
TurkLang • 2015
Tatarstan Academy of Sciences
PROCEEDINGS
OF THE INTERNATIONAL CONFERENCE
"TURKIC LANGUAGES PROCESSING"
TurkLang
2015
September 17-19, 2015, Kazan, Tatarstan, Russia
Kazan, 2015

PROCEEDINGS
OF THE INTERNATIONAL CONFERENCE
"TURKIC LANGUAGES PROCESSING"

TurkLang-2015

September 17–19, 2015, Kazan, Tatarstan, Russia

Kazan
2015

СОДЕРЖАНИЕ

SECTION 1. **MACHINE TRANSLATION TECHNOLOGIES**

SECTION 2 . **LINGUISTIC SOFTWARE**

RESEARCH OF PROBLEM OF THE SEMANTIC ANALYSIS AND SYNTHESIS OF PRETEXTS IN THE RUSSIAN-KAZAKH MACHINE TRANSLATION

Diana Rakhimova

Al-Farabi Kazakh National University
Almaty, Kazakhstan

Kazakh and Russian are very difficult languages and not so similar on lexical and syntactic structure. At creation of the Russian-Kazakh machine translation had identified difficulties. This article offers the problem of the semantic analysis of pretexts of Russian and synthesis to Kazakh language and vice versa. At comparison of pretexts and communications for two languages defined the characteristic properties and rules of transformation based on lexical, syntactic and semantic analysis of sentences.

Введение

Предлог – это служебная часть речи, которые необходимы для связи слов в словосочетании. Предлоги выражают зависимость одних слов от других, и могут быть при существительных, местоимениях и числительных. Они имеют различные виды образования, структуры и смысловые значения (Рубашкин В.Ш).

При переводе роль предлогов русского языка в казахском языке выполняют аффиксы и вспомогательные слова. И если предлоги русского языка стоять перед определяемым словом, то в казахском языке после: *Уехал на два месяца – екі айға кетіп қалды. Работать до вечера- кешке дейін жұмыс істеу;* Вроде бы никаких проблем с переводом предложных связей, т.е. определяется предлог в тексте из словаря и при переводе должен будет расположен после определяемого слова. Но при разработке машинного перевода (МП) с русского на казахский язык столкнулись с многозначностью предлогов в контексте предложения. В разных лексических конструкциях и с разными падежами предлоги могут иметь разные значения. Например: *забыл на столе* (пространственное значение), *отлучился на минуту* (временное значение), *верить на слово* (значение образа действия). Выше предложный метод перевода может быть не достаточен для полноты текста, т.к. имеет только стуктурную характерность в синтаксическом анализе и генерации предложении в машинном переводе. В данной работе

будет рассмотрены предложные связи, которые имеют семантические свойства.

При разработке русско-казахского словаря предлоги были разделены на две группы:

1) однозначные – предлоги и словообразования у которых есть определенный точный перевод: *до завтра -ертеңге дейін, в течение года – жыл бойы, через один час – бір сағаттан кейін, накануне наурыза- наурыз қарсаңында*;

2) многозначные – предлоги и словообразования которые имеют несколько значении перевода: *в декабре – желтоқсанда, в доме-үйдің ішінде. под вечер – кешке қарай, под стол – үстел астында.*

Синтаксическое и семантическое описание предлогов практически значимо как раздел машинного словаря. С этой точки зрения состав словарного описания предлогов должен определяться исключительно востребованной алгоритмами анализа функциональностью словаря.

1. Модель структуры предложений русского и казахского языков

Формальные модели синтаксиса простых предложений казахского языка ориентированы на выделение трех типов фразовых структур: субъектной фразовой структуры(SP), глагольной фразовой структуры(VP) и объектной фразовой структуры(OP). Основой субъектной фразовой структуры является подлежащее, основой глагольной фразовой структуры является глагол и основой объектной фразовой структуры является объект действия. ниже представлена формальные модели синтаксиса простых предложений казахского языка с использованием аппарата формальных грамматик ().

