

V МЕЖДУНАРОДНАЯ КОНФЕРЕНЦИЯ  
ПО КОМПЬЮТЕРНОЙ ОБРАБОТКЕ  
ТЮРКСКИХ ЯЗЫКОВ  
«TURKLANG 2017»

Труды конференции

Том 2

КАЗАНЬ  
2017

УДК 004.8+81'32  
ББК 81.1

**Организаторы:**

**Академия наук Республики Татарстан**  
Институт прикладной семиотики

**Казанский (Приволжский) федеральный университет**  
Высшая школа информационных технологий  
и информационных систем  
Институт вычислительной математики  
и информационных технологий

**Евразийский национальный университет имени Л. Н. Гумилёва**  
**Министерства образования и науки Республики Казахстан**  
НИИ «Искусственный интеллект»

**Международная Тюркская академия**  
**Российская ассоциация искусственного интеллекта**

Издание осуществлено при финансовой поддержке  
Российского фонда фундаментальных исследований  
(проект №17-47-161033)

**Научные редакторы:**  
академик АН РТ, профессор, д.т.н. Д. Ш. Сулейманов,  
к.т.н. А. Р. Гатиатуллин

**Пятая Международная конференция по компьютерной  
обработке тюркских языков «TurkLang 2017».** – Труды кон-  
ференции. В 2-х томах. Т 2. – Казань: Издательство Академии  
наук Республики Татарстан, 2017. – 327 с.

ISBN 978-5-9690-0407-8

Сборник содержит материалы Пятой Международной конференции по  
компьютерной обработке тюркских языков «TurkLang-2017» (Казань, Татар-  
стан, Россия, 18–21 октября 2017 г.)

Для научных работников, преподавателей, аспирантов и студентов, спе-  
циализирующихся в области компьютерной лингвистики и ее приложений.

УДК 004.8+81'32  
ББК 81.1

ISBN 978-5-9690-0407-8

УДК 004.89

**BUILDING A BILINGUAL DICTIONARY OF POLYSEMOUS WORDS FOR THE MACHINE TRANSLATION OF THE KAZAKH LANGUAGE****U. Tukeyev** ualsher.tukeyev@gmail.com,**Z. Zhumanov** z.zhake@gmail.com,**A. Karibayeva** a.s.karibayeva@gmail.co,**D. Amirova** amirovatdina@gmail.com,**A. Sundetova** sun27aida@gmail.com,**B. Abduali** balzhanabdualy@gmail.com*Al-Farabi Kazakh National University, Almaty, Kazakhstan*

In this article we consider the creation of linguistic resources for Kazakh-English and Kazakh-Russian language pairs to find the correct translation in a specific context. These language pairs represent an interesting task of translating ambiguous words in the text. To perform this task, it is important to build linguistic data for the Kazakh language. One of such construction methods is based on large amounts of annotated data. These data are the linguistic basis for machine translation systems, such data can be in form of dictionaries, corpora. Quality of machine translation depends on the volume of language database and on the depth of description of natural languages. Dictionaries have a great influence on work of machine translation systems. An important problem of a dictionary is its transparency and limitations; it can not be absolutely complete because the lexical composition of a language is always subject to change. Thus, the process of creating a dictionary becomes infinite, as each dictionary can improve daily.

Nowadays there are many machine translation systems for the language pair. But the lack of reliable mechanisms for determining the meaning of a word reduces the accuracy (quality) of the translation since a word can have not one but a whole series of meanings in another language. And automatically determining the correct translation of words that depend on the context is a very difficult task. The solution of lexical disambiguation problem for a long time was conceived as the main task, the completion of which will make it possible to achieve almost perfect machine translation.

