

9. Espla-Gomis, Miquel, and Mikel Forcada. "Combining content-based and url-based heuristics to harvest aligned bitexts from multilingual sites with bitextor". The Prague Bulletin of Mathematical Linguistics 93 (2010): 77-86.

10. Washington, Jonathan, Ilmar Salimzyanov, and Francis M Tyers. "Finite-state morphological transducers for three Kyrgyz languages". LREC 2014: 3378-3385.

УДК 681.3

АЛГОРИТМ ОБРАЗОВАНИЯ СЛОВОФОРМ ДЛЯ АВТОМАТИЗАЦИИ ПРОЦЕДУРЫ ПОПОЛНЕНИЯ БАЗЫ ДАННЫХ СЛОВАРЯ

Бакасова Пери Султановна, магистрант КГТУ им. И. Раззакова, Кыргызстан, 720044, г.Бишкек, пр. Айтматова Ч., 66, e-mail: bakasovap@mail.ru

Исраилова Нелла Амантаевна, к.т.н., зав. кафедрой ИВТ, КГТУ им. И.Раззакова, Кыргызстан, 720044, г.Бишкек, пр. Айтматова Ч., 66, e-mail: inela@mail.ru

Цель статьи – исследовать алгоритм формирования словоформ по морфологическим признакам части речи. В данной статье рассматриваются принципы формирования форм глаголов исходя из следующих морфологических признаков: спряжение (первый, второй), лицо (первое, второе, третье), число (единственное, множественное число), время (прошлое, настоящее, будущее).

Ключевые слова: алгоритм, машинный перевод, морфологические признаки, часть речи, словоформа, морфологический анализ

ALGORITHM OF FORMATION WORD FORMS FOR AUTOMATION REPLENISHMENT OF DICTIONARY DATABASES.

Bakasova Peri Sultanovna, undergraduate KSTU. I. Razzakova, Kyrgyzstan, 720044, Bishkek, av. Aitmatov Ch, 66, e-mail: bakasovap@mail.ru

Israilov Nella Amantaevna, PhD, Head. the Department of ICT, KSTU. I.Razzakova, Kyrgyzstan, 720044, Bishkek, av. Aitmatov Ch, 66, e-mail: inela@mail.ru

The purpose of the article - the investigate of algorithm to the formation of word forms on morphological characteristics of the speech. This article discusses the principles of the verb forms the basis of the following morphological features: conjugation (first, second), a person (first, second, third), number (singular, plural), time (past, present, future).

Keywords: algorithm, machine translation, morphological features, part of speech, word forms, morphological analysis

Для обеспечения гибкости при переводе текстов (предложений, слов) необходим словарь, содержащий различные формы слова и их переводы. Для пополнения данного словаря используем алгоритм формирования слов по морфологическим признакам части речи, к которой относится исходное слово на русском языке(RU) и его перевод на кыргызском языке(KG). Суть алгоритма заключается в формировании разных форм исходного слова и его перевода по морфологическим признакам части речи, к которой относится исходное слово и слово- перевод.



Рис. 1 Схема формирования словоформ по морфологическим признакам части речи в русском языке

