

– ИТ В ИССЛЕДОВАНИЯХ АЛТАЙСКИХ ЯЗЫКОВ –

DOI: 10.25587/2782-662-2023-4-92-100

УДК 004.8+81'322

Когнитивные исследования лексико-грамматического потенциала татарского языка для создания новых интеллектуальных систем обработки данных

Д. Ш. Сулейманов, Р. А. Гильмуллин, Н.А. Прокопьев, М.Р. Галимов

Академия наук Республики Татарстан, Институт прикладной семиотики, г. Казань, Россия

E-mail: dvdt.slt@gmail.com

E-mail: rinatgilullin@gmail.com

E-mail: nikolai.prokopyev@gmail.com

E-mail: magl.galimov@gmail.com

Аннотация. В статье рассматриваются когнитивные свойства лексико-грамматических структур, удовлетворяющие показателям технологичности татарского языка, приводится классификация лексико-грамматических структур в соответствии с когнитивными механизмами представления и обработки знаний. Такие исследования особенно актуальны в пространстве агглютинативных языков, к которым относится татарский язык, который характеризуется достаточно сложной и, одновременно, практически регулярной морфологией, позволяющей в одной словоформе закодировать практически целую субъектно-предикативную ситуацию, описываемую в флективно-аналитических языках, таких как английский язык, множеством предложений. Одним из главных результатов исследований станет реализация инструментария, позволяющего описывать и формализовать лексико-грамматические структуры на основе семиотических универсалий. Разработанные при помощи данного ин-

СУЛЕЙМАНОВ Джавдет Шекветович – д. т. н., профессор, академик АН РТ, г. н. с. Академии наук Республики Татарстан, Институт прикладной семиотики.

E-mail: dvdt.slt@gmail.com

SULEYMANOV Djavdet – Doctor of Technical Sciences, Professor, Chief Scientific Associate, Academy of Sciences of the Republic of Tatarstan, Institute of Applied Semiotics.

ГИЛЬМУЛЛИН Ринат Абрекович – к. ф.-м. н., директор, Академия наук Республики Татарстан, Институт прикладной семиотики.

E-mail: rinatgilullin@gmail.com

GILMULLIN Rinat – Candidate of Physical and Mathematical Sciences, Director, Academy of Sciences of the Republic of Tatarstan, Institute of Applied Semiotics.

ПРОКОПЬЕВ Николай Аркадьевич – научный сотрудник, Академия наук Республики Татарстан, Институт прикладной семиотики.

E-mail: nikolai.prokopyev@gmail.com

PROKOPYEV Nikolay – research assistant, Academy of Sciences of the Republic of Tatarstan, Institute of Applied Semiotics.

ГАЛИМОВ Мансур Радикович – инженер-программист, Академия наук Республики Татарстан, Институт прикладной семиотики.

E-mail: magl.galimov@gmail.com

GALIMOV Mansur – software engineer, Academy of Sciences of the Republic of Tatarstan, Institute of Applied Semiotics.

струментария формальные модели когнитивных лексико-грамматических структур татарского языка могут служить основой для создания эффективных интеллектуальных систем и технологий на базе принципиально новой архитектуры систем обработки информации.

Ключевые слова: инфокоммуникационные технологии, лексико-грамматические структуры, татарский язык, когнитивные механизмы, морфология, рекурсия, активность знаний, интеллектуальные системы.

Исследования выполнены в рамках гранта российского научного фонда 23-21-10083.

Для цитирования: Сулейманов Д. Ш., Гильмуллин Р. А., Прокопьев, Н.А., Галимов М.Р. Когнитивные исследования лексико-грамматического потенциала татарского языка для создания новых интеллектуальных систем обработки данных // DOI: 10.25587/2782-662-2023-4-92-100

Cognitive researches of the Lexical and Grammatical potential of the Tatar language for the Development of New Intelligent Systems

D. Sh. Suleymanov, R. A. Gilmullin, N. A. Prokopyev, R. M. Gilmullin

Academy of Sciences of the Republic of Tatarstan, Institute of Applied Semiotics, Kazan, Russia

