

УДК 811

МНОГОФУНКЦИОНАЛЬНАЯ МОДЕЛЬ ТЮРКСКОЙ МОРФЕМЫ КАК БАЗА ДАННЫХ ДЛЯ ЛИНГВОПРОЦЕССОРОВ

© Джавдет Сулейманов, Айрат Гатиатуллин, Акмарал Альменова, Артур Баширов

A MULTIFUNCTIONAL COMPUTER BASED LINGUISTIC MODEL AS A DATABASE FOR LINGUISTIC PROCESSORS

Dzhavdet Suleymanov, Ayrat Gatiatullin, Akmaral Almenova, Artur Bashirov

The methods for the organization, formalization and structuring of data as elements of knowledge base are important to the linguistic resources used to create computer language processors. Currently, it is especially important while creating databases and knowledge bases for the computer processing of Turkic languages. This article proposes a conceptual and structural description of a multifunctional linguistic model of the Turkic morpheme, built on the basis of the pragmatic-oriented approach for constructing linguistic models.

The multifunctional model of the Turkic morpheme is information software, including technology tools for filling the database, used as a resource base for a number of practical applications.

Filled with the relevant content, the model may be used in various practical ways, primarily as a resource base for the computer processing software of Turkic languages, as well as an information and reference system of Turkic morphemes, and as a tool for Turkology scientists, researching morphology, in particular for a comparative analysis of Turkic language units.

The use of the morpheme model as a resource base of this kind is primarily caused by the significance of the morphological language level while processing natural language texts. This is especially relevant for the agglutinative languages with a rich morphology, which include all the languages of the Turkic family.

Keywords: Turkic morpheme, multifunctional model of Turkic morphemes, machine translation, computer-based processing of Turkic languages.

В лингвистических ресурсах, используемых для построения лингвопроцессоров, важное значение имеют способы организации, формализации и структуризации данных как элементов базы знаний. В настоящее время это особенно актуально при создании баз данных и знаний для компьютерной обработки тюркских языков. В данной статье предлагается концептуальное и структурное описание многофункциональной лингвистической модели тюркской морфемы, построенной на основе прагматически ориентированного подхода для построения лингвистических моделей.

Многофункциональная модель тюркской морфемы представляет собой информационно-программную оболочку, технологический инструментарий для заполнения базы данных, а также использования ее в качестве ресурсной базы для ряда практических приложений.

Данная модель, заполненная соответствующим контентом, может иметь различное практическое применение, прежде всего, как ресурсная база для программных продуктов, осуществляющих компьютерную обработку тюркских языков, а также как информационно-справочная система о морфемах тюркских языков и как инструментарий для исследований ученых-тюркологов в области морфологии, в частности, для сравнительного анализа тюркских языковых единиц.

Использование в качестве подобной ресурсной базы именно модели морфем обусловлено, в первую очередь, значимостью морфологического языкового уровня при обработке естественно-языковых текстов. Это особенно актуально для языков агглютинативного типа с богатой морфологией, к которым относятся все языки тюркского семейства.

Ключевые слова: тюркская морфема, многофункциональная модель тюркской морфемы, машинный перевод, тюркские лингвопроцессоры.

Несмотря на возросшую активность тюркских языков в киберпространстве как языков накопления и передачи информации, наличие электронных корпусов языков, тематических, много-

язычных и терминологических словарей, практически все тюркские языки, кроме турецкого, относятся к малоресурсным языкам. Имеющиеся электронные ресурсы не являются достаточными

для создания эффективных программ машинного перевода, машин поиска и систем понимания естественного языка, являющихся не только актуальными, но крайне необходимыми как для обработки все возрастающей информации в Интернете на тюркских языках, так и для развития самих языков.

Очевидно, процессы накопления лингвистических ресурсов и создания систем обработки информации на тюркских языках находятся в отношении прямой зависимости. Чем больше качественных лингвистических ресурсов, тем эффективнее, точнее будут работать лингвопроцессы, и наоборот, чем совершеннее программы морфологического, синтаксического, семантического анализа, извлечения информации (data mining, knowledge acquisition), тем быстрее будут накапливаться лингворесурсы.

Сегодня де-факто определились основные точки теоретической и практической активности в пространстве тюркских языков (см.: [Дыбо, Шеймович], [Желтов], [Сиразитдинов], [Сулейманов], [Труды Первой международной конференции «Компьютерная обработка тюркских языков»], [Разработка Англо-Казахского Переводчика На Базе Open-Source Платформы Apertium], [Aisha, Sun], [Altenbek], [Oflazer], [Orhun, Tantug, Adali], [Proceedings of the International Conference on Turkic Language Processing], [Proceedings of the International Conference “Turkic Languages Processing: TurkLang-2015”], [Qetin, Amasyali], [Synchronized liner tree for morphological analysis and generation of the Kazakh language], [Yildirim, Tantug]) и настало время для объединения усилий специалистов, а также формирования общих принципов и подходов в области компьютерной обработки тюркских языков.