The solution of the disambiguation problem is an important task of machine translation and requires a large number of linguistic data. In this problem, the meanings (possible translations) of ambiguous words are taken from a bilingual dictionary. A bilingual dictionary is a special dictionary used to translate words or phrases from the source language into the target language. In addition to translation, a bilingual dictionary usually defines part of speech and other char-

acteristics of words. Dictionaries are a fundamental element in the development of the linguistic base of natural languages. Linguistic resources, such as bilingual languages, are the basis for obtaining the correct translation in rule-based machine translation such as Apertium. The formation of linguistic resources for machine translation requires a wide and complete presentation of linguistic information about the natural language that needs to be processed. To this date, there are different types of machine translation, such as rule-based, statistical, hybrid, and others. In all systems of machine translation, the linguistic database of languages plays a very important role. Rule-based machine translation uses linguistic information contained in dictionaries and grammars of the source language and the target language. Statistical machine translation is based on a collection of aligned parallel corpora and includes calculation of this translation's probability and choice of the most probable one. The number of parallel corpora for the Kazakh language is small in comparison with the number of a monolingual text corpus. In any machine translation system, linguistic data is expanded through translations with different senses of the source language into the target language. Currently, one of the main problems of natural language processing is ambiguity. This paper will consider the creation of linguistic resources for solution of the ambiguous translation task in machine translation based on the rules in the free/open source Aперium platform and the results of extracting the necessary data, that is, alternative translations of ambiguous words from parallel natural language corpora based on a bilingual dictionary, will be demonstrated.

**Keywords:** machine translation; linguistic data; ambiguity; bilingual dictionary; corpus; Apertium.

## **ФОРМИРОВАНИЕ ДВУЯЗЫЧНОГО СЛОВАРЯ МНОГОЗНАЧНЫХ СЛОВ ДЛЯ МАШИННОГО ПЕРЕВОДА КАЗАХСКОГО ЯЗЫКА**

*У. А. Тукеев*, ualsher.tukeyev@gmail.com,  
*Ж. М. Жуманов*, z.zhake@gmail.com,  
*А. С. Карибаева*, a.s.karibayeva@gmail.com,  
*Д. Т. Амирова*, amirovatdina@gmail.com,  
*А. М. Сундетова*, sun27aida@gmail.com,  
*Б. А. Абдуали*, balzhanabdualy@gmail.com

*Казахский национальный университет им. аль-Фараби, Алматы,  
Казахстан*

В этой статье мы рассмотрим создание двуязычных словарей для казахско-английской и казахско-русской языковых пар на платформе Апертиум. Эти языковые пары представляют интересную задачу перевода многозначных слов в тексте. Существует множество разных методов

формирования лингвистических данных, используемые в машинном переводе. Один из таких методов построения основывается на больших аннотированных данных. Эти данные являются лингвистической базой для систем машинного перевода, такими данными могут быть словари, корпуса. Качество машинного перевода зависит от объема языковой базы данных и от глубины описания естественных языков. Словари оказывают большое влияние на работу систем машинного перевода. Важной проблемой словаря является его прозрачность и ограниченность; он не может быть абсолютно полным, потому что лексический состав языка всегда подлежит изменению. Таким образом, процесс создания словаря становится бесконечным, поскольку каждый словарь может улучшаться ежедневно.

В наше время существует множество систем машинного перевода для пары языков. Но отсутствие надежных механизмов определения значения слова снижает точность (качество) перевода, так как слово может иметь не одно, а целый ряд значений на другом языке. И автоматическое определение правильного перевода слов, зависящих от контекста – очень трудная задача. Разрешение лексической многозначности с давних пор задумывалось как главная задача, решение которой позволит добиться почти идеального машинного перевода, в котором двуязычный словарь играет важную роль.

Решение задачи многозначности – важная задача машинного перевода и требует большого количества лингвистических данных. В этой задаче смыслы (возможные переводы) многозначных слов берутся из двуязычного словаря. Двуязычный словарь – это специальный словарь, используемый для перевода слов или словосочетаний с исходного языка на целевой язык. В дополнение к переводу двуязычный словарь обычно определяет часть речи и другие характеристики слов.

На сегодняшний день существуют разные виды машинного перевода, такие как основанные на правилах, статистический, гибридный и другие. Во всех системах машинного перевода немало важную роль играет лингвистическая база данных языков.

