social life*, 1st ed. Oakland, CA: University of California Press, pp. 213–231.
- Jacobs, B. (2010). Upper Guinea Creole: Evidence in favor of a Santiago birth. *Journal of Pidgin and Creole Languages*, 25(2), pp. 289–343.
- Jørgensen, J.N., Karrebæk, M.S., Madsen, L.M. and Møller, J.S. (2011). Polylinguaging in superdiversity. *Diversities*, 13(2).
- Juffermans, K. and Tavares, B. (2017). South-North trajectories and language repertoires. In: C. Kerfoot and K. Hyltenstam, eds., *Entangled discourses: South-North orders of visibility*, 1st ed. London: Taylor & Francis, pp. 99–115.
- Juillard, C. (1995). *Sociolinguistique urbaine. La vie des langues à Ziguinchor (Sénégal)*. Paris: Presses du CNRS.
- Kießling, R. (2005). “Bak mwa me do”- Camfranglais in Cameroon. *Lingua Posnaniensis*, 47, pp. 87–107.
- Kopytoff, I. (1987). The internal African frontier: The making of African political culture. In: I. Kopytoff, ed., *The African frontier: The reproduction of traditional African society*, 1st ed. Bloomington, IN: Indiana University Press, pp. 3–84.
- Krajcik, C. (2019). *Exploring multilingualism in Senegal: A multimodal approach to the expression of caused motion*. PhD. School of Oriental and African Studies, University of London.
- Léglise, I. and Alby, S. (2016). Plurilingual corpora and polylinguaging, where corpus linguistics meets contact linguistics. *Sociolinguistic Studies*, 10(3), pp. 357–381.
- Lentz, C. (1995). ‘Tribalism’ and ethnicity in Africa. *Cahiers des Sciences Humaines*, 31(2), pp. 303–328.
- Lentz, C. (2013). *Land, mobility, and belonging in West Africa*. Bloomington, IN: Indiana University Press.
- Lexander, K.V. (2010). Voeux plurilingues électroniques – Nouvelles pratiques, nouvelles fonctions pour les langues africaines ? *Journal of Language Contact, THEMA*, 3, pp. 205–228.
- Lexander, K.V. and Alcón, D. (forthcoming). Digital language and new configurations of multilingualism: Language use in a Senegal-based discussion forum. In: F. Lüpke, ed., *The Oxford guide to the Atlantic languages of West Africa*, 1st ed. Oxford: Oxford University Press.
- Li Wei. (2018). Translinguaging as a practical theory of language. *Applied Linguistics*, 39(1), pp. 9–30.
- Luffin, X. (2002). Language taboos in Kinubi: A comparison with Sudanese and Swahili cultures. *Africa: Rivista Trimestrale di Studi e Documentazione dell’Istituto Italiano per l’Africa e l’Oriente*, 57(3), pp. 356–367.
- Lüpke, F. (2010). Language and identity in flux: In search of Baïnouk. In: F. Lüpke and M. Chambers, eds., *Multilingualism and language contact in West Africa: Towards a holistic perspective: Journal of Language Contact THEMA*, 3, pp. 155–174.

- Lüpke, F. (2016a). Pure fiction – The interplay of indexical and essentialist ideologies and heterogeneous practices: A view from Agnack. *Language Documentation and Conservation*, Special Publication, 10, pp. 8–39.
- Lüpke, F. (2016b). Uncovering small-scale multilingualism. *Critical Multilingualism Studies*, 4(2), pp. 35–74.
- Lüpke, F. (2017). African(ist) perspectives on vitality: Fluidity, small speaker numbers, and adaptive multilingualism make vibrant ecologies (Response to Mufwene). *Language*, 93(4), pp. e275–e279.
- Lüpke, F. (2018). Multiple choice: Language use and cultural practice in rural Casamance between convergence and divergence. In: J. Knörr and W.T. Filho, eds., *Creolization and pidginization in contexts of postcolonial diversity: Language, culture, identity*, 1st ed. Leiden: Brill, pp. 181–208.
- Lüpke, F. (forthcoming a). Language contact in West Africa: From objects of enquiry to new ways of thinking and linking language and languaging. In: M. Aronoff, ed., *Oxford research encyclopedia in linguistics*, 1st ed. Oxford: Oxford University Press.
- Lüpke, F. (forthcoming b). Language contact in West Africa. *Oxford research encyclopedias*.
- Lüpke, F. (forthcoming c). Standardization in highly multilingual contexts: The shifting interpretations, limited reach, and great symbolic power of ethnonationalist visions. In: W. Ayres-Bennett and J. Bellamy, eds., *The Cambridge handbook of standard languages*. Cambridge: Cambridge University Press.
- Lüpke, F. and Chambers, M., eds. (2010). *Multilingualism and language contact in West Africa: Towards a holistic perspective: Journal of Language Contact THEMA*, 3.
- Lüpke, F., et al. (accepted). LILIEMA: Language-independent literacies for inclusive education in multilingual areas. *Language and Development Series of the British Council*.
- Lüpke, F., et al. (under revision). Comparing rural multilingualism in Lowland South America and Western Africa. *Anthropological Linguistics*.
- Manessy, G. (1994). *Les Français en Afrique noire : mythe, stratégies, pratiques*. Paris: L'Harmattan.
- Marten, L. and Petzell, M. (2016). Linguistic variation and the dynamics of language documentation: Editing in 'pure' Kagulu. *Language Documentation and Conservation*, Special Publication, 10, pp. 105–129.
- Matras, Y. (2009). *Language contact*. Cambridge: Cambridge University Press.
- Matras, Y. (2012). An activity-oriented approach to contact-induced language change. In: C. Chamoreau and I. Légise, eds., *Dynamics of contact-induced language change. (Language contact and bilingualism, 2)*. Berlin: De Gruyter Mouton, pp. 1–28.
- Mc Laughlin, F. (1997). Noun classification in Wolof – When affixes are not renewed. *Studies in African linguistics*, 26(1), pp. 1–28.
- Mc Laughlin, F. (2001). Dakar Wolof and the configuration of an urban identity. *Journal of African Cultural Studies*, 14(2), pp. 153–172.
- Mc Laughlin, F. (2008). On the origins of Urban Wolof: Evidence from Louis Descemet's 1864 phrase book. *Language in Society*, 37(5), pp. 713–735.
- Mc Laughlin, F. (2014). Senegalese digital repertoires in superdiversity: A case study from Seneweb. *Discourse, Context and Media*, 4(4), pp. 29–37.
- McMahon, A., Heggarty, P., McMahon, R. and Slaska, N. (2005). Swadesh sublists and the benefits of borrowing: An Andean case study. *Transactions of the Philological Society*, 103(2), pp. 147–170.
- Meyerhoff, M. and Stanford, J.N. (2015). "Tings change, all tings change": The changing face of sociolinguistics with a global perspective. In: D. Smakman and P. Heinrich, eds., *Globalising sociolinguistics: Challenging and expanding theory*, 1st ed. London: Routledge Taylor & Francis Group, pp. 1–15.
- Milroy, L. (1980). *Language and social networks*. Oxford: Blackwell.
- Minkailou, M. (2016). Exploring euphemism in standard Songhay. *Recherches Africaines. Annales de l'Université des Lettres et Sciences Humaines de Bamako*, 16, pp. 31–39.
- Moore, L.C. (2008). Body, text and talk in Maroua Fulbe Qur'anic schooling. *Text & Talk*, 28(5), pp. 643–665.
- Mufwene, S.S. (1996). The founder principle in creole genesis. *Diachronica*, 13(1), pp. 83–134.