С использованием нотации Бэкуса формальная модель структуры синтаксиса предложений казахского и русского языков будет иметь следующий вид.

S::= <SP><OP><VP>|<SP><VP><OP>|<OP><SP><VP>|<OP><SP><VP>|

<VP><SP><OP>|<VP><OP><SP>

(Данное правило представляет всевозможные варианты структур на уровне введенных фразовых структур)

<SP>::=<N>|<Adj><SP>|<Num><SP>|<N><SP>
<SP>::=<SP><Conn><SP>
<VP>::=<V>|<Aux><VP>|<Adv><VP>
<OP>::=<N>|<Adj><OP>|<OP><Conn><OP>

Здесь Adj –прилагательное, Num – числительное, Conn – союзы, Aux – вспомогательные глаголы, Adv – наречие.

Для русского языка добавляется правило с учетом предлогов:

<OP>::= <Ртер><OP>|

Вышеуказанная модель преобразований структур предложений русского языка в структуры предложений казахского языка и наоборот используются при создании системы МП. По этим моделям разработаны алгоритмы, и разработана программа генератора

2. Семантический анализ и синтез предлогов в системе машинного перевода

Семантическая интерпретация предлога должна рассматриваться как частный случай более общей задачи – задачи семантической интерпретации синтаксических связей. Если между словами (в общем случае – текстовыми элементами) *W1* и *W2* парсером обнаружена синтаксическая связь, ставится вопрос о ее семантическом свойстве. В случае предложной связи вопрос может быть сформулирован так же, но в этом случае речь идет о конструкции вида *W1*---> *P*---> *W2*, где *W1* – (возможный) синтаксический хозяин, *W2* – синтаксический слуга, а предлог *P* маркирует связь между *W1* и *W2*, которая и является объектом семантической связи.

Задача машинного понимания текста сводится к переводу с естественного языка на язык представления знаний (ЯПЗ), в котором точно описаны правила построения и правила вывода (Рубашкин).

Синтезом предлогов (если он однозначный) является его перевод и структурное преобразование на целевой язык. Отношение предлогов рассматривается некой смысловой связью между объектами, которую при синтаксическом и семантическом анализе преобразуется в ЯПЗ. При генерации на целевой язык информация будет считываться с промежуточного ЯПЗ. Для предложных связей аксиомами могут быть некий набор правил, по которым может быть разрешена многозначность предлогов в МП. К отдельным

экземплярам относятся все предложные связи и словосочетания с предлогами, которые не поддаются аксиомам и правилам отношении. Это может быть художественные выражения, литературное высказывание или фразеологизмы. И такие отдельные экземпляры имеет свой полный смысловой перевод на выходной язык.

Так как предлог является корневым элементом своей предложной группы, были выделены его прямые аргументы из соответствующего поддерева семантического дерева предложения. Для обобщения переводной формулы предлога его аргументы были заменены семантическими классами атрибутов и грамматическими признаками.

Надо учитывать что организация связей и словоформ казахского языка отлична от русского языка. При переводе на казахский язык в различных случаях можно использовать определенный вид синтеза. Предлоги и предложные связи в казахском языке преобразуется с помощью присоединении аффиксов к основе слова (вариант 1) и/или вспомогательного слова стоящие после определяемого (вариант 2).

$$<pw_i> ::= <w_j> | <w_j\ w^*_{j+1}>$$

Где p – предлог, w_i – определяемое слово входного языка, w_j – генерируемое слово выходного языка, w^*_{j+1} – вспомогательное слово. Разработан полный анализ предлогов русского языка и их преобразование на казахский язык. В таблице 1 проиллюстрированы примеры преобразования предлогов на казахский язык.