Машинный перевод, основанный на правилах, использует лингвистическую информацию, содержащуюся в словарях и грамматиках исходного языка и целевого языка. Статистический машинный перевод основан на коллекции выровненных параллельных корпусов и включает в себя вычисление вероятности данного перевода и выбор наиболее вероятного. Количество параллельных корпусов для казахского языка невелико по сравнению с количеством одноязычного текстового корпуса. В любой системе машинного перевода лингвистические данные расширяются за счет многозначных переводов исходного языка на целевой.

Словари являются основополагающим элементом в разработке лингвистической базы естественных языков. Лингвистические ресурсы, такие как двуязычные языки, являются основой для получения правильного перевода в машинном переводе основанным на правилах, таких как Апертиум.

Платформа Апертиум насчитывает около 40 автоматических переводчиков и еще больше находится в процессе разработки. Команда Апертиум проявляет большой интерес к региональным языкам и активно поддерживает работу по созданию новых систем машинного перевода с лингвистической базой данных словарями и правилами, указанными в формате XML.

Формирование лингвистических ресурсов для машинного перевода требует широкого и наиболее полного представления лингвистической информации о естественном языке, который требуется обработать.

В настоящее время, одной из основных проблем обработки естественного языка является многозначность. В данной статье будет рассматриваться создание многозначных двуязычных словарей для решения задачи лексического выбора в машинном переводе, основанный на правилах в открытой/свободной платформе Апертиум и будут продемонстрированы результаты извлечения слов или лемм, то есть альтернативных переводов многозначных слов из параллельных корпусов естественного языка и двуязычного словаря.

**Ключевые слова:** машинный перевод; лингвистические данные; многозначность; двуязычный словарь; корпус; Апертиум;

## 1. Введение

Одной из основных проблем обработки естественного языка является многозначность, в которой немаловажную роль играют словари. Проблема разрешения многозначности как отдельная задача была сформулирована еще в конце 40-х годов XX века, практически одновременно с появлением машинного перевода. Начиная с того времени было разработано немало методов решения данной задачи, тем не менее она остается актуальной и по нынешний день.

Во всех развитых языках присутствуют как однозначные, так и многозначные слова. Способность слов выступать лишь в одном значении называется однозначностью или моносемией. Примеры таких слов: «бинокль», «троллейбус», «suitcase», «pouu». Однако большинство слов имеют не одно, а несколько значений. Они называются многозначными или полисемантическими. Способность лексических единиц иметь несколько значений называется многозначностью или полисемией.

В данной статье будут показаны методы для формирования двуязычного словаря Апертиума, используемые при решении задачи лексического выбора при комбинированной технологии в

англо-казахском, казахско-английском, русско-казахском и казахско-русской языковой пары на платформе Апертиум.

Двуязычные словари многозначных слов являются частью общих двуязычных словарей языковых пар. Так как проблема многозначности слов существенно сказывается на качестве машинного перевода, вопрос полноты двуязычных словарей многозначных слов является очень актуальным.

В данной работе рассматривается комплексная технология формирования двуязычных словарей многозначных слов машинного перевода с использованием как существующих электронных двуязычных словарей, так и формирование двуязычных словарей многозначных слов через обработку параллельных двуязычных корпусов.

## 2. Обзор работ

Формирование лингвистических ресурсов имеет большое значение при приобретении знаний. Существуют некоторые подходы к извлечению лингвистических данных из корпусов и двуязычных словарей.

Многие подходы касаются автоматического приобретения лингвистических ресурсов для машинного перевода (МП). Самый ранний метод для МП просто включал двуязычный словарь, в котором каждое слово было просмотрено и перенесено на его эквивалент на целевом языке. Этот подход полезен для перевода фраз, но не для предложений (Victor M. Sanchez-Cartagena, Miquel Espla-Gomis и др.)