- Mufwene, S.S. (2001). *The ecology of language evolution*. Cambridge: Cambridge University Press.
- Mufwene, S.S. (2017). Language vitality: The weak theoretical underpinnings of what can be an exciting research area. *Language*, 93(4), e202-e223.
- Myers-Scotton, C. (1993). *Social motivations for codeswitching: Evidence from Africa*. Oxford: Oxford University Press.
- Ndecky, A. (2011). *Pratiques et représentations des parlers mancagnes de Goudomp (Sénégal)*. PhD. Université de Picardie Jules Verne.
- Ngom, F. (2017). West African manuscripts in Arabic and African languages and digital preservation. *Oxford Research Enclopedias*. doi:10.1093/acrefore/9780190277734.013.123.
- Nunez, J.F. (2015). *L'alternance entre créole afro-portugais de Casamance, français et wolof au Sénégal : une contribution trilingue à l'étude du contact de langues*. PhD. Institut National des Langues et Civilisations Orientales.
- Nunez, J.F. and Léglise, I. (2016). Ce que les pratiques langagières plurilingues au Sénégal disent à la linguistique du contact. In: M. Auzeanneau, M. Bento, and M. Leclère, eds., *Pratiques plurilingues, mobilités et éducation : éclairages d'Afrique ou d'ailleurs*. Paris: Édition des Archives Contemporaines, pp. 99–119.
- Nyst, V. (2010). Sign languages in West Africa. In: D. Brentari, ed., *Sign languages*. Cambridge: Cambridge University Press, pp. 405–432.
- Pozdniakov, K. (2007). Etudes atlantiques comparatives: Questions de méthodologie. In: A. Lemaréchal, ed., *Tradition et rupture dans les grammaires comparées de différentes familles de langues* 1st ed. Louvain-la-Neuve: Peeters, pp. 93–120.
- Pozdniakov, K. and Segerer, G. (2010). *Classification des langues atlantiques mise à jour*. Paper presented at *Sénélangues Workshop*, Ndayane, Sénégal.
- Pozdniakov, K. and Segerer, G. (forthcoming). A genealogical classification of Atlantic languages. In: F. Lüpke, ed., *The Oxford guide to the Atlantic languages of West Africa*, 1st ed. Oxford: Oxford University Press.
- Ranger, T. (1983). The invention of tradition in colonial Africa. In: E.J. Hobsbawm and T.O. Ranger, eds., *The invention of tradition*, 1st ed. Cambridge: Cambridge University Press, pp. 211–262.
- Rosch, E. (1975). Cognitive representations of semantic categories. *Journal of Experimental Psychology: General*, 104(3), pp. 192–233.
- Rosch, E. and Mervis, C. (1975). Family resemblance: Studies in the internal structure of categories. *Cognitive Psychology*, 7, pp. 573–605.
- Sagna, S. (2008). *Formal and semantic properties of the Gújjolaay Eegimaa nominal classification system*. PhD. University of London.
- Sagna, S. (forthcoming). A typological overview of Eegimaa (Jóola Banjal). *Studies in African Linguistics*.
- Sambou, P.M. (1979). *Diola kaasa esuulaalor. Phonologie, morphophonologie et morphologie*. PhD. Université Cheikh Anta Diop de Dakar.
- Sapir, J.D. (1965). *A grammar of Diola-Fogny*. Cambridge: Cambridge University Press.
- Sapir, D.J. (1975). Big and thin: Two Diola-Fogny meta-linguistic terms. *Language in Society*, 4(1), pp. 1–15.
- Silverstein, M. (2003). Indexical order and the dialectics of sociolinguistic life. *Language & Communication*, 23, pp. 193–229.
- Smakman, D. and Heinrich, P., eds. (2015). *Globalising sociolinguistics: Challenging and expanding theory*. London: Routledge Taylor & Francis Group.
- Souag, L. (2010). Ajami in West Africa. *Afrikanistik Online*, 7, pp. 1–11. Available at: <http://eprints.soas.ac.uk/13429/>.
- Swigart, L. (1994). Cultural creolisation and language use in post-colonial Africa: The case of Senegal. *Africa*, 64(2), pp. 175–189.
- Tamari, T. and Bondarev, D., eds. (2013). Qur'anic exegesis in African languages. *Journal of Qur'anic Studies*. School of Oriental and African Studies, University of London.
- Torres Cacoullous, R. and Travis, C.E. (2018). *Bilingualism in the community*. Cambridge: Cambridge University Press.