От интеграции усилий мы вправе ожидать хороших результатов, прежде всего благодаря тому, что тюркские языки имеют практически идентичную лексико-грамматическую структуру, соответственно, идентичные лингвистические модели и программные модули для их обработки.

Одной из перспективных задач в настоящее время видится создание единой технологической базы по компьютерной обработке тюркских языков, которая должна включать единую многоязычную систему машинного перевода для тюркских языков, а также единую систему многоязычного информационного поиска в сети Интернет на тюркских языках. Для разработки этих программных продуктов необходим целый ряд лингвистических ресурсов, моделей и словарей. В данной статье предлагается один из таких ре-

сурсов – это многофункциональная модель тюркской морфемы.

1. Описание модели

Многофункциональная модель тюркской морфемы представляет собой информационно-программную оболочку, технологический инструментарий для заполнения базы данных, а также использования ее в качестве ресурсной базы для ряда практических приложений.

Данная модель, заполненная соответствующим контентом, может иметь различное практическое применение, прежде всего, как ресурсная база для программных продуктов, осуществляющих компьютерную обработку тюркских языков, а также в качестве информационно-справочной системы о морфемах тюркских языков и инструментария для исследований ученых-тюркологов в области морфологии, в частности, для сравнительного анализа тюркских языковых единиц.

Использование в качестве подобной ресурсной базы именно модели морфем обусловлено, в первую очередь, значимостью морфологического языкового уровня при обработке естественно-языковых текстов. Это особенно актуально для языков агглютинативного типа с богатой морфологией, к которым относятся все языки тюркской семьи.

Авторами данной статьи разработана компьютерная структурно-функциональная модель татарских аффиксальных морфем, описанная в работе [Сулейманов, Гатиатуллин]. Опыт создания и использования структурно-функциональной модели для татарских аффиксальных морфем показал, что ценность и эффективность использования модели существенно возрастет при ее многоязычном использовании для нескольких тюркских языков, а также расширении модели для описания корневых морфем.

Многофункциональная модель тюркских морфем представляет собой прагматически-ориентированное структурно-функциональное описание элементов морфологии [Сулейманов, Гатиатуллин], [Сулейманов, Гатиатуллин, Карабаева] и позволяет осуществить полную «инвентаризацию» тюркских морфем с описанием характеристик и ситуаций их проявления на всех языковых уровнях (фонологическом, морфологическом, синтаксическом, семантическом).

Многофункциональная модель тюркской морфемы имеет иерархическую структуру, состоящую из множества подмоделей. Модель морфемы каждого из тюркских языков является составной частью общей модели. Общая схема модели представлена на рис. 1.

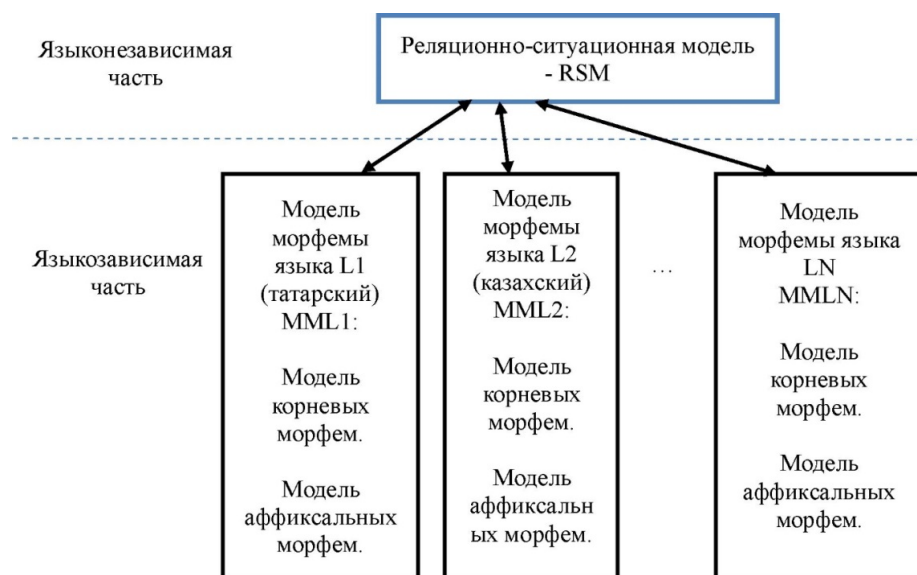


Рис. 1 Общая схема модели

Концептуальным ядром многофункциональной прагматически ориентированной лингвистической модели, служащим для сравнительного анализа, является реляционно-ситуационная модель, которая предназначена для описания контекстно-семантических характеристик морфем.