Во (Michael Rosner, Kurt Sultana, 2014) путем объединения двух методов для расширения двуязычного словаря авторы получают верные словарные записи. Они используют два метода для расширения словаря из сопоставимых корпусов. Первый метод – это вектор контекста, а второй – метод эскиза. Последний метод показывает более высокую точность, но он очень чувствителен к размеру корпуса. В то время как векторный метод контекста хорошо работает даже с небольшими корпусами по сравнению с методом эскиза и дает лучшие результаты.

Метод обогащения двуязычных окончательных переводов без каких-либо предварительных знаний представлен в (Ajay Dubey и др.). Авторы представили автоматический метод создания межъязычного словаря, который был сгенерирован небольшим количе-

ством текстов из средств массовой информации, при использовании сопоставимых пар.

Существует много методов автоматической генерации двуязычных словарей. Эти методы создают двуязычные словари, извлекая двуязычные словосочетания из корпусов.

Создание двуязычных словарей для определенных областей, определенных параллельным корпусом, описаны в (Angelina Ivanova и др). Авторы предлагают метод, который состоит из 4 шагов. Первый шаг – это выравнивание предложений, который устанавливает эквивалентности между предложениями параллельных текстов. Следующий шаг – морфологический анализ, собирающий морфологическую информацию для слов. Идентификация эквивалентности между одиночными и многословными выражениями в параллельных предложениях выполняется на третьем этапе – лексическое выравнивание. Четвертый шаг отвечает за выбор более релевантных записей для словаря ().

Авторы (L. Hilgert, L. Lopes и др., 2012) показали, что двуязычные словари также могут быть изучены или обогащены из неприсоединившихся сопоставимых корпусов. Они предлагают комбинированные лингвистические и статистические методы. В статье описана двухэтапная модель перевода

Двуязычной терминологии из сопоставимых корпусов, неоднозначности и выбора лучших вариантов перевода на основе их морфологических знаний.

Турдаков Д. Ю. (2010) использовал названия статей, которые используются для поиска терминов в тексте для создания словаря. После всех терминов, найденных в тексте, они представлены в виде серии наблюдений и значений, соответствующих состояниям в скрытом марковской модели. Для создания словаря терминов они использовали имена всех статей, описывающих соответствующие понятия и имена всех страниц, перенаправляемых к статье. В статье будут рассмотрены методы по формированию списка многозначных слов по частотности по корпусу.

### **3. Описание методов формирования двуязычного словаря**

Технология формирования лингвистических данных подразумевает несколько этапов создания данных со словарей и корпусов.



Для формирования многозначного словаря используются следующие данные:

- Двухязычный словарь данной языковой пары (apertium-eng-kaz.eng-kaz.dix);

Для других языковых пар используется такой же двухязычный словарь в формате .dix. Эти файлы написаны на языке XML, форматы заполнения записей приведены ниже.

- Параллельные корпуса исходного и целевого языка;

Параллельные корпуса для казахского языка собирались из разных источников, таких как электронные сайты правительства Республики Казахстан, художественной литературы, библии и т.д. На сегодняшний день имеется около 26000 предложений в параллельных корпусах для англо-казахского(и обратно) языка, для русско-казахского(и обратно) первоначально было собрано около 2500 предложений.

Второй подход формирования многозначного словаря реализуется из существующего двухязычного словаря системы Апертиум.

Шаги для формирования многозначного словаря из двухязычного словаря( файл с расширением.dix) следующие:

- Пройтись по двухязычному словарю от начала до конца;
- Нахождение леммы исходного языка, которые имеют несколько переводов в целевом;
- Извлечение слова с лексической многозначностью перевода из словаря.

Для формирования из корпусов аналогично извлечению из словаря и корпусов, и включает следующие этапы:

- Сделать список частотных слов из одноязычного корпуса исходного языка;
- Выбор слова с множеством перевода из списка частот, используя двухязычный словарь;
- Создать многозначный словарь с использованием словаря;
- Найти перевод заданных многозначных слов в параллельных корпусах;
- Добавить перевод из целевого корпуса в словарь, при его отсутствии.

