

2. Esplà-Gomis M., Sánchez-Cartagena V.M., Pérez-Ortiz J.A., Sánchez-Martínez F., Forcada M.L., Carrasco R.C. An efficient method to assist non-expert users in extending dictionaries by assigning stems and inflectional paradigms to unknown words. //Proceedings of EAMT 2014, The Seventeenth Annual Conference of the European Association for Machine Translation (Dubrovnik, June 16–18, 2014). – 2014. – P.19–26.

3. Abduali B., Sundetova A., Zhanbussunov N., Musabekova Zh., Study of the problem of creating structural transfer rules for the Kazakh-English and Kazakh-Russian machine translation systems on Apertium platform. //Вестник Казну им. Аль-Фараби, том 20 (Messenger of Al-Farabi KazNU, vol. 20) by proceedings of the International Conference “Computational and Informational Technologies in Science, Engineering and Education” (CiTech-2015, 24-27 September, 2015). - Almaty: Al-Farabi Kazakh National University Press, 2015. - P.77-82

4. Рахимова Д.Р., Қалдашбеков Е.Е., Мұсабекова Ж.Ф., Абақан М., Кызырканова С., Жамалиева А., Абдуали Б. Apertium платформасы негізінде орыс тілінен қазақ тіліне аудару жүйесін құру. //Материалы международной научной конференции студентов и молодых ученых "Фараби Әлемі», Алматы, Қазақстан, 13-16 апреля, 2015 г. - Алматы, 2015. - С.175.

УДК 621.3

ОНТОЛОГИЧЕСКАЯ МОДЕЛЬ ИМЕНИ СУЩЕСТВИТЕЛЬНОГО ДЛЯ СИСТЕМЫ КАЗАХСКО-КИТАЙСКОГО МАШИННОГО ПЕРЕВОДА

Unzila Kamanur Евразийский Национальный Университет им.Л.Н.Гумилева, Астана, 010000, Қазақстан Unzila.88@mail.ru

Алтынбек Шарипбай Евразийский Национальный Университет им.Л.Н.Гумилева, Астана, 010000, Қазақстан

Гүлмира Бекманова Евразийский Национальный Университет им.Л.Н.Гумилева, Астана, 010000, Қазақстан

Лена Жеткенбай Евразийский Национальный Университет им.Л.Н.Гумилева, Астана, 010000, Қазақстан

Цель статьи - В этой работе будет рассмотрен создание онтологической модели имен существительных казахского, русского языков, предназначенных для системы машинного перевода. Эта модель даст нам возможность сравнить сходства и различие двух языков. Также можно применить для разработки информационного поиска, машинного перевода и автореферирования, диалоговых и других систем.

Ключевые слова: агглютинативные языки, морфология, онтологии.

ONTOLOGICAL MODEL OF NOUNS SYSTEM FOR KAZAKH-CHINESE MACHINE TRANSLATION

Unzila Kamanur L.N. Gumilyov Eurasian National University, Astana, 010000, Kazakhstan Unzila.88@mail.ru

Altynbek Sharipbay L.N. Gumilyov Eurasian National University, Astana, 010000, Kazakhstan

Gulmira Bekmanova L.N. Gumilyov Eurasian National University, Astana, 010000, Kazakhstan

Lena Zhetkenbay L.N. Gumilyov Eurasian National University, Astana, 010000, Kazakhstan

Abstract. In this work, the machine translation for jüyecine Kazakh, Chinese nouns establishment of the ontological model. This model allows us to compare the similarities and differences between the two languages. Information retrieval, machine translation and avtoreferattaw, and other systems can be used to create a dialogue.

Keywords: agglutinative languages, morphology, ontology.

1. Введение

В настоящей работе строятся онтологические модели имен существительных казахского и китайского языков для создания системы машинного перевода между ними.

Казахский язык входит в кыпчакскую группу [1], а Китайский язык - ветвь сино-тибетской языковой семьи.

Китайский язык очень древний язык, который считается одним из самых древних языков. Кроме того, язык постоянно развивается и является одним из совершенствующих себя языков. Китайский язык входит в ряд хрупких языков. В качестве любых грамматических показателей используются символы иероглифы. В различных частях речи отдельные слова тоже пишутся иероглифами. Важным фактором непрерывного развития этого языка является использование иероглифов в качестве языкового инструмента. Китайский язык является одним из 6 рабочих и официальных языков Организации Объединенных наций. Если взглянуть с исторической стороны, то этот язык является языком национальности хань, который входит в состав национальностей КНР.