- Trudgill, P. (2011). *Sociolinguistic typology: Social determinants of linguistic complexity*. Oxford: Oxford University Press.
- van Coetsem, F. (2000). *A general and unified theory of the transmission process in language contact*. (*Monographien zur Sprachwissenschaft, Bd. 19*). Heidelberg: Winter.
- Vydrine, V. (1999). Language contact: The case of Manding and Jola Fogny. *Mande Studies*, 1, pp. 141–157.
- Watson, R. (2017). *Data management workflow*. [online] Available at: <http://soascrossroads.org/2017/05/> [Accessed 13 Feb. 2017].
- Watson, R. (2018). Patterns of lexical correlation and divergence in Casamance. *Language & Communication*, 62(B) pp. 170–183.
- Watson, R. (2019). Language as category: Using prototype theory to create reference points for the study of multilingual data. *Language and Cognition*, 11, pp. 125–164.
- Watson, R. (in preparation). A distributed workflow for the creation, annotation, and archiving of multilingual data in collaborative, multisited documentation projects.
- Weidl, M. (2019). *Nga ne!? [What did you say!]: The role of Wolof in multilingual conversations in the Casamance: Fluidity of linguistic repertoires*. PhD. School of Oriental and African Studies, University of London.
- Winford, D. (2007). Some issues in the study of language contact. *Journal of Language Contact, THEMA*, 1.
- Woodbury, A.C. (2005). Ancestral languages and (imagined) creolisation. *Language Documentation and Description*, 3, pp. 252–262.



Taylor & Francis

Taylor & Francis Group

<http://taylorandfrancis.com>

Index

Note: page numbers in *italic* indicate a figure and page numbers in **bold** indicate a table.

- 4-M model 185; borrowing and 243; convergence in 76–79; critical issues and topics in 66–75; current contributions and research in 75–81; historical overview of 64–66; language production in 66–67; mixed languages and 336; mixed languages in 79–80
- Acholi 69, 72–73
- AECs *see* African English-lexifier contact languages (AECs)
- African-American Vernacular English 134, 307–308
- African English-lexifier contact languages (AECs) 135
- Afro-Bolivian Spanish 136–137
- ageing: bilingualism and 177–178; language learning and 172
- Ahtna 505–506
- Aki 494, 532
- Alabama 519
- Albanian 281, 386–388, 393, 395–396
- AMTB *see* Attitude and Motivational Test Battery (AMTB)
- Anatolia: critical issues and topics in 408, 408–411, 411; current contributions and research in 411–421, **413**; future directions with 426–427; historical background in 404–406; methodological and theoretical issues in **421**, 421–426; morphological isoglosses in 414–419; phonetic and phonological isoglosses in 412–414, 413; Sprachbund in 409; syntactic isoglosses in 419–421; word order and 422–423; zoning attempts in 410–411, 411
- Anatolian Arabic 406
- Angloromani 195
- annotation of data 12–14, **13**; automated 14–15
- Antioch Arabic 406
- Apalachee 519
- Arabic 65, 73, 160, 240, 246–247, 372, 405–406, 408, 408, 414–415, 532
- Aramaic 405–406, 408, 414, 418, 423
- Arizona Tewa 509
- Armenian 387–388, 404–406, 408, 413, 416–417, 419–421, **421**, 422–426
- Aromanian 392–393, 397
- Arta 490
- Arvanitika 281
- Asian region: climate and 441–442; computer-mediated communication and 452–454; critical issues in 435–451; current research in 435–451; future directions with 451–455; geopolitics and 441–442; globalization and 451–452; historical overview of 435–451; Pearl River Delta 449; phenomena defining 434; trade and 441–442, 454–455
- Asmat 484
- Asymmetry Principle 66–67
- Atakapa 513, 515, 519
- Athapaskan 505–506, 507
- Attitude and Motivational Test Battery (AMTB) 224
- attrition, language 179; critical issues and topics in 224–225; current contributions and research in 226–227; emotions and 224; future directions with 228–229; globalization and 225; historical overview of 222–224; and length of residence 223; Principle Component Analysis and 224; Radical Embodied Cognition and 226; research methods with 227–228; retrieval induced forgetting and 226–227; variation in 222–223
- automated annotation methods 14–15
- Azerbaijani 422
- Baba Malay (BM) 443–445, 448
- baby talk hypothesis 305–306
- Bainouk Gubeeher 534, 535, 540
- Balkan Armenian 387
- Balkans: analytic comparatives in 392; analytic perfect in 394; anterior future in 392; clitic order in 393; critical issues and topics with

- 391–395; current contributions and research in 396; declensional system in 393; definition of, as geographic area 385–386; definition of, as linguistic area 386; Essentially Rooted in Conversation loanwords in 394; etymology of 385; evidentials in 393; future directions with 397–398; historical overview of 387–391, **390**; languages of 386–387; languages of, external histories and 388–389; lexicon 394–395; linguistics of, as field 389–391, **390**; morphosyntax in 391–394; object reduplication in 392–393; phonology in 395; postposed definite article in 392; research methods with 396–397; resumptive clitic pronouns in 392–393; as Sprachbund 385; ten in 394
- Barwar 413
- Basque 206–208
- Batavian Malay 445
- Bayot 534
- Bazaar Malay (BZM) 442–445
- Belarusian 278–279
- Bella Bella 507
- Bella Coola 506, 507
- Betawi Malay 445
- bilingual advantage hypothesis 176–177
- Bilingual Bootstrapping Hypothesis 207
- Bilingual Navajo 522–523
- bilingual speakers: ageing and 177–178; critical period hypothesis and 172; early 171; executive control in 176–177; forms of individual bilingualism 171–173; grammar in action 268–269, 269; as intertwined, languages of 29–31; language contact in mind of 29–33; language processing in 173–175; late 171; second language learning and 30–31; simultaneous 171–172; in structuralism 170; subsystems hypothesis and 173–174; successive 172; *see also* children, bilingual
- bilingual tag 92
- BillingBank 10
- Biloxi 519
- Bilua 486
- Blocking Hypothesis 189
- BM *see* Baba Malay (BM)
- BNC *see* British National Corpus (BNC)
- borrowability 241–242; of morphology 246–247
- borrowing: 4-M model and 243; code-switching *vs.* 239–240; constructional 119–120; critical issues and topics in 239–244; culture and 242; current contributions and research in 244–247; defined 237; future directions with 248–249; glottochronology and 243–244; historical overview of 238–239; implicational hierarchies of, and convergence 289; importation in 238; as innovative reproduction 238; integration and 240–241; language genealogy and 243–244; Markedness Hypothesis and 244; matter 117, 287; motivations for 242–243; Natural Language Processing and 249; nonce 256–258, 257, 259; pattern 117, 287; pragmatic 159–162; prestige and 243; research methods with 247–248; social motivations for 247; Specificity Hypothesis and 244–245; substitution in 238; typology in 186–188; in usage-based approaches 110, 116, 119–120; usage frequency and 245–246; Utilitarian motivation and 245; *see also* lexical borrowing
- Brazilian Portuguese 136
- bridges 73–74
- British National Corpus (BNC) 10
- Broken Slavey 519
- Bulgarian 387, 393, 397
- burstiness 18, 21
- BZM *see* Bazaar Malay (BZM)
- Caddo 519
- Cafeteria Principle 306
- Cahuilla 510–511
- calques 117, 238, 287, 293n2, 412, 540
- Cameroon Pidgin 135, 305
- Cantonese 210, 447
- Canton System 449
- Cappadocian 415, 417, 424–425
- Cappadocian Greek 406
- Categorical Congruence 190
- CDST *see* complex dynamic systems theory (CDST)
- CEMP *see* Central Eastern Malayo-Polynesian (CEMP)
- Central Cost Salish 507
- Central Eastern Malayo-Polynesian (CEMP) 496
- Central Malayo-Polynesian (CMP) 496
- Centre of Gravity Hypothesis 191–192
- Chehalis 518
- Chickasaw 519
- children, bilingual: Bilingual Bootstrapping Hypothesis and 207; critical issues and topics in 204–206; cross-linguistic influence in 204–206, 208–209, 211–212; cross-linguistic priming and 211–212; cross-linguistic structures as relief strategy in 207–208; current approaches and research with 206–212; future directions with 213–214; grammatical maturation in 206–207; grammatical overlap and 209–210; historical overview of 202–204; interface areas and 210–211; Interface Hypothesis and 210–211; Ivy Hypothesis and 207; language dominance and 208–209; lexical mixing in 201; lexical transfer in 202; Mixed Languages and 340–341; morphological mixing in