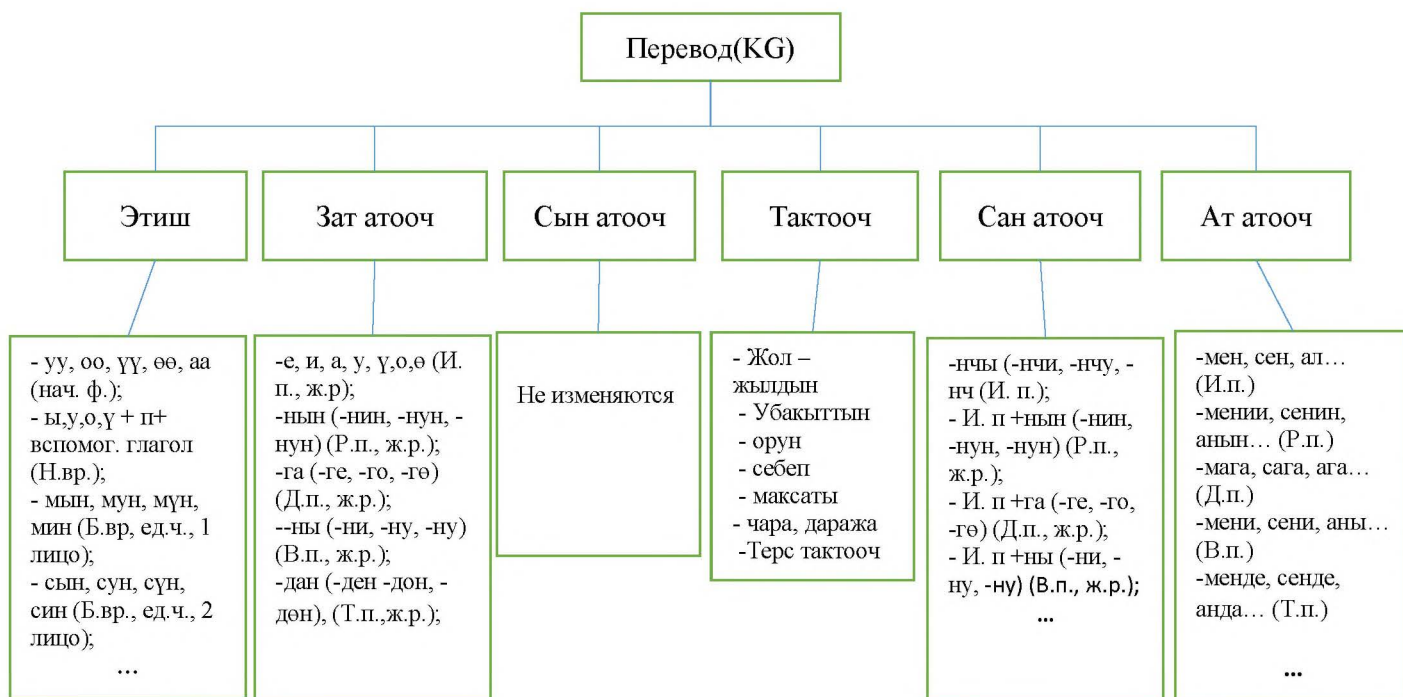


Рис. 2 Схема формирования словоформ по морфологическим признакам части речи в кыргызском языке

Формирование форм глаголов в русском языке

Алгоритма формирования слов по морфологическим признакам глаголов позволяет распознать к какому спряжению относится искомый глагол и образует формы слова следующим образом.

Формы глаголов в русском языке образуются спряжением глаголов по числам и лицам.

Формы глагола первого спряжения настоящего времени образуются спряжением глагола по лицам и числам добавляя личные окончания глаголов, показанные в таблице 1:

Таблица 1. – Личные окончания глаголов первого спряжения в настоящем времени

	Ед. ч.	Мн. ч.
1 лицо	ю	ем
2 лицо	ешь	ете
3 лицо	ет	ют

Формы глагола второго спряжения настоящего времени образуются спряжением глагола по лицам и числам добавляя личные окончания глаголов, показанные в таблице 2:

Таблица 2. - Личные окончания глаголов второго спряжения в настоящем времени

	Ед. ч.	Мн. ч.
1 лицо	ю	им
2 лицо	ишь	ите
3 лицо	ит	ят

Глаголы в прошедшем времени спрягаются по родам и числам с помощью окончаний показанных в таблице 3.

Таблица 3. – Окончания глаголов в прошедшем времени

	Ед. ч.	Мн. ч.
м.р.	л	ли
ж.р.	ла	ли
ср.р.	ло	ли

Формы глаголов прошедшего совершенного вида образуются добавлением приставки “с” и личных окончаний глаголов, показанных в таблице 3.

Глаголы в будущем времени имеют те же окончания что и глаголы в настоящем времени, но к ним еще и добавляются такие приставки как – *с, по, раз, у* и т.д.