Задача перевода многозначного слова требует много ручной работы в корпусах и словарях. Все переводы с параллельного корпуса будут добавлены к двухязычному словарю с описанием частей речи.

При формировании частотного списка переводов корпуса были очищены от слов, не имеющих важность при переводе, такие как цифры, знаки препинания и т.д, то есть “стоп” слов. Список наиболее часто встречающихся форм слов был отсортирован и сокращен до основ леммы при помощи скриптов, в результате чего был составлен список наиболее часто встречающихся лемм. Лемма – это каноническая форма слова, слово без грамматической информации. Например, леммой слова “cats” является “cat”. В английском языке лемма рассматриваемого существительного как правило совпадает с его формой единственного числа. В русском и казахском языках лемма существительного имеет вид формы именительного падежа единственного числа. Для глаголов в английском языке лемма будет иметь вид инфинитива без “to” (или просто инфинитива в русском и казахском языках). Например, леммой слова “was” будет “be”, также как леммой слова “был” будет “быть”, или леммой слова “болды” будет “бол”.

Слова были добавлены в двуязычный словарь вручную, как описано ниже. Двуязычный словарь: содержит переводные соответствия слов и символов двух языков. Он называется так: **apertium-eng-kaz.eng-kaz.dix**.

Если слово имеет несколько переводов, то пишутся все альтернативные переводы:

```
<e><p><l>sister<s n=>n</l><r>сінді<s n="n"></r></p></e>
<e><p><l>sister<s n="n"></l><r>әпке<s n="n"></r></p></e>
<e><p><l>sister<s n="n"></l><r>қарындас<s n="n"></r></p></e>
```

В данном случае, слово «sister» на казахском языке имеет 3 перевода. Один из них выбирается позже, посредством лексической выборки. Лексический выбор – это нахождение правильного перевода по контексту, в котором слово исходного языка имеет несколько переводов, которые относятся к одной части речи. Важность задачи разрешения лексической многозначности все же очень велика. В том числе оно может быть полезным целому ряду приложений. Сформированный многозначный двуязычный словарь используется в разных сферах: для улучшения качества машинного перевода, повышения точности методов классификации и кластеризации текстов, в информационном поиске и других приложениях, а также мы используем этот словарь в решении задачи лексического выбора на основе комбинированной технологии, реализующая через правила и статистику.

### *3.1 Извлечение списка частотных слов с проверкой с помощью «традиционного» словаря*

Сделав список частот из одноязычного корпуса, мы стали искать среди них неоднозначные слова. В частности, мы искали слова с множеством переводов, относящиеся к одной части речи. Все возможные переводы были рассмотрены из «традиционного» словаря. Под «традиционным» словарем мы подразумеваем онлайн-словарь с большим объемом данных, таких как *lugat.kz*, *sozdik.kz* и т. д. Существует не так много словарей, которые переводят с и на казахский язык с разными вариантами переводов для многозначных слов.

*Lugat* – это комплекс словарей для нескольких языков. В словарях для англо-казахских и казахско-английских языковых пар содержится 113 318 слов для каждого направления. Словари для казахско-русского и русско-казахского языков содержат 95 118 слов для каждого направления. *Создик* – казахско-русский и русско-казахский онлайн-словарь.

Следующий шаг после выбора многозначных слов и нахождения переводов – заполнение двуязычного словаря не имеющих в двуязычном словаре.

Инструменты и ресурсы, которые использовались в работе:

- *apertium-eng-kaz.eng-kaz.dix* – двуязычный казахско-английский и англо-казахский словарь
- *apertium-kaz-rus.kaz-rus.dix* – двуязычный казахско-русский и русско-казахский словарь
- *Intertext* – программное обеспечение для управления параллельными текстами
- Инструменты Апертиума для морфологического анализа и лемматизации.
- *Sublime Text* – редактор с богатыми возможностями преобразования текста.

Для англо-казахских и казахско-русских языковых пар есть одноязычные и двуязычные словари на платформе Апертиум.