Модель морфемы каждого из тюркских языков включает модели корневых морфем и аффиксальных морфем. Связь между моделями морфем для отдельных языков осуществляется с помощью реляционно-ситуационной модели [Сулейманов, Гатиатуллин, Вагапов], которая используется для описания семантического аспекта моделей отдельных тюркских языков, включенных в единую модель.

Каждая из подмоделей, представленных на рис. 1 состоит из следующих аспектов:

1. идентификационный аспект;
2. морфонологический аспект;
3. морфологический аспект;
4. синтаксический аспект;
5. семантический аспект.

Рассмотрим содержание идентификационного, морфологического и морфонологического аспектов модели.

Идентификационный аспект

Морфема является не знаком, реально используемым в текстах, а абстрактной языковой единицей, классом знаков, который объединяет набор алломорфов, обладающих рядом одинаковых параметров. Этому абстрактному классу даются разные названия исходя из его состава или функций. Один и тот же класс разными специалистами называется по-разному, и для его обозначения могут быть использованы разные символы. В нашей модели Идентификационный аспект

объединяет эти разные типы обозначений и имеет следующую структуру:

- а. Обозначение морфемы
- б. Цифровой идентификатор
- с. Идентификатор морфологической категории для разметки корпуса (тэг)
- д. Название морфологической категории
4. Типологическое название
5. Название на русском языке
6. Название на национальном языке

а) Обозначение морфемы

Одно из этих названий морфемы выводится из названий, входящих в него алломорфов, по определенным правилам. Так, класс, объединяющий татарские алломорфы *-a, -ə, -ga, -gə, -ka, -kə*, обозначается, как *-[Г]А*. Это обозначение класса и задается в параметре «Обозначение морфемы». В казахском языке класс с аналогичным набором свойств имеет следующий набор алломорфов *-ga, -ge, -qa, -ke, -a, -e*. Этому классу в нашей модели присвоили обозначение *-[F]А*. Здесь и далее при обозначении морфемы заглавные буквы обозначают вариативность. Дефис '-' показывает связность данной морфемы слева, то есть в словоформе с данной аффиксальной морфемой слева от нее обязательно должна находиться другая морфема. А квадратные скобки [Г] показывают, что в этот класс входят алломорфы без данного символа. В приведенном примере у морфемы *-[Г]А* имеются алломорфы *-a* и *-ə*, состоящие всего из одного символа.

Для решения различных задач специалистами используются разные морфологические обозначения. Это может быть обозначение морфологической категории, для выражения которой используется данная морфема, или обозначение морфологического тэга при разметке электронно-

го корпуса. По разным языковым традициям одна и та же грамматическая категория может называться по-разному. Примером этому служит информация, приведенная в таблице 1. Это различные названия направительного падежа в грамматиках разных тюркских языков.

Таблица 1

Название	Язык
Юнәләш килеше	Татарский
Yönelme hâli	Турецкий
Барыс септік	Казахский
Барыш жөндөмө	Киргизский
Жўналиш келишиги	Узбекский
Догърултув келиши	Крымскотатарский

Таким образом, для заполнения идентификационного аспекта модели необходимо собрать и сопоставить идентифицирующую информацию о морфемах и морфологических категориях в тюркских языках.

б) Цифровой идентификатор

Цифровой идентификатор – это параметр, который автоматически генерируется системой при добавлении в базу данных описания новой морфемы и не может редактироваться пользователем. Этот параметр необходим для однозначной идентификации морфемы программными средствами и имеет следующий вид: 01.1.01.

Здесь первые две цифры обозначают номер языка, третья цифра (между точками) показывает тип морфемы (корневая (1) или аффиксальная (2)), а последующие цифры обозначают порядковый номер самой морфемы.

в) Идентификатор морфологической категории для разметки корпуса

Идентификатор морфологической категории для разметки корпуса представляет собой сокращенное обозначение морфологической категории, с помощью которого морфема представляется в электронном корпусе, то есть это общепринятое обозначение (tag), используемое при аннотировании текстов в электронном корпусе языка.

Например: DIR, ABL.

Если морфема используется для выражения нескольких морфологических категорий, то выбирается одна из этих категорий, наиболее частотная или более важная, по усмотрению эксперта. Однако здесь возникает другой вопрос – какие классы, состоящие из одних и тех же алломорфов, считать одной морфемой, а какие разными?

1.1. Название морфологической категории – содержит названия морфологической категории, употребляемые на разных языках или в разных

коллективах специалистов (позиция установления соответствия между категориями)

Например:

Типологическое: Directive
Русское: Направительный падеж

Национальное (татарское): Юнәләш килеш

1. Морфологический аспект. Порядок следования

Морфологический аспект содержит информацию о морфологических свойствах морфемы, где одним из основных свойств является параметр – Порядок следования.