- 201–202; research methods with 213; Separate Development Hypothesis in 203–204; Unitary Systems Hypothesis in 202–203
- China Coast Pidgin 449
- Chinese 190, 352
- Chinese Pidgin English (CPE) 449–451
- Chinookan 507
- Chinook Jargon 516–518
- Chintang 261
- Chipewyan 519
- Chitimacha 519
- Chitty Malay 445
- Choctaw 519
- Chumash 510
- CIA *see* Copper Island Aleut (CIA)
- Circassian 406
- circumscribing variable context 46
- Clackamas 518
- Clatskanie 507
- CLI *see* cross-linguistic influence (CLI)
- CMC *see* Computer Mediated Communication (CMC)
- CMI *see* code-mixing index (CMI)
- CMP *see* Central Malayo-Polynesian (CMP)
- CMs *see* content morphemes (CMs)
- Cocos Malay 443
- code copying 287
- code-mixing index (CMI) 18
- code-switching (CS) 156–159; borrowing *vs.* 239–240; community norms and 260–262; composite 67; constraint-based program with 89–90; constraint-free approach with 90, 97–102, 101; contextual distribution via 255; counting 262–264, 264; critical issues and topics in 97–101, 101, 256–262, 257, 259–260; current contributions and research in 102–104, 262–269, 264, 267, 269; defined 88; embedded language in 64–65; Equivalence Constraint in 90–91, 257–258; Free Morpheme Constraint and 91–92; Functional Head Constraint and 95–96; future directions with 104, 269–270; Government Constraint and 94; grammar of 88–104, 101; grammatical similarity and 265–268, 267; harmonization in 190; historical overview 89–97; lenition and 255–256, 256; lexical borrowing and 157–158; lexical borrowing *vs.* 47; matrix language in 64, 67; and mechanisms of change 253–256, 256; minimalist model of 101; Mixed Language and 338–340; motivations in 259–262, 260; nonce borrowing and 256–258, 257, 259; non-lexicalist approaches to 102–103; Null Theory of 93; phrase structure rules and 92–93; prepositions in 70–71; psycholinguistics and 31–32; for rhetorical purposes 261; syntactic boundary and 262–263; typology in 188–190, 189; in usage-based approaches 116; verbs in 68–70; word-internal 103–104
- cognate facilitation effect 35
- cognates 35
- cognition: social psychology *vs.*, in convergence 283–284
- cognitive factors: bilingual advantage hypothesis 176–177; critical issues and topics in 171–175; current contributions and research in 175–178; historical overview of 169–171; of language contact 168–180; research methods with 178; types *pf* 168
- Cognitive Linguistics 113
- cognitive reserve 177
- communities of practice 49
- community, speech 49–50
- community norms 49, 260–262
- comparative, analytic 392
- competition: and cross-linguistic priming in child grammars 211–212
- Complementizer system 210
- complex dynamic systems theory (CDST) 178–179
- Complexity Factor 18
- Computer Mediated Communication (CMC) 376, 452–454
- conflict sites 53
- Congo Shaba Swahili 69
- Congo Swahili 72
- ConLL-U 12–13
- connectionism 174
- constraint hierarchy 53
- Construction Grammar 113
- contact-induced grammaticalization 254, 287
- contemporary urban vernacular 368
- content morphemes (CMs) 64–68, 70–71, 81, 189–190
- contextual distribution: via code-switching 255
- convergence, grammatical 47; 4-M model and 76–79; area 282; area size and 291; assessment of 53; code copying and 287; code-switching and 156–159; cognition and 283–284; comparison of approaches in 288–289; contact-induced grammaticalization and 287; correlations between types of, and contact conditions 284–285; critical issues and topics in 283–286; current contributions in 286–290; defined 276, 293n2; diastratic 291–292; diastratic leveling and 279–280; diatopic 291–292; discourse pragmatics and 280; distance and, quantification of 289–290; factors 283–284; historical overview of 277–283; identification models and 280–281, 286–289; implicational hierarchies of, and borrowing 289; and interlingual identification of linguistic subsystems 288; koineisation and 279–280; in Linguistic Melanesia 489–490;