Таблица 1

Структурное соответствие предлогов в русском и казахском языке

Представление предлогов в русском языке	Преобразование предлогов на казахский язык. (1 вариант)	Преобразование предлогов на казахский язык. (2 вариант)	Пример
На сущ(ед\м.р) + е	зат+ жат.с.(да.де..)	Зат+төуел үстінде	На стол+е

На сущ(ед\ж.р) + е	зат+жат.с.(да.де..)	Зат+тәуел үстінде	На книг+е
На сущ(ед\с.р) + е	зат+жат.с. (да.де..)	Зат+тәуел үстінде	На окн+е
На сущ(ед\ж.р) + у	зат+бар.с. (ға.ге..)	Зат+тәуел үстіне	На книг+у
На сущ(ед\м.р) + у	зат+жат(да.де..)		на берег+у
На сущ(ед\м.р)	зат+бар.с. (ға.ге..)	Зат+тәуел үстіне	на стол
На сущ(мн\ж.р) + ах	зат+көп+ жат.с. (да.де..)	Зат+ көп +тәуел үстінде	на книг+ах
На сущ(мн\м.р) + ах	зат+көп+ жат.с. (да.де..)	Зат+ көп +тәуел үстінде	На стол+ах
На сущ(мн\с.р) + ах(ях)	зат+көп+ жат.с. (да.де..)	Зат+ көп +тәуел үстінде	На окн+ах(ях)
У сущ(ед\с.р)+а(я)	Зат+жат.с. (та.те..)	–	У окн+ а(мор+я)
У сущ(ед\ж.р)+ы	Зат+жат.с. (та.те..)	–	У ват+ы
У сущ(мн\ж.р)	Зат+ көп+ жат.с. (та.те..)	–	У книг
У сущ(мн\м.р)+ов	Зат+ көп+ жат.с. (та.те..)	–	У стол+ов
У сущ(мн\с.р)	Зат+ көп+ жат.с. (та.те..)	–	У окон
У сущ(мн\с.р)+ей	Зат+ көп+ жат.с. (та.те..)	–	У мор+ей
...			

К примеру переведем следующие простое предложение: “*Книга лежит на столе*» данный пример с предлогом можно перевести

в двух вариантах «*кітап үстелде жатыр*» или «*кітап үстелдің үстінде жатыр*». Конечно, тот или иной вариант не ошибочен и может использован по усмотрению пользователя. Для многозначных предлогов метод трансформации предложных связей на казахский язык не всегда можно описать с помощью синтаксической маркировки. В данном случае значение надо выбирать по контексту предложения. Например «*я пришел под вечер*» предлог «*под*» обычно описывает место и переводится как «*астында*», но в данном смысловое значение указывает время выполнения действия и должно иметь следующий перевод – «*мен кешке қарай келдім*». Основываясь на исследованиях семантических отношениях и многозначных предлогов была создана сопоставительная семантико-синтаксическая структура предложных связей русского и казахского языка.

Рассмотрим примеры с многозначными предлогами и определим их семантические отношения:

в декабре → желтоқсанда;
в доме →үйдің ішінде;
под вечер→кешке қарай;
под стол → үстел астында;

В примерах с предлогм «*в*» мы видим, что по лексическим и морфологическим признакам отличить разницу трудно, т.к. определяемые слова являются одной частью речи (имя существительное) и имеют одинаковое окончание «*е*». Но по смысловому значению первый пример описывает время, а второй – место. В реализации морфологического анализа и синтеза машинного переводчика словари будут оснащены дополнительными грамматическими и семантическими свойствами. И будут маркированы в базе данных для удобства реализации.

На этапе семантического анализа текста (Тукеев У.А., Рахмнова Д.Р. 2012) будут определены семантические атрибуты определяемого слова с предлогом, с помощью которых будут применены семантические правила. Для слов «*декабрь*» и «*вечер*» будут определены семантический атрибут времени и для фраз с помощью семантических ролей (правил) правильно определено смысловое отношение и корректно сгенерировано на выходной язык. В таблице 2 проиллюстрированы некоторые варианты.