E-mail: dvdt.slt@gmail.com

E-mail: rinatgilullin@gmail.com

E-mail: nikolai.prokopyev@gmail.com

E-mail: magl.galimov@gmail.com

Abstract. In this article researches the cognitive properties of lexico-grammatical structures that corresponding the technological indicators of the Tatar language, provides a classification of lexico-grammatical structures in accordance with the cognitive mechanisms of knowledge representation and processing. Such studies are especially relevant in the space of agglutinative languages, which include the Tatar language, which are characterized by a fairly complex and, at the same time, almost regular morphology, which allows encoding in one word form almost the entire subject-predicative situation described in inflectional-analytic languages such as English, many offers. One of the main results of the research will be the implementation of tools that make it possible to describe and formalize lexical and grammatical structures based on semiotic universals. Formal models of cognitive lexico-grammatical structures of the Tatar language developed using this toolkit can serve as the basis for creating effective intelligent systems and technologies based on a fundamentally new architecture of information processing systems.

Keywords: IT, lexical and grammatical structures, Tatar language, cognitive mechanisms, morphology, recursion, knowledge activity, intelligent systems.

This study was partially funded by the Russian Foundation for Basic Research (grant no. 23-21-10083).

For citation: Suleymanov D. Sh., Gilmullin R. A., Prokopyev N. A., Gilmullin R. M. Cognitive researches of the Lexical and Grammatical potential of the Tatar language for the Development of New Intelligent Systems // DOI: 10.25587/2782-662-2023-4-92-100

Введение

В исследовании естественных языков (ЕЯ) можно выделить три аспекта: когнитивный, коммуникативный и технологический. Когнитивный аспект – это характеристика естественного языка с точки зрения возможностей описания модели мира, представления знаний. Коммуникативный аспект отражает потенциал естественного языка для кодирования, приема и передачи, семиотической обработки информации, организации диалога. Технологический аспект определяет формальный и концептуальный потенциал естественного языка для реализации средств эффективной обработки, адекватного описания и компактного хранения информации на данном языке, создания эргономичных технических средств, учитывающих специфику языка (например, частотность букв при разработке клавиатуры), а также для разработки интеллектуального программного инструментария, включая операционные системы. Очевидно, в основе искусственных языков и систем программирования лежат глубинные структуры, соответственно, ментальность естественного языка и, таким

образом, эти системы реализуют описательный и вычислительный потенциал соответствующего естественного языка.

Основной задачей фундаментальных исследований когнитивного потенциала татарского языка для создания интеллектуальных технологий являются разработка и использование семиотических моделей лексико-грамматических конструкций тюркских языков как формальной основы интеллектуальных технологий, а также лексического корпуса как базы данных. Среди важных признаков интеллектуальности систем принятия решений, как правило, выделяются такие свойства, как активность знаний, то есть первичность анализа данных и вторичность принятия решения на основе этого анализа; возможность оперировать нечеткой информацией, семантически управляемой контекстом, и исполнять нечеткие команды. В данной работе на конкретных примерах показывается перспектива исследования в этих целях татарского языка как языка агглютинативного типа, обладающего такими важными свойствами, как регулярность, рекурсия, активность знаний и ряд других. Работы этого направления создают перспективу построения универсального языка общения систем Искусственного Интеллекта (ИИ) между собой и с человеком, а также интеллектуальной операционной системы, как когнитивно-коммуникативной системы ИИ.

Классификация когнитивных лексико-грамматических структур, определяющих технологичность языка