§ Формирование форм глаголов в кыргызском языке

Формы глаголов в кыргызском языке, как и в русском образуются спряжением глаголов по числам и лицам.

Формы глагола настоящего времени образуются с помощью двух глаголов — основного и вспомогательного: к основному глаголу в исходной форме — в повелительном наклонении — прибавляется окончание -п, если глагол оканчивается на гласную, или -ып (-ип, -уп, -үп), если глагол оканчивается на согласную.

В качестве вспомогательных чаще всего берутся глаголы жат - лежи и жүр - иди; реже глаголы тур - встань и отур - садись в форме настоящего времени, которые теряют свое значение и на русский язык не переводятся.

При спряжении глаголов настоящего времени основной глагол не изменяется, а вспомогательный принимает личные окончания будущего времени,

Формы глаголов будущего времени образуются путем прибавления к основе глагола (повелительному наклонению) личных окончаний

Таблица 4. – Личные окончания глаголов в настоящем времени

	Единственное число	Множественное число
1 лицо	-мың (-миң, -муң, -мүң)	-быз (-биз, -буз, -бүз)
2 лицо	-сың (-сиң, -суң, -сүң)	-сыңар (-сиңер, -суңар, -сүңөр)
2 лицо (в.ф.)	-сыз (-сиз, -суз, -сүз)	-сыздар (-сиздер, -суздар, -сүздөр)
3 лицо	-т	-шат (-шет, -шот, -шөт)

При спряжении глагола в будущем времени между основой глагола (повелительным наклонением) и личным окончанием появляется звук -Й-, если глагол оканчивается на гласную, или звук -А- (-Е-, -О-, -Ө-), если глагол оканчивается на согласную. Если исходная форма глагола заканчивается на Й, то при сочетании этих глаголов с личными окончаниями будущего времени во всех лицах вместо ЙО пишется Ё, вместо ЙА — Я.

В кыргызском языке глаголы прошедшего определенного времени образуются путем прибавления к основе глагола суффиксов -ды (после гласных и звонких согласных) или -ты (после глухих, согласных) плюс личные окончания глагола.

Таблица 5. – Окончания глаголов в прошедшем совершенном времени

	Единственное число	Множественное число
1 лицо	-дым (-дим, -дум, -дүм)	-дык (-дик, -дук, -дүк)
2 лицо	-дың (-диң, -дуң, -дүң)	-дыңар (-диңер, -дуңар, -дүңөр)
2 лицо (в.ф.)	-дыңыз (-диңиз, -дүңүз, -дүңүз)	-дыңыздар (-диңиздер, -дүңүздөр, -дүңүздөр)
3 лицо	-ды (-ди, -ду, -дү)	-шты (-шти, -шту, -штү)

Прошедшее обычное время образуется при помощи суффикса -ган (-ген? -гонг -ген), если основа глагола оканчивается на гласный и звонкий согласный, или -кан (-кен, кон, -кон), если основа глагола оканчивается на глухой согласной.

Таблица 6. – Окончания глаголов в прошедшем обычном времени

	Единственное число	Множественное число
1 лицо	-ганмың (-генмиң, -гонмуң, -гөнмүң)	-ганбыз (-генбиз, -гонбуз, -гөнбүз)
2 лицо	-гансың (-генсиң, -гонсуң, -гөнсүң)	-гансыңар (-генсиңер, -гонсуңар, -гөнсүңөр)
2 лицо (в.ф.)	-гансыз (-генсиз, -гонсуз, -гөнсүз)	-гансыздар (-генсиздер, -гонсуздар, -гөнсүздөр)
3 лицо	-ган (-ген, -гон, -гөн)	-ышкан (-ишкен, -ошкон, -үшкөн)

Если глагол оканчивается на согласную, то в 3 л. мн, числа между основой и личным окончанием вставляется показатель совместимости -ыш (-иш, -уш, -уш).