Как было сказано выше, морфема является не одним знаком, а классом, объединяющим несколько алломорфов. Следовательно, у разных алломорфов одной морфемы могут быть разные правила следования. Например, у той же морфемы $-[Г]А$ алломорфы $-a$ и $-ә$ в словоформе следуют только после алломорфов из классов притяжательности $-[Ы]м$, $-[Ы]ң$ и $-[с]Ы[н]$, а алломорфы $-ка$ и $-кә$, наоборот, никогда не используются после алломорфов притяжательности. Алломорфы $-га$ и $-гә$ могут следовать после алломорфов классов притяжательности $-[Ы]бЫз$, $-[Ы]гыз$. Таким образом, правила следования морфем также являются неким абстрактным правилом, которые показывают не строгие правила следования, а потенциальную возможность того, что некоторые алломорфы этой морфемы могут следовать после алломорфов другой морфемы.

Этот параметр уже присутствовал в структурно-функциональной модели татарской аффиксальной морфемы и содержал правила следования для аффиксальных морфем. В новой модели этот параметр содержит еще и правила сочетания корневых морфем с аффиксами. Форма представления в модели этих правил несколько отличается от представления правил следования аффиксальных морфем.

Корневых морфем в базе данных намного больше, чем аффиксальных, и они обладают набором совпадающих свойств по порядку следования в словоформе. Поэтому нет смысла описывать весь набор аффиксальных морфем, которые могут встретиться в словоформе, для каждой корневой морфемы. На основании этого в модели все корневые морфемы по типу сочетаемости с аффиксальными морфемами мы поделили на 4 морфотактических класса: N, V, D, S. Каждой корневой морфеме в модели присваивается его регулярный морфотактический класс. Рассмотрим содержание этих морфотактических классов.

1-й класс – N. Включает в себя корневые морфемы, справа от которых могут следовать все аффиксы именных категорий.

Например:

N = {-лАр, -[Ы]м, -[Ы]н, -[с]Ы[н], -[Ы]бЫз, -[Ы]гЫз, -[Г]А, -[н]Дан, -[н]ДА, -нЫ, -нЫн, ...}.

2-й класс –V. Включает в себя корневые морфемы, справа от которых могут следовать все аффиксы, присоединяемые к глагольным основам.

Например:

V = {-мА, -ЫЙ, -ДЫ, -ГАН, -АчАк . . .}

3-й класс – D. Корневые морфемы, справа от которых могут следовать аффиксы из категории N и плюс аффиксы, образующие формы числительного.

Например:

D = {-Ынчы, -[ш]Ар, -ГА, -[н]Дан, -[н]ДА, -нЫ . . .}

4-й класс – S. Корневые морфемы, справа от которых могут следовать только модальные аффиксы.

Например:

S = {-мЫ, -мЫни, -ДЫр}.

Представленные 4 класса используются для описания регулярных правил следования, которые присущи всем корневым морфемам этого класса. Однако есть такие аффиксальные морфемы, которые могут присоединяться только к отдельным корневым морфемам. К числу таких морфем относятся аффиксы возвратного и страдательного залогов для глаголов.

Например:

карал ‘чернеть’ – не может присоединять алломорфы, выражающие категории страдательного и возвратного залогов;

кара ‘смотреть’ – может присоединять аффиксы и страдательного, и возвратного залогов:

карал ‘(кара+ Бл) ‘быть просмотренным’

каран ‘(кара+ Бн) ‘осматриваться’

Такие нерегулярные сочетания в нашей модели описываются для каждой корневой морфемы индивидуально.

Например:

Наименование морфемы: кара

Морфотактический класс: V

Нерегулярные морфемы: -[Ы]л, -[Ы]н

Информацию этого параметра можно использовать в обучающих и информационно-справочных системах. В программах морфологического анализа и синтеза эти данные использовать неэффективно, так как придется выполнять действия в три этапа:

- выбирать морфему по правилам следования;

- выбирать алломорф по правилам выбора алломорфа из морфемы;

- проверять правила чередования букв.

Для этих целей эффективнее использовать правила следования алломорфов, которые описываются в морфонологическом аспекте модели морфем.

Морфонологический аспект. Правила следования алломорфов

Морфонологический аспект состоит из следующих параметров:

1. таблица алломорфов;

2. правила сочетания алломорфов:

a. сочетание алломорфов корневой и аффиксальной морфем;

b. сочетание алломорфов аффиксальных морфем.

Таблица алломорфов имеет следующий вид:

Таблица 2

Алломорф	Код	Терминальность
сы	01.2.004.01	+
се	01.2.004.02	+
ы	01.2.004.03	+
е	01.2.004.04	+
сын	01.2.004.05	-
сен	01.2.004.06	-
ын	01.2.004.07	-
ен	01.2.004.08	-

Код алломорфа – параметр, необходимый для однозначной идентификации алломорфа, и используется программными модулями типа морфологического анализатора.