*InterText* использовался для работы с параллельными текстами на английском, казахском и русском языках. *InterText* – инструмент управления для выровненных параллельных текстов. Он помогает редактировать и управлять выравниваниями нескольких версий текста на нескольких языках на уровне предложений, а

также поддерживает пользовательские XML-документы и набор символов Юникод(<http://wanthalf.saga.cz/intertext>).

Инструмент в основном используется лингвистами и экспертами, которые занимаются работой с текстами, особенно с выравниванием текстов, с которыми трудно справиться без глубоких знаний и технических навыков.

#### 4. Результаты

В статье описаны методы формирования многозначного двуязычного словаря необходимых для решения задачи лексической многозначности на свободной/открытой платформе Апертиум. Работа по созданию словаря необходима для реализации комбинированной технологии лексического выбора в Апертиуме, на стадии формирования семантического куба для каждого многозначного слова. Семантический куб – это многомерная таблица данных для каждого многозначного слова с частотой появления с определенным контекстом из параллельного корпуса. Размеры словаря для англо-казахских (и наоборот) и русско-казахских составляют ~ 21 165 и ~ 12 923 записей соответственно. Результаты формирования многозначного словаря исходных языков казахского, русского и английского языка показаны ниже в таблице 1.

*Таблица 1. Результаты по созданию списка многозначных слов из словарей и корпусов*

	Объем параллельного корпуса	Объем двуязычного словаря Апертиум	Количество полученных слов из словаря Апертиума	Количество полученных слов из параллельного корпуса
English-Kazakh	25 324	21 165	432	197
Kazakh-English	25 324	21 165	748	246
Russian-Kazakh	2 256	12 923	91	71
Kazakh-Russian	2 256	12 923	87	63

Данные таблицы 1 показывают формирование многозначного двуязычного словаря из параллельных корпусов и общего двуязычного словаря Апертиума. Количество полученных многозначных слов с двуязычного словаря больше чем количество изъятых из корпуса, так как в параллельном корпусе предложения из официальных сайтов государства, в которых многозначность перевода присутствует в очень малой степени. С увеличением записей в словаре будет повышаться качество и точность перевода.

## 5. Заключение

Формирование лингвистических данных, а именно многозначного словаря будет представлять интерес для лингвистов, специалистам по прикладной и вычислительной лингвистике, и другим специалистам, интересующихся увеличением качества машинного перевода, информационного поиска и других приложений.

В этой статье был представлен подход для извлечения многозначных слов из разных лингвистических ресурсов: параллельных корпусов и словаря. Полученные лингвистические данные могут использоваться в различных задачах обработки языка, таких как задача лексического выбора в машинном переводе. Формирование более полного двуязычного словаря многозначных слов позволит повысить качество систем машинного перевода казахско-английской и казахско-русской языковых пар. Несомненно, данная тема актуальна и требует продолжения исследований в будущем.

## ЛИТЕРАТУРА

1. Victor M.Sanchez-Cartagena, Miquel Espla-Gomis, Juan Antonio Perez-Ortiz. Source-Language Dictionaries Help Non-Expert Users to Enlarge Target-Language Dictionaries for Machine Translation.

2. Michael Rosner, Kurt Sultana. Automatic Methods for the Extension of a Bilingual Dictionary using Comparable Corpora. LREC 2014 :3790-3797.

3. Ajay Dubey, Parth Guptay, Vasudeva Varma, Paolo Rossoy. Enrichment of Bilingual Dictionary through News Stream Data. LREC 2014 :3761-3765.

4. Khang Nhut Lam and Feras Al Tarouti and Jugal Kalita. Automatically Creating a Large Number of New Bilingual Dictionaries.

5. Angelina Ivanova. Charles University in Prague Faculty of

Mathematics and Physics M.Sc. Thesis. Creating a Bilingual Dictionary using Wikipedia.

6. L. Hilgert, L. Lopes, A. Freitas, R. Vieira, D. N. Hogetop, A. A. Vanin. Building Domain Specific Bilingual Dictionaries. LREC 2012 : 2772-2777.