Из-за того что китайский язык является неприменимой формой, при формировании слов в структуре не происходит большие изменения. Казахский язык богат окончаниями и суффиксами, а в китайском языке более 70 % слов составляет двуслоговые слова. Также в китайском языке один иероглиф является одним слогом, вследствие этого большинство морфем будет одинарным. Например: 山, 海, 江, 我, 人. Вместе с этим они бывают двуслоговые и многослоговые, например: 蜘蛛, 蝴蝶, 仿佛, 巧克力, 阿司匹林. Каждый слог многословных слов не дает никакого значения, только каждый слог соединяясь с другим дает какой-нибудь смысл, при произношении не имеет значений. Например: 巧 иероглиф имеет единственное значение, а 巧克力 не имеет никакого смысла, и поэтому не морфема. Соединение три иероглифов дает одну морфему. Слова, состоящие из двух или нескольких морфем называются слитными словами. Например: 英雄, 劳动, 哈萨克, 文化局.

Слова на китайском языке во многих случаях по внешнему виду не отражает морфологических признаков. Только в составе предложения можно точно определить, какой части речи относятся слова. По типу связей двух слов делается морфологический анализ. Эта отличительная особенность предложений написанных этими иероглифами и показывают гибкость использования иероглифов. [5-8]

В настоящее время существуют различные виды систем машинного перевода. Применение тех или иных систем машинного перевода может быть продиктовано сложностью формализации того или иного естественного языка и наличием национальных языковых корпусов естественного языка.

Проблеме машинного перевода с китайского языка на другие тюркские языки посвящены [9-15]. Проблемы формализации грамматических правил казахского языка решены в работах. Эти результаты существенно облегчают создание системы казахско-китайского машинного перевода.

2 Онтологическая модель имени существительного для системы машинного перевода

В настоящее время онтология является мощным и распространенным инструментом моделирования отношений между объектами различных предметных областей. Принято классифицировать онтологии по степени зависимости от задач или прикладной области, по модели представления онтологических знаний и его выразительным возможностям и другим параметрам [Gruber,1993][Gruber1995].

Основным смыслом использования онтологии в области технической науки – показать концептуальным чертежом обобщенные и отдельные формализации охватывающие все множества данных по одной определенной образовательной сфере. В основе концептуального чертежа дается множества понятий и данные о понятиях (особенность,

связь, ограничения, аксиомы и закрепление понятий, вся эта информация необходима для описания процесса решения задачи по выбранной предметной области)

Почему возникает необходимость в создании онтологии? Обычно онтология используется в зависимости от следующих причин:

- Для общего понимания структуры информации людьми и программными агентами;
- Для возможности повторного использования знаний в предметной области;
- Для четких предположений в предметной области;
- Для разделения знаний предметной области от оперативных знаний;
- Для проведения анализа знаний в предметной области;

Использование общего понимания структуры информации людьми и программными агентами вместе являются одной из общих целей создания онтологии.

Для создания онтологии необходимо определить его предметную область и масштабы. Итак ответим на несколько основных вопросов:

1. Какую область будет охватывать онтология? Ответ: Имена существительные.
2. Для чего нам необходима онтология? Ответ: нужна для создания сравнительной онтологической модели по именам существительным казахского и китайского языков.
3. На какие виды вопросов должна отвечать информация в онтологии? Ответ: для определения спряжения и видов окончания имени существительного по содержанию и значению.
4. Кто будет использовать и поддерживать онтологию? Ответ: Лингвисты и программисты.

В соответствии с этими ответами сравнительная онтологическая модель имени существительного будет такой: $O(X, R, I)$, X – это наименования, входящие в структуру имени существительного, R – связи между наименованиями, а I – множества наименований структур и отношений.

Для создания онтологической модели используется языки описания онтологии и редакторы. Разные онтологические языки дает возможность использовать разные возможности. Сравнительная онтологическая модель имени существительного была создана одной из последних обработок стандартных языков описывающих онтологию - Protege (<http://protege.stanford.edu>). Protege OWL дает возможность описывать не только понятия, но и также конкретные объекты. Наименования и знаки использованные в сравнительной онтологической модели имени существительного казахского и китайского языков описаны в 1-таблице

Таблица 1 – имена и знаки использованные в сравнительной онтологической модели имен существительных из казахского и китайского языков