Известно, что разработка моделей ЕЯ, исследование их возможностей для разработки языков искусственного интеллекта входят в число базовых проблем в области построения интеллектуальных систем. Такие задачи, как компьютерная обработка больших массивов ЕЯ-текстов, ЕЯ-диалог с системой, создание больших банков информации (бигдата) на основе ЕЯ, разработка языков-посредников в многоязычной информационной среде, базирующихся на более развитых лингвистических моделях, приобретают особую актуальность в связи с развитием глобальных компьютерных сетей и формированием больших объемов распределенных данных. В исследовании естественных языков можно выделить три аспекта: когнитивный, коммуникативный и технологический. Когнитивный аспект – это характеристика естественного языка с точки зрения возможностей описания модели мира, представления знаний. Коммуникативный аспект отражает потенциал естественного языка для кодирования, приема и передачи, семиотической обработки информации, организации диалога. Технологический аспект определяет формальный и концептуальный потенциал естественного языка для реализации средств эффективной обработки, адекватного описания и компактного хранения информации на данном языке, создания эргономичных технических средств, учитывающих специфику языка (например, частотность букв при разработке клавиатуры), а также для разработки интеллектуального программного инструментария, включая операционные системы. Очевидно, в основе искусственных языков и систем программирования лежат глубинные структуры, ментальность естественного языка и, таким образом, эти системы реализуют описательный и вычислительный потенциал соответствующего ЕЯ. На наш взгляд, современные средства накопления и обработки знаний на естественном языке малоэффективны и практически не справляются с такими задачами, как эффективный поиск и отбор информации в распределенных базах данных, извлечение знаний, семантический анализ текстовой информации. Причиной этому, прежде всего, служит то, что они изначально являются неинтеллектуальными, созданы на основе примитивных искусственных языков программирования, практически представляющих собой подмножество флективно-аналитических языков или искусственных конструкций, созданных на их основе.

В связи с этим перспективным представляется разработка нового программного инструментария по следующей технологии: 1) исследование и выявление естественных грамматических (морфологических, синтаксических, семантических) конструкций в различных языках, достаточно регулярных и обладающих естественной сложностью, в целях создания на их базе языков искусственного интеллекта нового поколения; 2) разработка языка-посредника на основе подмножеств и конструкций языков с определенными свойствами, позволяющими наиболее адекватно и сжато описывать контекст и быстро обрабатывать тексты на ЕЯ.

Как известно, для систем обработки знаний важны следующие характеристики, определяющие их эффективность и интеллектуальность: 1) время обработки; 2) объем памяти для хранения информации; 3) компактность хранения и передачи смысла; 4) возможность кодирования и обработки нечеткой информации; 5) активность знаний. Причем первые три параметра описывают эффективность, а параметры 4 и 5 – интеллектуальность систем и технологий. Как показывают исследования [2], татарский язык, являясь одним из тюркских языков, имеет богатую, сложную, но достаточно регулярную морфологию, обладает потенциалом, позволяющим эффективно кодировать и компактно хранить информацию, а также реализовывать на уровне аффиксальных морфем такие явления, как рекурсия, «нечеткость».

В объектно-предикативной модели мира именные группы, как правило, маркируют некое состояние объекта или объектов, в то время как действие, отношения между объектами и группой объектов описываются глагольной группой. Соответственно, выделяются когнитивные механизмы, реализуемые в рамках именной группы и когнитивные механизмы, реализуемые в рамках глагольной группы. Кроме того, сама структура текста, которая определяется синтаксическими закономерностями языка, служит одним из когнитивных механизмов языка, управляющих в тексте такой важной характеристикой, как активность знаний, естественным образом реализуя логическую схему: анализ-действие (известно, что активность знаний есть один из важных признаков интеллектуальной системы).

Далее проиллюстрируем на примерах соответствующие когнитивные формализмы, выделенные в татарском языке.

Регулярная и естественная сложность морфологии

Когнитивные механизмы при описании состояния объектов. Как известно, татарская морфология является регулярной, почти автоматной [2], и в то же время имеет естественную сложность, которая заключается, прежде всего, в следующем: 1) возможность присоединения определенных аффиксальных морфем, присоединяясь справа, «переключающих» слова одного типа в другой, например, превращающих именную словоформу в глагольную или в форму прилагательного, и наоборот (уй-уйла: мысль-размышляй; кара-карау-караучы: наблюдай-наблюдение-наблюдатель); 2) морфологическое (синтетическое) задание признаков модальности, настроения, эмоционально-личностного отношения к ситуации, объекту или процессу, описываемых данной словоформой (кояш-кояштай: солнце-словно солнце; алды-алдымыни: взял-неужели взял); 3) контекстное разнообразие значений аффикса. Известно, что именная группа, как правило, кодирует некую семантическую ролевую ситуацию, а глагольная группа – контекстные отношения над этими ролями. Таким образом, возможность перехода с именной формы к глагольной и, наоборот, через присоединение соответствующих аффиксов позволяет описывать одновременно в пределах одной словоформы как сложную ролевую ситуацию, так и контекстные отношения между семантическими ролями. Тем самым обеспечивается компактность описания и хранения информации. Синтетический, аффиксальный способ словоизменения обеспечивает коди-

рование в рамках одной словоформы некоторого значения, описываемого на флективно-аналитических языках (например, на английском) несколькими словосочетаниями и даже предложениями.