Параметр «Терминальность» показывает, может ли данный алломорф быть последним в словоформе или после него обязательно должен быть другой алломорф. В приведенном примере для алломорфов морфемы -[с]Ы[н] последние 4 алломорфа являются нетерминальными и используются только в тех словоформах, где после них идет алломорф падежной морфемы.

Например:

алма+сын+а ‘его яблоку’

алма+сын+нан ‘из его яблока’

алма+сын+да ‘в его яблоке’

Этот параметр также необходим для программных модулей, работающих с правилами сочетания алломорфов, в частности, для морфологического анализатора.

Правила сочетания алломорфов

Правила сочетания алломорфов состоят из правил двух видов:

- правила сочетания аффиксальных алломорфов;

Таблица 3

- правила сочетания корневого алломорфа с аффиксальным алломорфом.

Правила сочетания аффиксальных алломорфов аналогичны правилам сочетания аффиксальных морфем. Разница только в том, что в таблице приводятся сочетания не морфем, а алломорфов. Плюс данного варианта представления в том, что он оптимизирует работу программных модулей, работающих на морфологическом уровне. Модуль морфологического анализа вместо трех действий, описанных выше при работе с правилами следования морфем, производит только одно, сразу выбирает нужный алломорф из таблицы следования. Возможные фонетические чередования также заложены в самой таблице.

Правила сочетания корневого алломорфа с аффиксальным алломорфом также похожи на правила сочетания корневых и аффиксальных морфем. Все корневые алломорфы классифицируются по морфонологическим типам, и для каждого типа составляется своя таблица следования. Для словаря татарских корневых морфем это 83 морфонологических класса. Для словаря казахских корневых морфем – 55 морфонологических класса. Причем общее количество алломорфов в обоих языках примерно одинаково. Эта цифра показывает, что в татарском языке намного больше исключений, для которых нужно вводить отдельный морфонологический класс.

Например, такие исключения, как:

йөр 'ходи' – йөри 'ходит';

халык 'народ' – халкы 'его народ'

хольк 'характер' – холькы 'характер';

милек 'собственность' – милке 'собственность'.

Семантический аспект. Синонимия

Семантический аспект модели содержит параметр – «Синонимия». Этот параметр служит в модели для связи между собой синонимичных морфем. Все морфемы, как корневые, так и аффиксальные, всех языков, описанных в модели, делятся на синонимичные классы, и каждому классу присваивается свой уникальный номер. В этом параметре модели для каждой морфемы указывается номер его синонимичного класса. Так, в таблице 3 приведен список корневых морфем, у которых в параметре «Синонимия» указан 734-й синонимичный класс. Все корневые морфемы в таблице принадлежат разным языковым подмоделям многофункциональной модели, и информация о названии языковой подмодели приведена в поле «Язык».

Морфема	Язык	Синонимия
арлекин	татарский	734
балаганчы	татарский	734
кәмитче	татарский	734
комик	татарский	734
мәзәкче	татарский	734
мәсхәрәче	татарский	734
шамакай	татарский	734
арлекин	казахский	734
қалжыңқой	казахский	734
қуақы	казахский	734
қылжақпас	казахский	734
масқарапаз	казахский	734
arleken	турецкий	734
soytarı	турецкий	734
masxaraboz	узбекский	734
mazaxchi	узбекский	734

Аффиксальные морфемы также классифицируются по синонимичным группам, и в таблице 4 приведен пример списка аффиксальных морфем разных тюркских языков, имеющих одинаковое значение идентификатора синонимического класса.

Таблица 4

Морфема	Язык	Синонимичный класс
-У	Татарский	35
-[Ы]ш	Татарский	35
-У	Казахский	35
-[У]в	Крымскотатарский	35
-[Ы]ш	Крымскотатарский	35
-мА	Крымскотатарский	35
-мА	Турецкий	35

Параметр синонимичного класса может быть использован в различных прикладных программах: системах многоязычного поиска, системах машинного перевода среди тюркских языков.

Для заполнения содержимого многофункциональной модели тюркской морфемы создана База данных и программное обеспечение, которое находится на стадии разработки и реализуется в виде веб-ресурса. Параллельно производится заполнение базы данных модели для трех тюркских языков: татарского, казахского и крымскотатарского. На основе базы данных, заполненных для этих языков, разработан морфологический анализатор (Рис.2), программная часть которого языконезависима и универсальна для всех тюркских языков. Для запуска его с другим языком достаточно только указать, файлы какого языка использовать в процессе анализа.