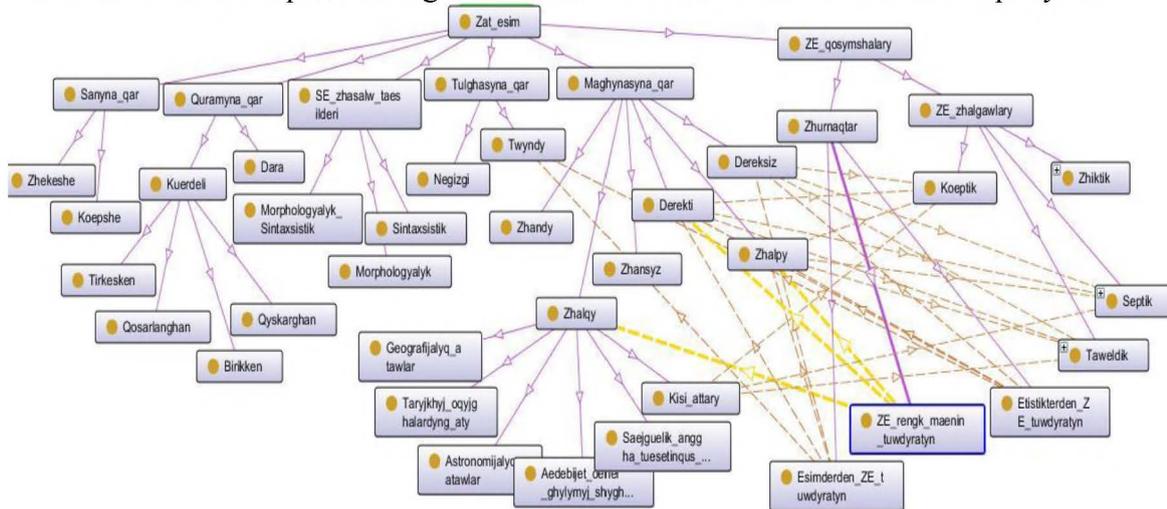
N	На казахском		На китайском	English	UNIFIED
	Существительные			Noun	
1.	По составу			Structure	
1.1		Единственное		Simple	
1.2		Сложное		Complex	
1.2.1		Объединенное		Compound words	
1.2.2		Парное слово		Reduplicates	
1.2.3		Словосочетание		Compound words	
1.2.4		Сокращенное слово		Abbreviations	
2.	По форме				
2.1		Основное		Derivations	DERIVNS
2.2		Производное		Derivatives	DERIV

2.3					
3.	По значению		Виды по значению 名词的种类	According to the meaning	
3.1		Одушевленное	表示人的 (Biǎoshì rénde) имена существительные означающие имена людей	Animate	ANIM
3.2		Неодушевленное	表示事物的 (Biǎoshì shìwù de): Имена существительные означающие предметы и явления	Inanimate	INANIM
3.3		Общее	表示时间: (Biǎoshì shíjiān) Имена существительные означающие период	Common	COM
3.4		Собственное	表示方位的: (Biǎoshì fāngwèi de) Имена существительные означающие место и расположение	Proper	PROP
4.			表示处所的 (Biǎoshì chùsuǒ de): Имена существительные означающие место		
		Конкретное		Concrete	CON
		Абстрактное		Abstract	ABS
5.	По количеству				
5.1		Единственное		Singular	SG
5.2		Множественное		Plural	PL
5.3				Collective	
6.	По окончанию				