Вместе с тем морфология в большой степени регулярна, близка к автоматной, с небольшим количеством исключений из правил, что обеспечивает минимизацию ёмкостных и временных функций при обработке текстов на татарском языке, а также упрощает анализ структуры и значения словоформы, несмотря на естественную сложность морфологии.

Важным свойством татарской морфологии, наряду с ее регулярностью, является «жесткое», фиксированное, размещение аффиксов в последовательности аффиксальных морфем. Регулярность морфологии означает, что одна и та же схема сочетания морфем (морфотактика) присуща всем или почти всем именным и глагольным группам. Это дает возможность по одной и той же схеме, практически автоматически образовывать словоформы с одинаковыми глубинными значениями аффиксов.

Например:

1) кул, куллар, кулларым, кулларыма – ('рука, руки, мои руки, моим рукам');

2) бакча, бакчалар, бакчаларым, бакчаларыма – ('сад, сады, мои сады, моим садам').

Именные корневые морфемы кул ('рука'), бакча ('сад') имеют одни и те же последовательности аффиксальных морфем с идентичными значениями. Обобщенно приведенные парадигмы описываются следующими схемами:

Х (Имя сущ.), Х (Имя сущ.) + лар (афф. мн.), Х (Имя сущ.) + лар (афф. мн.) + ым (афф. притяж., 1 л., ед. ч.), Х (Имя сущ.) + лар (афф. мн.) + ым (афф. притяж., 1 л., ед. ч.) + а (афф. падежн., дат. падеж').

Позиции аффиксальных морфем, составляющих словоформу, неизменны относительно друг друга. Таким образом, мы определили два первых когнитивных механизма татарского языка: 1) регулярность морфотактики и 2) фиксированность позиций соответствующих типов аффиксальных морфем.

Морфологический эллипсис

Следующая возможность в татарской морфологии, которая может быть отнесена к третьему когнитивному механизму (компактность), называется морфологический эллипсис, это: возможность пропуска последовательности аффиксов при однородных именных словоформах с сохранением ее в последней словоформе.

То есть, возможность вывода любой последовательности аффиксов, общих для однородных членов, вправо, за последовательность однородных членов и присоединение их к последнему справа однородному члену.

Например:

Ишек алды тавыкларга, казларга, сарыкларга тулы = Ишек алды тавык, каз, сарыкларга тулы (Двор полон кур, гусей, овец).

Мин кырларыбызга, урманнарыбызга, елагларыбызга шатланам = мин кыр, урман, елагларыбызга шатланам (Я радуюсь нашим полям, лесам, рекам).

Морфологическая рекурсия

Одним из важных и интересных когнитивных механизмов в татарском языке является рекурсия, т.е. возможность циклического порождения нового значения путем последовательного применения одной и той же «формулы», т.е. повторного присоединения одного и того же аффикса.

Таковыми свойствами обладают аффиксальные морфемы –ДАГЫ (локатив2) и –НЫКЫ (притяжат.), которые можно назвать также аффиксами неопределенности, т.е. аффиксами, придающими неопределенность к присоединенным лексемам (здесь и далее заглавные

буквы в аффиксальных морфемах обозначают вариативность символов в соответствующих позициях согласно закону сингармонизма). Например, пусть задана лексема ком ('песок'). Присоединение аффикса –дагы порождает новые объекты или свойства, являющиеся неопределенными: комдагы – 'нечто на песке'; комдагыдагы – 'нечто в нечто на песке'; комныкы – 'то, что принадлежит/присуще песку'; комныкыныкы – 'то, что принадлежит тому, что принадлежит/присуще песку'.