- matter borrowing and 287; mechanisms of 286–289; metatypy and 280–281; and new dialect formation 285–286; pattern borrowing and 287; pivot matching and 287; polysemy copying and 288; research methods with 286–290; social psychology and 283–284; sociolinguistics and 280; transitional dialects and secondary 278–279; in typology 281–283; in West Africa 529–530, 538–541
 Coosans 507
 Copper Island Aleut (CIA) 79–80, 520
 corpus: as data source in variationist perspective 51
 CPE *see* Chinese Pidgin English (CPE)
 CPH *see* critical period hypothesis (CPH)
 Cree 191–192, 328, 519–520
 creole exceptionalism 313–314
 creole languages: in 4-M 80–81; baby talk hypothesis and 305–306; Cafeteria Principle and 306; critical issues and topics in 305–313; current contributions and research in 313–315; demography and 132; future directions in 315–317; historical overview of 301–305; koineisation and 279; metatypy and 280–281; Mixed Languages *vs.* 325–326; origin of term 305; relexification hypothesis and 306–307; social factors in 130, 135; universalist hypothesis and 309; *see also* pidgin languages
 creolization 135
 critical period hypothesis (CPH) 172
 Croatian 121–122
 cross-language activation 30
 cross-linguistic influence (CLI): bilingual children and 204–206; defined 204; grammatical maturation and 206–207; grammatical overlap and 209–210; interface areas and 210–211; language dominance and 208–209
 cross-linguistic priming 211–212
 cross-variety comparison 52–56
 CS *see* code-switching (CS)
 Cupeño 510–511
 Czech 16, 16–17, 17, 19, 161, 278–279

 Danish 77–78
 data annotation 12–14, 13
 data sources 9–12
 demography 132
 demonstratives 160
 Dene Slave 519
 Devanagari 452
 Dhivehi 436
 diagnostic difference 254
 diastratic leveling 279–280
 Diegueño 510–511
 Differential Access Hypothesis 67, 74–75
 diffusion 277

 discourse markers 159–160
 discourse pragmatics 280
 Distributed Morphology (DM) 102–103
 Diyarbakir 413
 DM *see* Distributed Morphology (DM)
 Domari 246–247
 Doms 407
 Dravidian culture 435–437
 Dutch 70, 119, 178, 190, 206, 245–247, 370–375

 Eastern Polynesia: consonant inventories in 464; critical issues and topics in 465–466; current contributions and research in 466–475; dominance in 466; French Polynesia 471–474; future directions with 475–476; historical overview of 463, 463–465, 464, 464; identity in 466; indirect *vs.* direct contact in 465–466; linguistic relationships in 464; Rapa Nui 474–475; socio-political pressure in 462; variation in 466
 EEG *see* electroencephalography (EEG)
 EL *see* Embedded Language (EL)
 electroencephalography (EEG) 39–40, 228–229
 Embedded Language (EL) 64–65, 68–70, 96
 emotions: language attrition and 224
 Enga 484
 Enggano 490
 English 134
 enslaved peoples 136
 entropy score 18
 equivalence: usage-based approaches and 117–118
 Equivalence Constraint 90–91, 189, 257–258
 ERIC *see* Essentially Rooted in Conversation (ERIC) loanwords
 ERP *see* event-related potentials (ERP)
 Esselen 510
 Essentially Rooted in Conversation (ERIC) loanwords 394
 Estonian 203
 Europarl 18
 event-related potentials (ERP) 39–40, 228
 evidentials 393
 Eyak 507
 eye-tracking: during reading 36–37; during spoken language comprehension 37–39

 Farsi 93, 98–99, 102
 Fernando Po Creole English 448
 FHC *see* Functional Head Constraint (FHC)
 Finnish 12, 67, 339
 founder principle 306, 443, 448
 Francoprovençal 160
 Free Morpheme Constraint 91–92
 French 69, 72–73, 76, 80, 160, 191, 205, 209, 308–309, 519, 534

- French Polynesia 471–474
 Functional Head Constraint (FHC) 9, 95–96
 Functional Turnover Hypothesis 335
- Gagauz 387
 Garo 341
 gaze duration 36
 genealogy, language: borrowing and 243–244
 Georgian 422
 German 77–78, 119, 151, 155–156, 161, 209,
 238–239, 284, 519
 Ghanaian Pidgin English 135, 139
 Gitksan 507
 globalization: Asian region and 451–452;
 language attrition and 225
 glottochronology 243–244
 Gorna Belica 393
 Government Constraint 94
 graffiti 357
 grammatical maturation 206–207
 grammatical overlap 209–210
 grassroots heteroglossia 134
 Greek 241, 247, 281, 387, 397, 406–408, 408,
 413, 417
 Gullah 306
 Gurindji Kriol 191–192, 195, 333–334, 339,
 341–342
 Gwich'in 519
- Haida 504, 507
 Haihais 507
 Haisla 507
 Haitian Creole 70, 80, 139, 306
 harmonization 190
 Hawai'i Creole English 469–470
 Hawai'i English 462, 471
 Hebrew 154, 160
 Hemshinli Armenian 406
 Heritage Language Variation and Change in
 Toronto 10
 Highest Alemannic 160–162
 Hindi 447
 HKE *see* Hong Kong English (HKE)
 Hokkien 447–448
 Hong Kong English (HKE) 447, 454
 Hungarian 67
- identification models 280–281
 Igbo 54
 Ilongot 490
 immediacy assumption 36
 Indian Ocean 434
 Indo-Aryan culture 435–437
 Indonesian 245
 insertion 9
 Integration Index 18, 20
- Interface Hypothesis 210–211
 Intonational Unit (IU) 16, 263–264
 Inuktitut English Pidgin 518–519
 Istanbulite Greek 407
 Italian 65, 75–77, 160–162, 209
 Ivy Hypothesis 207
 Ixcatec 121–122, 241
- Jakarta Malay 445
 Japanese 136
 Jargon Loucheux 519
 Java Malay 446
 Java Port Malay 446
 Javindo 331–332
 Jenisch 330–331
 Jireček line 388
 Joola Banjai 534, 535–536, 536, 540
 Joola Fogny 540
 Joola Kasa 534, 538–539
 Joola Kujireray 534, 538, 540
 Judeo-Spanish 407
 Judezmo 390
- Kalaallisut 519
 Kalapuyans 507
 Kannada 158
 Kartvelian 415
 Kazakh 158
 Kiezdeutsch 369
 Kiliwa 511
 Kiowa Apache 519
Kitche Duits 153–154
 Koasati 519
 Kobon 484–485
 Køge Corpus 10
 koineisation 279–280
 Komi-Russian 13, 13
 Komi-Zyrian 12
 Korean 136, 226
 Krio 135, 532
 Kriol 534
 Kriolu 535
 Kuot 486
 Kurdish 407, 408, 414, 416, 418
 Kurmanji 413–414, 416–417
 Kutenai 512
 Kwakiutl 507
 Kwalhioqua 507
- L2 stylization 370–372
 LA *see* Lexical Array (LA)
 Lake Miwok 509
 language attrition 179
 language contact: actions and activity types
 as drivers of 150–156; cognitive factors of
 168–180; consequences of, in bilingual mind