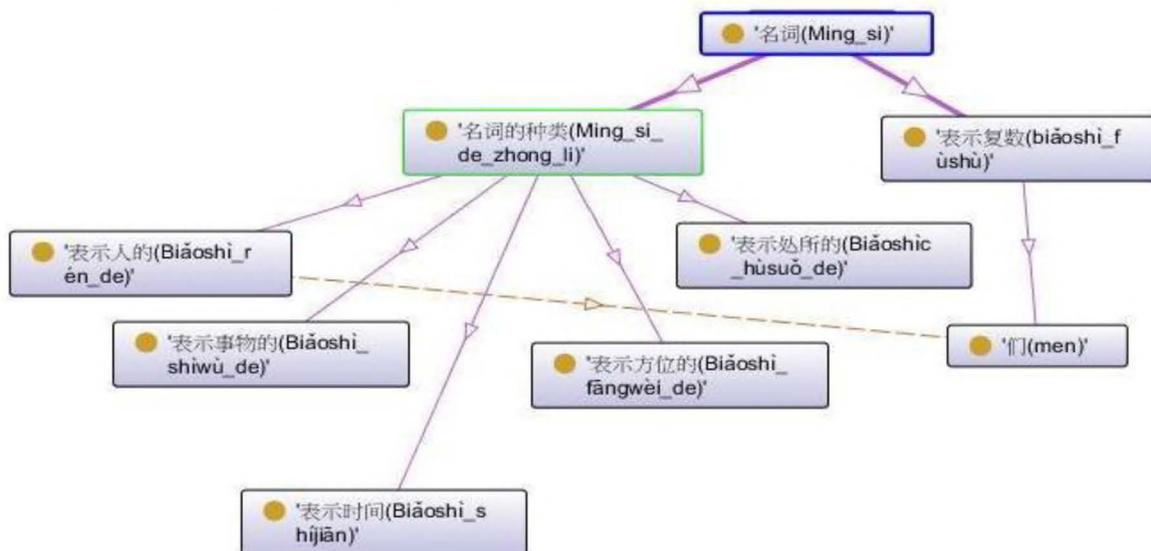
6.1	По падежному окончанию			Caseending	Cases
6.1.1		Именительное		Nominativecase	NOM
6.1.2		Родительный		Genitivecase	
6.1.3		Дательный		Direction-dativecase	DAT
6.1.4		Винительный		Accusativecase	ACC
6.1.5		Местный Жатыс		Locativecase	LOC
6.1.6		Исходный Шыгыс		Ablativecase	ABL
6.1.7		Творительный		Instrumentalcase	
6.2	По множественному числу		表示 复数biǎoshì fùshù 表示人的: именам существительным означающим имена людей присоединя окончание men получаем множественное значение		Plural
6.2.1				Singular	SG
6.2.2				Plural	PL
6.3	По личному окончанию			Personalending	Pers_end
		1 лицо единственное		1 personalsingular	PERS.1SG
		2 лицо единственное		2 personalsingular	PERS.2SG
		3 лицо единственное		3 personalsingular	PERS.3SG
		2 лицо единственное вежливое		2 personalsingularformal	
		1 лицо множественное		1 personalplural	PERS.1PL
		2 лицо множественное		2 personalplural	PERS.2PL
		3 лицо множественное		3 personalplural	PERS.3PL
		2 лицо множественное вежливое		2 personalpluralformal	
6.4	По притяжательной форме			Possesiveending	Poss_end
		1 лицо		1 Possesivesingular	POSS.1SG

		ЕДИНСТВЕННОЕ			
		2 лицо ЕДИНСТВЕННОЕ		2 Possivesingular	POSS.2SG
		3 лицо ЕДИНСТВЕННОЕ		3 Possivesingular	POSS.3SG
		2 лицо ЕДИНСТВЕННОЕ ВЕЖЛИВОЕ		2 Possivesingularfor mal	
		1 лицо МНОЖЕСТВЕННОЕ		1 Possiveplural	POSS.1PL
		2 лицо МНОЖЕСТВЕННОЕ		2 Possiveplural	POSS.2PL
		3 лицо МНОЖЕСТВЕННОЕ		3 Possiveplural	POSS.3PL
		2 лицо МНОЖЕСТВЕННОЕ ВЕЖЛИВОЕ		2 Possivepluralfor mal	

Созданная сравнительная онтологическая модель имени существительного казахского и китайского языков в среде Protege с помощью этих значений показан в 1-рисунке



1-рисунок. Онтологическая модель имени существительного казахского языка



2-рисунок. Онтологическая модель имени существительного китайского языка

Таким образом, сравнительная онтологическая модель имени существительного для системы машинного перевода охватывает все компоненты и множества морфологического анализа, а именно в казахском языке по составу имени существительного делятся на единственное и множественное, по значению – общее, собственное, конкретное, абстрактное, одушевленное и неодушевленное, а в китайском – не предусмотрено такое разделение, по значению рассмотрены 表示人的: имена существительные означающие имена людей, 表示事物的: имена существительные означающие предметы и явления, 表示方位的: имя существительное означающее место и направление, 表示时间: имена существительные означающие период, 表示处所的: имена существительные означающие место. В казахском языке слова означающие период и расположение рассматривается как наречие, а в китайском языке как имя существительное. В китайском языке имена существительные не спрягаются, только одно множественное окончание «men», а также его круг пользования узок. Его специфичность в том, что этот предмет делится на группы по виду, форме, внешнему виду, качественному свойству. Например: предмет широкий, длинный, в комочек и т.д. Вот такие имена предметов разделенные на группы обязательно используется с именем числительным сочетающимся с местоимением, а между ними ставятся слова означающие величину.

Предметы и явления окружающей среды воспринимаются и узнаются по разному. В китайском языке задается вопрос кто? именам существительным человеку и его деятельности, профессий и работе. В этом языке очень много слов определяющее наименование предметов, отвечающее на вопросы кто? что? В общем эти все являются наименованиями предметов, явлений, профессий, и каждый предмет имеет смысловой, объяснительный, понятийный и другие свойства. Обычно “不, 很” имена существительные не согласуются напрямую с наречиями, а также с именем числительным, между ними без слов понадобится слово дающее значение величины. Имена существительные в предложении часто бывают подлежащим и дополнением. [4]

3 Выводы: Созданная онтологическая модель для компьютерной обработки казахского и русского языков является важнейшим шагом при сравнительном исследовании двух языков. И поэтому исследование структуры и значений имен существительных казахского и китайского языков и их результаты сравнений безусловно влияет на систему машинного перевода и создание системы обработки естественного языка.