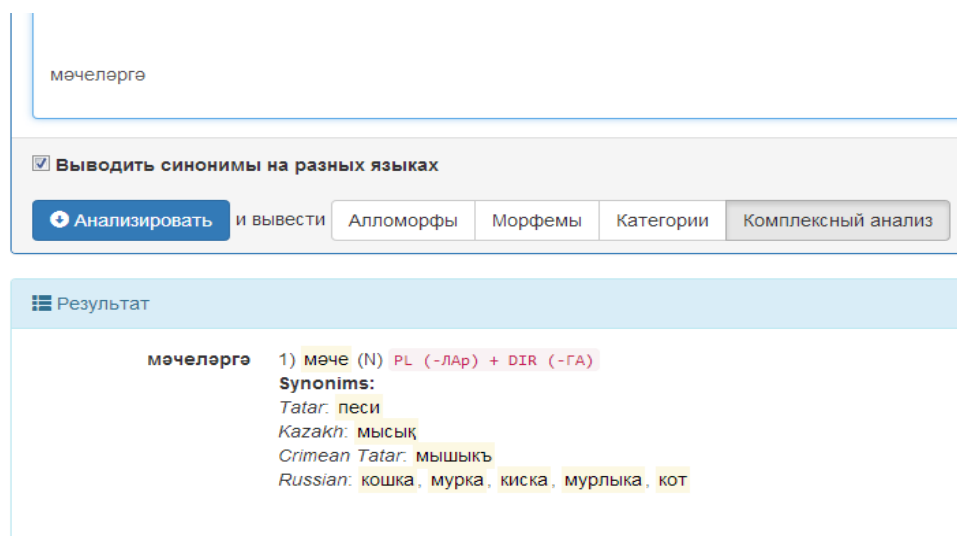


Рис.2. Фрагмент интерфейса морфологического анализатора

На рисунке 2 видно, что модуль морфологического анализатора, кроме того, что производит морфологический анализ, еще и выдает семантические эквиваленты на разных языках с аналогичным синонимичным классом.

Заключение

В статье дается концептуальное описание многофункциональной лингвистической модели тюркских морфем как информационно-программной оболочки, предназначенной для использования ее в качестве ресурсной базы для программных продуктов, осуществляющих компьютерную обработку тюркских языков, в качестве информационно-справочной системы, содержащей практически полную информацию о тюркских морфемах, а также в качестве инструментария для исследований ученых-тюркологов. Весьма конструктивным и продуктивным представляется использование данной многофункциональной и многоязычной модели тюркских морфем в качестве одного из центральных, ядерных, модулей в едином веб-портале для тюркских языков. Авторы статьи выражают также надежду, что данный проект послужит интеграции усилий ученых-тюркологов для расширения базы данных описаниями различных тюркских языков, что обеспечит эффективное использование многофункциональной модели в качестве технологического инструментария и межъязыкового модуля в системах компьютерной обработки тюркских языков.

Список литературы

Дыбо А. В., Шеймович А. В. Автоматический морфологический анализ для корпусов тюркских языков // Филология и культура. Philology and Culture. 2014. №2. С. 20–27.

Желтов П. В. Морфологический анализатор чувашского языка // Материалы международной конференции студентов и аспирантов по фундаментальным наукам «Ломоносов 2002», М.: изд-во. МГУ, 2002. С.11.

Сиразитдинов З. А. Алгоритмическая грамматика словоизменения башкирского языка. URL: <http://212.193.132.98/mfbl/res/bashdb/algram/algram.htm> (дата обращения: 1.04.2016).

Сулейманов Д. Ш. К разработке эффективной морфологии среднетюркского языка / В сб. трудов междунар. конф. Проблемы создания среднетюркского языка Ортаюрк. Ташкент: [б.и.], 1993. С. 85–86.

Сулейманов Д. Ш., Гатиатуллин А. Р. Структурно-функциональная компьютерная модель татарских морфем. Казань: Фэн, 2003. 220с.

Сулейманов Д. Ш., Гатиатуллин А. Р., Карабаева С. Ж. К разработке функционально-структурной модели аффиксальных морфем языков тюркской группы (на примере киргизского и татарского языков) // Казанская наука. 2013. № 6. Казань: Изд-во Казанск. Изд. дом, 2013. С. 220–223.

Сулейманов Д. Ш., Гатиатуллин А. Р., Ваганов Д. Р. Семантико-синтаксическая модель татарского предложения в контексте реляционно-ситуационной системы // В сб. трудов «Открытые семантические технологии проектирования интеллектуальных систем» = Open Semantic Technologies for Intelligent Systems (OSTIS-2013): матер. III Междунар. научн.-техн. конф. (Минск, 21-23 февраля 2013г.) / редкол.: В.В. Голенков (отв. ред) [и др.]. Минск: БГИУР, 2013. С.329–332.

Труды Первой международной конференции «Компьютерная обработка тюркских языков». Астана: ЕНУ им. Л.Н. Гумилева, 2013. 345с.

Тукеев У. А., Шормакова А. Н., Айткулова А., Сундетова А. Разработка Англо-Казахского Переводчика На Базе Open-Source Платформы Apertium. // Сулейман Демирел Атындагы Университет Хабаршысы. 2014, Том №28, С. 99 – 111.

Batuer Aisha and Maosong Sun. A Uyghur Morpheme Analysis Method based on Conditional Random Fields, International Journal of Asian Language Processing, 2009, 19(2): pp. 69–83.