По такой формуле возможно образование словоформы практически неограниченной длины. Естественно, такие длинные последовательности морфем в нормальной речи почти не используются. Это прежде всего связано с проблемами глубины памяти человека, удобства общения между людьми. Тем не менее подобное словоизменение является совершенно корректным с точки зрения грамматики татарского языка, и словоформа, образованная присоединением любой последовательности, гипотетически всегда имеет смысл, конкретное значение приобретается при «погружении» словоформы в определенный контекст. Приведем пример со следующей словоформой: бакчаныкындагыныкыныкындагы, которая однозначно раскладывается на составляющие: бакча+ныкы+ндагы+ныкы+ныкы+ндагы 'сад' (имя сущ., притяж.+ локатив2+притяж.+притяж.+локатив2). Значение данной словоформы таково: 'нечто, которое находится на/в нечто, которое само принадлежит нечему, которое само принадлежит нечему, которое само находится на/в нечто, которое принадлежит сад'.

Нетрудно заметить, что, эксплицитно задавая параметры после каждой морфемы, можно получить контекстную определенность словоформы. В реальных случаях такие параметры задаются имплицитно (т. е. неявно), наполняясь конкретным значением в зависимости от контекста речи. Рассмотрим следующий пример для иллюстрации изложенного утверждения. Пусть после каждого аффикса неопределенности стоят параметры: бакча + ныкы(x_0) + ндагы(x_1) + ныкы(x_2) + ныкы(x_3) + ндагы(x_4) + ныкы(x_5), где x_i – контекстные объекты, т. е. объекты, которые либо приобретают конкретное значение из контекста, либо их задает пользователь ($i = 1, \dots, 4$). Тогда, придавая значения параметрам: x_0 = «вишня», x_1 = «ветка», x_2 = «птичка», x_3 = «крыло», x_4 = «перо», мы получаем следующее контекстное значение: «нечто (значение x_5 , придаваемое параметру последним аффиксом, осталось неопределенным) на пере, что принадлежит крылу, что принадлежит птичке, что на ветке, что принадлежит вишне».

Даже на этом коротком примере элементарный расчет показывает, что применение рекурсивных аффиксов приводит к сжатию информации и существенной экономии памяти.

Когнитивные механизмы при описании действий

Известно, что поверхностное, лексическое описание предикатов (действий, отношений), как правило, осуществляется глагольными группами. Свойства *регулярности морфотактики и фиксированности позиций соответствующих типов аффиксальных морфем* в татарском языке присущи также и глагольным группам.

В дополнение к когнитивным механизмам, выделенным на основе именных словоформ, рассмотрим следующие естественные когнитивные механизмы, проявляющиеся в глагольных словоформах.

Возможность рекурсивно задавать нечеткие команды и описывать нечеткие действия

Данное свойство кодируется глагольными аффиксами, занимающими позицию залога, т. е. сразу же после глагольной основы – ГАлА, – штЫр.

Например:

кара ('смотри') – 'смотри' (3 лицо, ед.ч., повел. накл.);

караштыр ('смотри/поглядывай время от времени');
 кара ('смотри') + итыр ('время от времени');
 караштыргала ('смотри время от времени, время от времени – реже');
 кара ('смотримой') + итыр ('время от времени') + гала ('время от времени');
 караштыргалаитыргалаитыр... ('смотри время от времени, время от времени, время от времени – и еще реже...');
 кара (смотри, корень, 3 лицо, ед.ч., повел.накл.) + итыр ('время от времени-изредка') + гала ('время от времени-еще реже') + итыр (еще реже) + гала (еще реже) + итыр (еще реже)...

Сам факт, насколько редко требуется мыть – определяется, исходя из контекстной информации (уход за ребенком – это один контекст, а за котом – другой).

Возможность рекурсивно описывать в рамках одной словоформы действия, относящиеся к целой ролевой ситуации

Реализация данного свойства обеспечивается рядом специальных глагольных аффиксов, занимающих также залоговую позицию: -н, -Ыш, -т, -ДЫр.

Рассмотрим изменения ролевой ситуации при присоединении соответствующих аффиксов на примере с глагольной словоформой кара ('смотри').

Участники действия: субъект S , объект-предмет O_k , где $k \geq 1$.

Для словоформы сора ('спроси') ролевая ситуация следующая:

S воздействие на O_k

Присоединение аффиксов -н, -Ыш, -т, -ДЫр приводит к изменениям, описанным ниже.

1) **-н:** каран – кара + н ('оглядывайся').