- 32–33; critical issues and topics in 156–162; defined 148; future directions in 162–163; historical overview of 150–156; migration and 150–151; in mind and brain of bilingual speakers 29–33; research methods 33–40
- language dominance: cross-linguistic influence and 208–209
- language index 94
- Language Policy and Planning (LPP) 351, 354
- language processing: in bilinguals 173–175
- language tag 92
- Lankan 437, 438, 440–441
- Laz 407, 417, 421, 422, 425–426
- Laze 408
- LDC *see* Linguistic Data Consortium (LDC)
- learning, second-language, bilingualism and 30–31
- legal nonword 34
- Lekoudesch 154, 330
- lenition 255–256, 256
- lesser-used languages (LUL) 353–354
- Levantine Arabic 414–415
- Lexical Array (LA) 98, 101
- lexical borrowing 157–158; code-switching *vs.* 47; comparative method with 53–54; defined 47; *see also* borrowing
- lexical copying 509
- lexical decision 34
- Lexical Functional Grammar 194
- lexical item (LI) 98
- lexical mixing 201
- lexical transfer 202
- LI *see* lexical item (LI)
- Light Warlpiri 191–192, 332–333, 337–339, 342
- Lingua Franca 303
- Linguistic Data Consortium (LDC) 10
- Linguistic Landscape (LL): critical issues and topics in 353–355; current contributions and research in 355–358; defined 349–350; English in 354; future directions in 359–361; historical overview of 351–353; language awareness and 357–358; lesser-used languages and 353–354; origin of term 351; research methods with 358–359; in semi-public spaces 357; transgressive urban semiotics and 357; urban multilingualism and 349, 352–353
- Linguistic Melanesia: convergence in 489–490; critical issues in 492–494; current and future directions of research in 494–497; defined 480, 481; historical overview of 482–492, 483, 485–491; lexical features in 488–491, 489–491; multilingualism in 492–493; phonological features in 483, 483–485, 485–486; semantic features in 488–491, 489–491; word order features in 486–487, 487–488
- LL *see* Linguistic Landscape (LL)
- loanblends 238
- loanshifts 238
- loanwords 238
- Loms 407
- Low German 77
- LPP *see* Language Policy and Planning (LPP)
- Luiseno 510–511
- LUL *see* lesser-used languages (LUL)
- Luxemburgish 152
- Ma'a 79, 242, 331
- Macanese 304
- Macaronic Latin 154
- Macedonian 387, 393, 395, 397
- Macro-Sudan belt 530, 531
- Makah 507
- Malay 190, 193
- Maldivian 436
- Mandarin 447
- Mandinka 540
- Māori 247, 467
- Māori English 466–468
- Marathi 158
- Mardin 413
- marginalized peoples 136
- Marind 484
- Maritime Polynesian Pidgin (MPP) 304–305
- Markedness Hypothesis 244
- masked priming 34
- matrěški*-principle 282
- Matrix Language (ML) 64–65, 68–70, 96
- Matrix Language Frame (MLF) Hypothesis 9, 64–67, 96–98, 185, 188–189, 335
- Matrix Language Turnover 191–192
- Matter Borrowing 117, 287
- Mauritian 81
- Media Lengua 328–329, 342
- Mednyj Aleut 334, 340, 520–521
- Meglenoromanian 393, 397
- Melanesian Pidgin English 305
- memory 18–19
- mentalists turn 169
- metatypy 280–281
- Miami Corpus 19
- Michif 191–192, 328, 519–520
- Micmac Pidgin English 519
- migration 150–151
- Mikasuki 512
- M-index 20
- Minimal Default Grammar 209
- Minimalist Program (MP) 98, 102–103, 240
- Miwok 506
- Mixed Language(s) 79–80; 4-M and 336; in adults *vs.* children 340–341; autonomy and 337–338; code-switching and 338–340; as conscious decision 335, 340; converted languages 334–335; creoles *vs.* 325–326;

- critical issues and topics with 327; current contributions and research in 327–341; emergence of 335–338; formation 190–192; future directions with 343; historical overview of 326–327; lexicon-grammar 328–332, 339; Matrix Language Frame and 336; in North America 519–532; noun-verb split 332–334; outcome of, vs. bilingualism 338–341; research methods with 341–343; terminology for 326; transitory phase in 339; types of language combination in 327–335
- ML *see* Matrix Language (ML)
- MLF *see* Matrix Language Frame (MLF)
- Hypothesis
- Mobilian Jargon 519
- modularity: usage-based approaches and 117
- Molise Slavic (SVM) 20
- Monache 506
- Mono 506
- MOP *see* Morpheme Order Principle (MOP)
- Moroccan Arabic 65, 73, 157
- Moroccan Flavoured Dutch 370–372, 375, 377
- Morpheme Order Principle (MOP) 64, 189
- morphemes: content 64–68, 70–71, 81, 189–190; and Free Morpheme Constraint and 91–92; language production and 66–67; system 64–68, 70–75, 81, 189, 189–190; and System Morpheme Principle 64; types of 67–68; typological profile of 185
- morphological mixing 201–202
- morphology: agglutinative 186; in Anatolia 414–419; in Asian region 437, 438; autonomy and 337–338; borrowability of 246–247; borrowing and 194; contextual 194; derivational 160, 242; distributed 102–103; double 72–73; emergence and 335–336; Igbo 54; inflectional 186–191, 193, 195, 242, 416; lexical borrowing and 56, 241; in North America 511–512; and noun-verb split languages 332; pragmatic borrowing and 150
- morphology borrowability 246–247
- morphosyntax: 4-M model and 66, 336; in Balkan languages 391–394; convergence and 76; creole languages and 80, 305; heteroglossia and 134; L2 stylization and 371; lexical borrowing and 53–54; mixed language and 340; pidgin languages and 80, 305
- MP *see* Minimalist Program (MP)
- MPP *see* Maritime Polynesian Pidgin (MPP)
- multi-ethnolect 368
- Multilingual index 18
- multilingual speakers: as intertwined, languages of 29–31
- Munsee 516
- Nahuatl 97, 101–102
- Naija 532
- Natchez 519
- Natural Language Processing (NLP) 8, 10–11, 14, 249
- Navajo 522–523
- Negro Pidgin French 308–309
- NENA *see* North-Eastern Neo Aramaic (NENA)
- Neo Aramaic 408
- New Mexican Spanish 122
- New Zealand English 247, 286, 466–468
- Nez Perce 511–512
- Nigerian English 448
- Nigerian Pidgin 135, 300, 305
- Nishga 507
- nonce borrowing 256–258, 257, 259
- Nootkans 507
- norms, community 49
- North America: alignment in 512–515, 514; contact languages in 515–523; critical topics in 505–509, 507–508; current contributions and research in 509–523, 514; future directions with 523; historical overview in 504–505; jargons in 516–519; mixed languages in 519–532; morphology in 511–512; phonology in 510–511; pidgins in 516–519
- North-Eastern Neo Aramaic (NENA) 405, 418
- Northern Wakashan 506
- North Western Neo-Aramaic 406
- Norwegian 186, 206
- Nouchi 368, 374, 377
- Nuuchahnulth 518
- Nuxalk 506
- object reduplication 392–393
- Ofo 519
- Old Rapa 474
- One Parent One Language 120
- Oowekeeno 507
- Optimality Theory (OT) 99
- OT *see* Optimality Theory (OT)
- outsiders 73–74
- Paipai 511
- Päkehā 467
- Pangloss Collection 10, 19
- Papia Kristang 304, 446
- paralexical insertion 372–374, 373
- Para-Romani varieties 329–330
- part-of-speech (POS): annotation 8; tagging 11–12, 14–15
- Pattern Borrowing 117, 287
- Patwin 509
- Pazih 490
- PDM *see* Pidgin-derived Malay (PDM)
- Penn TreeBank Tagset 15
- Persian Arabic 414–415
- Petjo 331–332