Gulila Altenbek. Automatic Morphological Tagging of Contemporary Uighur Corpus. The IEEE International Conference on Information Reuse and Integration, 2006, Hawaii, USA, 2006, pp. 557–560.

Kemal Oflazer. Two-level Description of Turkish Morphology. Literary and Linguistic Computing, Vol. 9, No 2, 1994. pp. 137–148.

Orhun, M., Tantug A. C., Adali E. Morphological Disambiguation Rules For Uyghur Language // IEEE International Conference on Software Engineering and Service Sciences (ICSESS), Beijing, China, 2010. pp. 542–547.

Proceedings of the International Conference on Turkic Language Processing (TURKLANG-2014). (Istanbul, November 6-7, 2014). Istanbul: Ozkaracan Matbaacilik-Bagcilarr, 2014. 135 p.

Proceedings of the International Conference «Turkic Languages Processing: TurkLang-2015» Kazan: Academy of Sciences of the Republic of Tatarstan Press, 2015. 488 p.

Qetin, M., Amasyali, M. F. Active Learning for Turkish Sentiment Analysis // In Innovations in Intelligent Systems and Applications (INISTA), IEEE International Symposium. (Albena, Bulgaria, June 19-21, 2013). pp. 160–164.

Sharipbay A. A., Bekmanova G., Yergesh B., Mukanov A. Synchronized liner tree for morphological analysis and generation of the Kazakh language. - Proceedings of the international conference “Turkic languages processing”, TurkLang 2014, Istanbul, pp. 113–117.

Yildirim E., Tantug A.C. “The feasibility analysis of re-ranking for N-best lists on English-Turkish machine translation”, IEEE International Symposium on Innovations in Intelligent Systems and Applications. (Albena, Bulgaria, June 19-21, 2013). pp. 179–184.

References

Dybo, A. V., Sheimovich, A. V. (2014). *Avtomaticheskii morfologicheskii analiz dlia korpusov tiurkskikh iazykov* [Automatic Morphological Analysis for Corpora of Turkic Languages]. Filologiya i kul'tura. Philology and Culture, 2014. No. 2, pp. 20–27. (In Russian)

Batuer, Aisha and Maosong Sun. (2009). *A Uyghur Morpheme Analysis Method Based on Conditional Random Fields*. International Journal of Asian Language Processing, 2009, 19(2), pp. 69–83. (In English)

Gulila Altenbek. (2006). *Automatic Morphological Tagging of Contemporary Uighur Corpus*. The IEEE International Conference on Information Reuse and Integration, 2006, pp. 557–560. Hawaii, USA. (In English)

Kemal, Oflazer. (1994). *Two-level Description of Turkish Morphology*. Literary and Linguistic Computing, 1994, Vol. 9, No 2, pp. 137–148. (In English)

Orhun, M., Tantug, A. C., Adali, E. (2010). *Morphological Disambiguation Rules for the Uyghur Language* // IEEE International Conference on Software Engineering and Service Sciences (ICSESS), Beijing, China, pp. 542–547. (In English)

Proceedings of the International Conference on Turkic Language Processing (TURKLANG-2014). (Istanbul, November 6-7, 2014). 135 p. Istanbul, Ozkaracan Matbaacilik-Bagcilarr. (In English)

Proceedings of the International Conference “Turkic Languages Processing: TurkLang-2015”. 488p. Kazan, Academy of Sciences of the Republic of Tatarstan Press. (In English, in Russian)

Qetin, M., Amasyali, M. F. (2013). *Active Learning for Turkish Sentiment Analysis*. In Innovations in Intelligent Systems and Applications (INISTA), IEEE International Symposium. (Albena, Bulgaria, June 19-21, 2013). Pp. 160–164. (In English)

Sharipbay, A. A., Bekmanova, G., Yergesh, B., Mukanov, A. (2014). *Synchronized Liner Tree for Morphological Analysis and Generation of the Kazakh Language*. Proceedings of the international conference “Turkic languages processing”, TurkLang, Istanbul, pp. 113–117. (In English)

Sirazitdinov, Z. A. Algoritmicheskaya grammatika slovoizmeneniia bashkirskogo iazyka [Algorithmic Inflection Grammar of the Bashkir Language]. URL: <http://212.193.132.98/mfbl/res/bashdb/algram/algram.htm> (accessed: 1.04.2016). (In Russian)

Suleimanov, D. Sh. (1993). K razrabotke effektivnoi morfologii srednetiurkskogo iazyka [On the Development of an Effective Morphology of the Middle Turkic Language]. V sb. Trudov mezhd. konf. «Problemy sozdaniia srednetiurkskogo iazyka Ortatiurk». Pp.85–86. Tashkent, [b.i.]. (In Russian)