§ Запись в словарь

Слова записываются в словарь в виде:

Слово1(RU) | Перевод1(KG) /

Слово2(RU) | Перевод2(KG) /

Также образованные формы слова сохраняются в базе по частям речи.

Например, глаголы на русском языке хранятся в базе глаголов, а их переводы на кыргызском в базе глаголов кыргызского языка в следующем формате:

Глагол(RU) / Форма1 / Форма2 / Форма3 / ...

Этиш(KG) / Форма1 / Форма2 / Форма3 / ...

Вывод:

Алгоритм формирования слов по морфологическим признакам части речи обеспечивает быстрое пополнение базы словаря за счет автоматического словообразования на основе искомого слова и его перевода и обеспечивает гибкость при переводе.

Недостатком является возможность возникновения ошибки при формировании словоформ. Поэтому дальнейшие оптимизации и модификации данного алгоритма являются актуальными.

Список литературы

1. Грамматика кыргызского языка: краткий справочник для студентов, Бишкек 2002
2. Глазунова О. Грамматика русского языка в упражнениях и комментариях (Морфология + Синтаксис).
3. Ө. Калыева Разговорник русско-кыргызский - Орусчакыргызча сүйлөшмө
4. Исраилова Н.А. Принципы организации морфологического анализатора в трансляторе. /Известия КГТУ им. И. Раззакова-2010, №20, С233-236
5. Исраилова Н.А. Организация морфологического анализа в трансляторах. /Вестник Восточно-Казахстанского государственного технического университета им. Д. Серикбаев, Усть-Каменогорск, №1, март, 2012 г.-С 97-101
6. Исраилова Н.А. Алгоритмы отладки процесса трансляции. /Известия КГТУ им. И. Раззакова-2011, №22, С278-279

УДК 004.432.4

AUTOMATING OF THE TEXT GENERATION WITH A GIVEN REPRESENTATIVENESS PHONETIC UNITS BY A FORMALIZATION OF PHONOLOGICAL RULES

Aigerim Buribayeva, PhD, L.N.Gumilyov Eurasian National University, 010008, Astana, Pushkin str. 2, e-mail: buribayeva@mail.ru

Arman Kaliyev, L.N.Gumilyov Eurasian National University, 010008, Astana, Pushkin str. 2, e-mail: kaliyev.arman@yandex.kz

Banu Yergesh, L.N.Gumilyov Eurasian National University, 010008, Astana, Pushkin str. 2, e-mail: saturn_banu@mail.ru

The article describes formalization of phonological rules of Kazakh language and use it to automate the process of formation carried the text of the material with a given phonetic units of representativeness (in particular diphones). This is necessary, particularly in the development of automatic speech synthesis. Using this versatile program, it will be able to get the text material with full coverage of all possible diphones for all Turkic language.

Keywords: Speech synthesis, text analyzer, sound units, diphones, statistics, acoustic database, text body.

АВТОМАТИЗАЦИЯ ФОРМИРОВАНИЯ ТЕКСТОВОГО МАТЕРИАЛА С ЗАДАННОЙ ПРЕДСТАВИТЕЛЬНОСТЬЮ ФОНЕТИЧЕСКИХ ЕДИНИЦ С ПОМОЩЬЮ ФОРМАЛИЗАЦИИ ФОНОЛОГИЧЕСКИХ ПРАВИЛ

Бурибаева Айгерим Кеулимжаевна, PhD, ЕНУ им. Л.Н. Гумилева, 010008, Астана, Пушкина 2, e-mail: buribayeva@mail.ru

Калиев Арман Куанышевич, ЕНУ им. Л.Н. Гумилева, 010008, Астана, Пушкина 2, e-mail: kaliyev.arman@yandex.kz

Ергеш Бану Жантуганкызы, ЕНУ им. Л.Н. Гумилева, 010008, Астана, Пушкина 2, e-mail: saturn_banu@mail.ru

В статье описана формализация фонологических правил казахского языка и на ее основе осуществлена автоматизация процесса формирования текстового материала с заданной представительностью фонетических единиц (в частности дифонов). Это