- PFIC *see* Phonetic Form Interface Condition (PFIC)
- Pharasiot Greek 407, 425
- Philippine Hybrid Hokkien 342
- Phonetic Form (PF) 98, 101
- Phonetic Form Interface Condition (PFIC) 100, 102
- phonology: in Anatolia 412–414, 413; attrition and 223; in Balkans 395; borrowing and 240; code-switching and 99; in Hawai'i Creole English 469–470; identity and 156; and Indo-Aryan and Dravidian admixture 435; lexical 101; in Linguistic Melanesia 483, 483–485, 485–486; in Māori English 467–468; metatypy and 280; in North America 510–511
- PI *see* Pidginization Index (PI)
- Pichi 135, 448
- picture-word Stroop 34
- Pidgin Delaware 516
- Pidgin-derived Malay (PDM) 443–445
- Pidginization Index (PI) 309
- pidgin languages: in 4-M 80–81, 151–152; baby talk hypothesis and 305–306; Cafeteria Principle and 306; contacts in 302; critical issues and topics in 305–313; current contributions and research in 313–315; expanded 302–303, 309; future directions in 315–317; historical overview of 301–305; incipient 303; as *lingua franca* 303; in North America 516–519; origin of term 305; relexification hypothesis and 306–307; social factors and 130; universalist hypothesis and 309; *see also* creole languages
- pivot-matching 287
- Polish 151, 278
- Polissja 278
- polysemy copying 288
- Pomak 154–155
- Pomoan 509
- Pontic Greek 407, 416, 421
- Portuguese 245, 304, 334–335
- postposed definite article 392
- pragmatic borrowing 159–162
- pragmatics: defined 148–149; identity and 149; in Māori English 468; research questions in 149–150
- prepositions: in code-switching 70–71
- Principle Component Analysis 224
- Principle of Accountability 46–47
- pronouns: personal 160; resumptive clitic 392–393
- Proto-Austronesian (PAN) 490
- Proto-Takic 510
- Proto-Uto-Aztecan 510
- Proto-Yuman 510–511
- psycholinguistics: usage-based approaches and 115–116
- Puluwat 487
- quantitative analysis 52
- quantitative reasoning 52
- Quechua 186, 241, 243, 328–329, 335
- Quichua 342
- Quileute Chemakum 507
- race 389
- Radical Embodied Cognition 226
- reading: eye-tracking during 36–37; self-paced 35–36
- reading-moving window 35–36
- reification 529, 536–538, 538
- relexification hypothesis (RH) 306–307
- Reo Papa 474
- research methods 17–21, 21
- response-time techniques 34–35
- resumptive clitic pronouns 392–393
- retrieval induced forgetting (RIF) 226–227
- RH *see* relexification hypothesis (RH)
- RIF *see* retrieval induced forgetting (RIF)
- Romani 67, 79–80, 121–122, 154–155, 240, 242, 246, 329–330, 342, 387, 389, 395–396, 426
- Romanian 99, 389–390, 392
- Romeica 407, 421–422, 424
- Rotwelsch 154
- Rumelian Turkish 387, 390, 396
- Russian 12–13, 13, 79, 119, 152, 158, 160, 520–521
- Russonorsk 152–154
- Sabah Pidgin Malay 445
- Salinan 510
- schematic units 118
- second language learning: bilingual speakers and 30–31
- self-paced reading 35–36
- semantic relatedness judgments 34
- sensitive period 172
- Separate Development Hypothesis 203–204
- Serbian 395
- SgE *see* Singapore English (SgE)
- Sheng 374, 377
- shift-break language 474
- Shona 72
- SHWNG *see* South Halmahera-West New Guinea (SHWNG)
- signs *see* Linguistic Landscape (LL)
- Siirt 413
- Singapore English (SgE) 447–448
- Sinhala 436–437, 439, 440
- Siuslawans 507