Suleimanov, D. Sh., Gatiatullin, A. R. (2003). Strukturno-funksional'naiia komp'iuternaia model' tatarskikh morfem [Structural and Functional Computer Model of the Tatar Morpheme]. 220 p. Kazan', Fen. (In Russian)

Suleimanov, D. Sh., Gatiatullin, A. R., Karabaeva, S. Zh. (2013). K razrabotke funktsional'no-strukturnoi modeli affiksial'nykh morfem iazykov tiurkskoi gruppy (na primere kirgizskogo i tatarskogo iazykov). / Kazanskaia nauka, No. 6, pp. 220–223. Kazan', Izd-vo Kazansk. Izd. Dom. (In Russian)

Suleimanov, D. Sh., Gatiatullin, A. R., Vagapov, D. R. (2013). Semantiko-sintaksicheskaya model' tatarskogo predlozheniia v kontekste reliatsionno-situatsionnoi sistemy [Semantico-Syntactic Model of the Tatar Sentence in the Context of Relational-Situational System]. V sb. trudov «Otkrytie semanticheskie tekhnologii proektirovaniia intellektualnikh sistem» = Open Semantic Technologies for Intelligent Systems (OSTIS-2013): mater. III Mezhd. nauchn.-tekhn. konf. (Minsk, 21-23 fevralia 2013g.) / redkol.: V.V. Golenkov (otv. red) [i dr.]. Pp. 329–332. Minsk, BGUIR. (In Russian)

Trudy Pervoi mezhdunarodnoi konferentsii «Komp'iuternaia obrabotka tiurkskikh iazykov» (2013) [Proceedings of the First International Conference "Computer Processing of Turkic Languages"]. 345 p. Astana, ENU im. L.N. Gumileva. (In Russian)

Tukeev, U. A., Shormakova, A. N., Aitkulova, A., Sundetova, A. (2014). Razrabotka Anglo-Kazakhskogo Perevodchika Na Baze Open-Source Platformy Apertium [The Development of English-Kazakh Translator Based on the Open-Source Platform Apertium]. Suleiman Demirel Atyndagy Universitet Khabarshysy. T. 28, pp. 99 – 111. (In Russian)

Yildirim, E., Tantug, A.C. (2013). "The Feasibility Analysis of Re-ranking for N-best Lists on English-Turkish Machine Translation", IEEE International Symposium on Innovations in Intelligent Systems and Applications. (Albena, Bulgaria, June 19-21, 2013). Pp. 179–184. (In English)

Zhel'tov, P. V. (2002). Morfologicheskii analizator chuvashskogo iazyka [Morphological Analyzer of the Chuvash Language]. Materialy mezhdunarodnoi konferentsii studentov i aspirantov po fundamental'nym naukam «Lomonosov 2002». Pp. 11. Moscow, izd-vo MGU. (In Russian)

The article was submitted on 12.04.2016
Поступила в редакцию 12.04.2016

Сулейманов Джавдет Шевкетович,
доктор технических наук,
директор НИИ «Прикладная семиотика» АН РТ,
420111, Россия, Казань,
Лево-Булачная, 36а
dvdt.slt@gmail.com

Suleymanov Dzhavdet Shevketovich,
Doctor of Technical Sciences,
Director of the Research Institute of Applied Semiotics, Tatarstan Academy of Sciences,
36a Levo-Bulachnaya Str.,
Kazan, 420111, Russian Federation.
dvdt.slt@gmail.com

Гатиатуллин Айрат Рафизович,
кандидат технических наук,
заведующий отделом,
НИИ «Прикладная семиотика» АН РТ,

420111, Россия, Казань,
Лево-Булачная, 36а
agat1972@mail.ru

Gatiatullin Ayrat Rafizovich,
Ph.D. in Technical Sciences,
Head of the Department,
the Research Institute of Applied Semiotics, Tatarstan Academy of Sciences,
36a Levo-Bulachnaya Str.,
Kazan, 420111, Russian Federation.
agat1972@mail.ru

Альменова Акмарал Байжановна,
младший научный сотрудник,
НИИ «Прикладная семиотика» АН РТ,

420111, Россия, Казань,
Лево-Булачная, 36а
almen_akmaral-baijan@mail.ru

Almenova Akmaral Baizhanovna,
junior researcher,
the Research Institute of Applied Semiotics, Tatarstan Academy of Sciences,
36a Levo-Bulachnaya Str.,
Kazan, 420111, Russian Federation.
almen_akmaral-baijan@mail.ru

Баширов Артур Маратович,
инженер, ООО «ТемирТех»,
заместитель директора
по техническому развитию,
420033, Россия, Казань,
Кулахметова, 10
a.basheerov@gmail.com

Bashirov Artur Maratovich,
engineer, TemirTech, LLC,
Vice-Director for Technical Development,

10 Kulakhmetova Str.,
Kazan, 420033, Russian Federation
a.basheerov@gmail.com