7. Sadat, F.; Yoshikawa, M.; Uemura, S.: Learning Bilingual Translations from Comparable Corpora to Cross-Language Information Retrieval: Hybrid Statistics-based and Linguistics-based Approach. In In Proc. IRAL 2003, 2003.

8. <http://wanthalf.saga.cz/intertext>.

9. Vondříčka, Pavel. Aligning parallel texts with InterText. LREC 2014:160-164.

10. Турдаков Д. Ю. Методы и программные средства разрешения лексической многозначности терминов на основе сетей документов. Москва, 2010. Thesis.

**СЕКЦИЯ 5. МАШИННЫЙ ПЕРЕВОД**

МОДЕЛИРОВАНИЕ СЛОВСОЧЕТАНИЙ ПО ЧАСТЯМ РЕЧИ В ПРОЦЕССЕ АНГЛО-УЗБЕКСКОГО МАШИННОГО ПЕРЕВОДА. <i>Н. Абдурахмонова</i> . . . . .	246
АЛГОРИТМ, ОСНОВАННЫЙ НА ЛИНГВИСТИЧЕСКОЙ МОДЕЛИ АНГЛО-УЗБЕКСКОГО МАШИННОГО ПЕРЕВОДА. <i>Н. Абдурахмонова, Х. Ахмедова</i> . . . . .	255
ОЦЕНКА КАЧЕСТВА МАШИННОГО ПЕРЕВОДА В СЦЕНАРИИ АССИМИЛЯЦИИ ДЛЯ АНГЛО-КАЗАХСКОЙ И КАЗАХСКО-РУССКОЙ ЯЗЫКОВЫХ ПАР. <i>Ж.М. Жуманов, Д.Т. Амирова</i> . . . . .	263
СИНТАКТИКО-СЕМАНТИЧЕСКИЕ ПРЕДСТАВЛЕНИЯ ДЛЯ РАЗВИТИЯ КАЗАХСКО-РУССКОГО ФРАЗЕОЛОГИЧЕСКОГО МАШИННОГО ПЕРЕВОДА. <i>Ж. Мейрамбеккызы, А.А. Хорошилов</i> . . . . .	274
О РАЗЛИЧНЫХ ПОДХОДАХ К МАШИННОМУ ПЕРЕВОДУ С РУССКОГО НА КАЗАХСКИЙ ЯЗЫК. <i>А. Макажанов, Б. Мырзахметов, Ж. Кожирбаев</i> . . . . .	288
ФОРМИРОВАНИЕ ДВУЯЗЫЧНОГО СЛОВАРЯ МНОГОЗНАЧНЫХ СЛОВ ДЛЯ МАШИННОГО ПЕРЕВОДА КАЗАХСКОГО ЯЗЫКА. <i>У.А. Тукеев, Ж.М. Жуманов, А.С. Карибаева, Д.Т. Амирова, А.М. Сундетова, Б.А. Абдуали</i> . . . . .	303
МАТЕМАТИЧЕСКИЕ МОДЕЛИ АНГЛИЙСКОГО ЯЗЫКА ДЛЯ СИСТЕМЫ МНОГОЯЗЫЧНЫХ СИТУАЦИЙ МАШИННОГО ПЕРЕВОДА. <i>М.Х. Хакимов</i> . . . . .	315

---

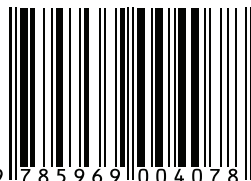
V МЕЖДУНАРОДНАЯ КОНФЕРЕНЦИЯ  
ПО КОМПЬЮТЕРНОЙ ОБРАБОТКЕ  
ТЮРКСКИХ ЯЗЫКОВ  
«TURKLANG 2017»

Труды конференции

Т о м 2

*В авторской редакции*

Подписано в печать 28.12.2017. Формат 60×84<sup>1</sup>/<sub>16</sub>.  
Усл. печ. л. 19,06. Тираж 100 экз.



9 785969 1004078