- Slave 519
 slavery 136, 311
 Slavey Jargon 519
 SLM *see* Sri Lanka Malay (SLM)
 SLP *see* Sri Lanka Portuguese (SLP)
 SMP *see* System Morpheme Principle (SMP)
 SMs *see* system morphemes (SMs)
 social factors: in creolization outcomes 135;
 critical issues and topics in 132–134; current
 contributions and research in 134–137; defined
 129; demography in 132; and expansive
 lingua franca 134–135; future directions with
 138–139; and grassroots heteroglossia 134;
 historical overview of 130–132; research
 methods with 137; socio-economic structure in
 132–133
 social motivations: for borrowing 247
 socio-economic structure 132–133
 sociolinguistics: convergence and 280; usage-
 based approaches and 113–115
 Sorbian 120, 122, 288
 sources of data 9–12
 Southern Coast Salish 507
 South Halmahera-West New Guinea
 (SHWNG) 496
 South Jutish 77
 Southwestern Cost Salish 507
 Southwestern Pomo 509
 Spanish 73, 75, 78–79, 93–94, 97, 99–101,
 121–122, 136–137, 186, 189, 205, 207–208,
 212, 226, 241–243, 252–256, 259–260, 260,
 262, 265–267, 267, 267–269, 269, 328–329,
 407, 519
 SPEAKING model 170, 174–175
 Specificity Hypothesis 244–245
 Spice Islands 441–442
 SpinTX 19–20
 Sprachbund 193, 282, 385, **390**, 409
 Sprachmischung 147, 169
 spread zones 281–282
 Sranan 190
 Sri Lanka **437**, 437–441, **440**
 Sri Lanka Malay (SLM) 193, 334–335, 437, **437**,
 437–440, **440**, 443
 Sri Lanka Portuguese (SLP) 334–335, 437, **437**,
 439–440
 Standard American English 134
 Strong Lexicalist framework 194
 structuralism 169
 subsystems hypothesis 173–174
 surprisal 18
 Swahili 64–65, 68–70, 72, 190
 Swedish 186
 Swiss German 65, 75–76
 syntactic boundary 262–263
 System Morpheme Hypothesis 97
 System Morpheme Principle (SMP) 64, 96
 system morphemes (SMs) 64–68, 70–75, 81,
 189, 189–190
 TAG *see* Tree Adjoining Grammars (TAG)
 Tagalog 342, 447
 Tahitian French 471–472
 Takelma 507
 Takia 194, 494
 Takic 510–511
 TalkBank 10
 Tamil 193, 439, **440**
 Tanaina 505–506
 Tangsi Malay 446
 Telugi 158
 Tetun Dili 245
 Thessaloniki Armenian 392
 Thrace Romani (RMN) 20
 Tillamook 507
 Tillo 413
 Tlingit 504, 507
 TNG *see* Trans-New Guinea (TNG)
 Tok Pisin 80–81, 261–262, 300
 tonal systems 483, 483–484
 Total Linguistic Fact 375
 Trabzon Turkish **421**, 422, 425–426
 Trader Navajo 519
 transgressive urban semiotics 357
 translanguaging studies 114–115
 Trans-New Guinea (TNG) 480, 495
 Tree Adjoining Grammars (TAG) 93
 Tsakonian 387
 Tsimshian 507
 Tunica 519
 Turkish 67, 70, 119, 154–155, 160–161, 190,
 206, 240, 245–247, 253, 339, 342, 387, 393,
 395–396, 405, 407, 413, 415–419, 425
 Turoyo Aramaic 414
 typological factors: in borrowing 186–188; in
 code-switching 188–190, 189; contact-induced
 typological change 193–194; critical issues
 and topics with 192–193; current contributions
 and research with 193–194; future directions
 with 194–195; historical overview of
 186–192; in mixed language formation
 190–192
 UD *see* Universal Dependencies (UD)
 Uisai 493
 Ukrainian 279
 Ulithian 487
 Unami 516
 Uniform Structure Principle (USP) 66
 Unitary System Hypothesis 202–203
 Universal Dependencies (UD) 12
 Universal Grammar 113

- universalist hypothesis 309
 Upper Chehalis 512
 Upper Inlet Tanaina 505–506
 Upper Sorbian 119
 urban multilingualism: lesser-used languages
 and 353–354; Linguistic Landscape and 349,
 352–353
 Urban Wolof 532
 Urban Youth Speech Styles (UYSS): Computer
 Mediated Communication and 376; critical
 issues and topics in 370–375, **373**; definitions
 in 367–369; future directions with 376–377;
 historical overview of 369–370; L2 stylization
 in 370–372; methodological matters in
 375–376; paralexical insertion in 372–374,
 373; social meaning of 374–375; Total
 Linguistic Fact concept and 375
 usage-based approaches: borrowing in 110, 116;
 code-switching in 110, 116; constructional
 borrowing in 119–120; critical issues and
 topics in 112–118; current contributions and
 research in 118–121; and ease of activation
 120–121; entrenchment in 110; equivalence
 and 117–118; frequency in 119; historical
 overview of 111–112; interference in 110;
 knowledge in 110; meaning in 118–119;
 modularity and 117; psycholinguistics and
 115–116; research methods in 121–122;
 sociolinguistics and 113–115; unit in 110;
 usage in 110
 usage frequency: borrowing and 245–246
 USP *see* Uniform Structure Principle (USP)
 UTF-8 13
 Utilitarian motivation 245
 UYSS *see* Urban Youth Speech Styles (UYSS)
 Valle d’Aostan French 76–77
 variationist perspective: analysis in 52–56;
 conflict sites in 53; constraint hierarchy in
 53; contact-induced change in 55–56; corpus
 in 51; cross-variety comparison in 52–56;
 current contributions and research in 56–57;
 defined 46; in Eastern Polynesia 466; future
 directions with 57–58; historical overview
 of 48; issues and topics in 48–49; language
 contact applications of 47; paradigm of 46–47;
 quantitative analysis in 52; quantitative
 reasoning in 52; research methods in 49–56;
 speakers in 50; speech community in 49–50;
 speech in 50–51
 Vedda 436
 verbs: in code-switching 68–70
 vernacular 50–51
 visual world paradigm 37
 WAPE *see* West-African Pidgin English (WAPE)
 Wappo 509
 Waskia 194, 494
 Welsh 211–212
 West Africa: convergence in 529–530, 538–541;
 critical issues and topics in 532–536, 534, **536**;
 current contributions and research in 536–541,
 538; defined 528–529; emblematicity in
 535–536; fragmentation and diffusion in 530;
 future directions with 541, 541–542; historical
 overview of 530–532, 531; language ecologies
 in 530–532, 531; language territorialization
 in 532–535, 534; reification in 536–538, **538**;
 research methods with 541, 541–542
 West-African Pidgin English (WAPE) 307–308
 Wolof 528, 534, 535, 540–541
 word count method 18
 Xhosa 70–72, 99
 Yenish 154
 Yiddish 160
 Yokuts 506
 youth *see* Urban Youth Speech Styles (UYSS)
 Yuman 510–511
 Zaza 416–417
 Zazaki 407, 408, 413, 422
 zooming effect 281