Список литературы

1. Казахская грамматика. Фонетика, словообразование, морфология, синтаксис. – Астана, 2002.
2. Нұрсайдов М.А. «Қытай тілінің өзіндік ерекшеліктеріне қысқаша сипаттама» Алматы, 2002 ж.
3. 卢福波《对外汉语教学实用语法》北京语言文化大学出版社2002年
4. Дүкен Мәсімханұлы “Қытай филологиясына кіріспе” (оқулық, 2004 ж., Астана).
5. Бекманова Г.Т., Махимов А.К. Графематический анализ казахского неструктурированного текста // Информатизация общества: труды 3-ой международной научно-практической конференции.– Астана, 2012.– С. 509-511
6. Бекманова Г.Т., Шарипбаев А. А. Формализация морфологических правил казахского языка с помощью семантической нейронной сети // Доклады Национальной академии наук Республики Казахстан. – 2009.– №4. – С. 11-16
7. Bekmanova G., Sharipbaev A. The synthesis of word forms of Turkic language using semantic neural networks // Modern problems of applied mathematics and information technologies: abstracts – Al Khorezmy, 2009.–P.145.
8. G.T.Bekmanova, A.A.Sharipbaevy, S.K.Buribayeva Formalization of morphological rules of inflection in the Kazakh language // Вестник. Астана: Евразийский национальный университет им. Л.Н.Гумилева, 2012. – Специальный выпуск.– С.18-26

9. Methods and models of Методы и алгоритмы распознавания слов казахского языка: монография // Астана: ЕНУ им. Л.Н.Гумилева, 2010. 132с.

10. A.Sharipbayev, G.Bekmanova, B.Yergesh, A.Buribayeva, and M.K.Karabalayeva. Intellectual morphological analyzer based on semantic networks. Processings of the OSTIS-2012, 2012, pp.397-400.

УДК 004.5:811.512.145

РЕЧЕВОЙ ЧЕЛОВЕКО-МАШИННЫЙ ИНТЕРФЕЙС НА ТАТАРСКОМ ЯЗЫКЕ

*А.Ф. Хусаинов, Институт прикладной семиотики Академии наук Республики Татарстан Казанский (Приволжский) федеральный университет, Казань, Россия
khusainov.aidar@gmail.com*

Аннотация

В работе приводятся результаты исследований по созданию речевого человеко-машинного интерфейса на татарском языке. Он включает в себя два основных элемента: систему автоматического распознавания речи и систему синтеза речи. Планируется использовать возможности речевого ввода и вывода информации на татарском языке при создании множества диалоговых систем: машинного переводчика, интеллектуального помощника и т.д.

Ключевые слова: распознавание речи, синтез речи, речевой интерфейс, татарский язык

SPEECH HUMAN-MACHINE INTERFACE FOR THE TATAR LANGUAGE

A.F. Khusainov, Institute of Applied Semiotics of the Tatarstan Academy of Sciences Kazan (Volga region) federal university, Kazan, Russia, khusainov.aidar@gmail.com

In this paper, we describe our recent work of creation speech human-machine interface for the Tatar language. Our work consists of two main elements: speech recognition system and speech synthesizer. These systems will be used in mobile and desktop applications, for instance, machine translation system, smart assistant.

Key words: speech recognition, speech synthesis, speech interface, the Tatar language

1 Введение

Увеличение числа информационных систем и условий их работы диктует развитие новых средств взаимодействия с ними. Одним из актуальных способов взаимодействия является использование произнесённых команд и сообщений на естественном языке. Речевой интерфейс получает всё большее распространение благодаря нескольким факторам: естественности использования речи для человека, удобству использования в определенных условиях, а также развитию технологий анализа, синтеза речи и понимания текста.

Данная работа описывает последние результаты, полученные при создании систем автоматического распознавания и синтеза речи в контексте использования татарского языка. Для решения обеих задач используется корпусный подход: система распознавания обучается на базе создаваемого многодикторного корпуса слитной речи, синтезатор речи – на основе студийных записей профессиональных дикторов.

2 Система распознавания слитной татарской речи

Система распознавания речи состоит из 4 основных элементов:

Акустические модели: модели произношения акустических единиц языка (фонем, дифононов и т.д.);