Таблица 2

Примеры многозначных предлогов русского языка и их преобразования на казахский язык

Структура предложенных связей в русском языке	Смысло-вое значение	Структуры преобразования предлогов на казахский язык
На <w_i>	место	w_j + жатыс септік жалғауы(да,де..)
		w_j + тәуелдік жалғау *үстінде* (көмектес сөз)
	время	w_j + барыс септік жалғауы (ка,ке...)
Под <w_i>	место	w_j + тәуелдік жалғау *астында* (көмектес сөз)
	время	w_j *қарай* (көмектес сөз)
За <w_i>	место	w_j +тәуелдік жалғау *артында* (көмектес сөз)
	время	w_j +жатыс септік жалғауы (та,те..)
	цель	w_i *үшін* (көмектес сөз)

Приведем пример практического применения:

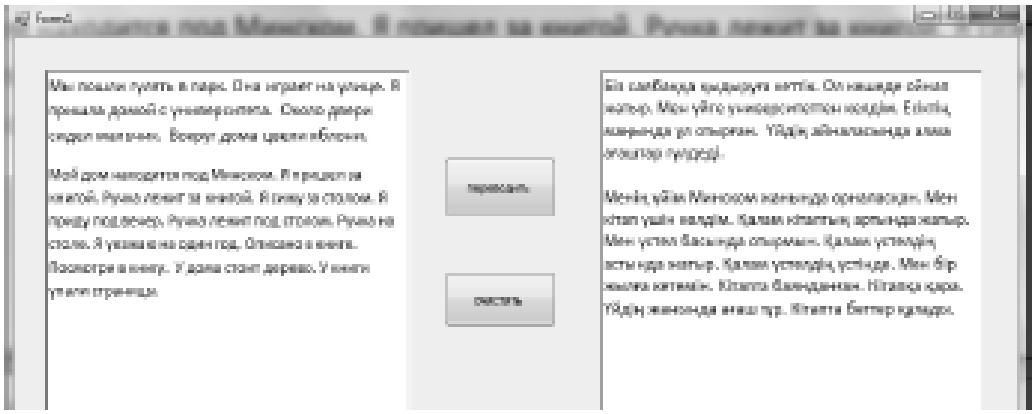

Рис. 1. Результат машинного перевода предложений с русского на казахский язык с многозначными предлогами

Как можно видеть на рисунке, было корректно определено смысловое значение предлогов и корректно подобрано соответствие на казахский язык.

Заключение

В данной работе была исследована проблема предлогов при машинном переводе. При анализе были учтены все возможные вариации слово изменений (часть речи, род, число, склонение, окончание и др.) в предложных связях, а так же были найдены и распознаны семантические свойства. Разработаны модель и алгоритмы анализа простых и многозначных предлогов русского языка с учетом характеристик синтеза на казахский язык. Реализована программа системы русско-казахского машинного перевода для простых предложений с предлогами.

ЛИТЕРАТУРА

Рубашкин В.Ш., *Семантической интерпретации предложных связей* [Электрон.ресурс].- url:http://www.dialog-21.ru/

Tukeyev U. Rakhimova D.R. (2012). *Augmented attribute grammar in meaning of natural languages sentences*. SCIS-ISIS 2012 The 6th International Conference on Soft Computing and Intelligent Systems. The 13th International Symposium on Advanced Intelligent Systems, (November 20–24), Kobe, Japan, 2012, – P. 1080–1084.

Рахимова Д.Р. (2014). *Построение семантических отношении в машинном переводе*. Вестник КазНУ, №1, 2014, – С. 90–101.

Тукеев У.А. *Разработка эффективных технологии компьютерного перевода казахского языка на английский и русский языки (и обратно) на основе методов формальных грамматик и статистических методов*: отчет о НИР (заключительный)/ ДГПНИИ ММ при КазНУ им аль-Фараби: рук. Тукеев У.А. Алматы, 2014. – 189 с. – № ГР 0112РК01467.

Jurafsky D., Martin J. *Speech and language processing: an introduction to nature language processing, computational linguistics, and speech recognition*. Pearson, Prentice hall. 2009, 988 p.

Кузнецов И.П., Сомин Н.В. (2010) *Особенности лексико-морфологического анализа при извлечении информационных объектов и связей из текстов естественного языка* [Электрон. ресурс]. – URL: http://www.dialog-21.ru/digests/dialog2010/materials/html/40.html.