Ролевая ситуация: S воздействие S

2) **-Ыш:** караш – кара + иш ('помогай смотреть-ухаживать/смотри-ухаживай вместе вместе').

Участники действия: субъект S , объект-актор $A_{i,j}$, объект-предмет O_k , где i – номер группы объекта-актора, $i \geq 1$; j – число участников в группе i , $j \geq 1$.

Ролевая ситуация: S воздействие (помощь) $A_{i,j}$ и (S & $A_{i,j}$) воздействие (бросить) O_k .

3) **-т, -ДЫр:** карат – кара + т ('сделай так, чтобы смотрел/смотрели-ухаживал/ухаживали').

Ролевая ситуация: S воздействие $A_{i,j}$ -> $A_{i,j}$ воздействие (смотреть) O_k . Здесь стрелка -> означает импликацию.

караттыр – кара + т + тыр ('сделай так, чтобы сделали так, чтобы смотрели').

Ролевая ситуация: S воздействие $A_{i,j}$ -> $A_{i,j}$ воздействие $A_{l,m}$ -> $A_{l,m}$ воздействие (смотреть) O_k .

караттырт – кара + т + тыр + т ('сделай так, чтобы сделали так, чтобы сделали так, чтобы смотрели').

Ролевая ситуация: S воздействие $A_{i,j}$ -> $A_{i,j}$ воздействие $A_{l,m}$ -> $A_{l,m}$ воздействие $A_{s,t}$ -> $A_{s,t}$ воздействие (смотреть) O_k .

По такой формуле, подставляя новые определенные аффиксы, можно создавать все новые и новые ролевые ситуации и описывать процессы на лексическом уровне. Например, добавление аффикса **-Ыл** к последней полученной словоформе: караттыртыл превращает сам субъект в объект-предмет, объект воздействия, т.е. $S = O_k$.

Получается следующая ролевая ситуация: S воздействие $A_{i,j}$ -> $A_{i,j}$ воздействие $A_{l,m}$ -> $A_{l,m}$ воздействие $A_{s,t}$ -> $A_{s,t}$ воздействие (бросить) S .

Активность знаний

Известно, что английские предложения строятся по схеме S-V-O (subject-verb-object: субъект-глагол-объект), а татарские – по схеме: S-O-V. То есть, англичанин, если говорит,

например, о намерении сходить в кино, сначала скажет, пойдет или не пойдет, и только после этого выдаст информацию – куда, какое, зачем, с кем, когда и т. д. («We'll go to the cinema «Attila» with my friends tomorrow»). Очевидно, здесь действие управляет ситуацией. После того, как высказано однозначно намерение субъекта We'll go – мы пойдем, дальнейшая информация становится пассивной, практически не влияет на выбор способа действия или усложняет его. А в тексте на татарском языке сначала дается информация и ее анализ, и только после этого определяется – положительное или отрицательное, само действие. («Без дуслар белән иртәгә буласы «Аттила» киносына барабыз/бармыйбыз» – буквально: 'Мы с друзьями на завтрашний фильм «Аттила» пойду/не пойду').

В системах искусственного интеллекта это называется активностью знаний, что служит одним из важных признаков интеллектуальности системы. Для подобных систем естественным и основополагающим является стиль мышления: анализ-действие, мышление-цели-алгоритмы, а не командный стиль: действие-анализ, алгоритм-цель, как это реализовано в современных технологиях, основанных на менталитете английского языка. То есть в тюркских языках сначала анализируется, обрабатывается информация, а затем осуществляется некое действие: подбирается соответствующая адекватная модель представления знаний или выбираются соответствующие алгоритмы и схемы реализации, оптимальность и эффективность которых во многом определяется корректностью и полнотой анализа информации. Это можно назвать событийным программированием.

Такие возможности, естественные для татарского языка и закрепленные в его грамматике, позволяют ставить задачу о разработке интеллектуальных программ накопления и извлечения знаний в глобальных компьютерных сетях и даже интеллектуальной операционной системы что, как известно, становится сверхактуальной задачей в современном информационном мире.

Заключение

Когнитивные лексико-грамматические структуры описания, запоминания, хранения, обработки и передачи информации на естественном языке являются теми технологическими признаками, которые позволяют на поверхностном уровне вербализовать, фиксировать и отображать эксплицитно глубинные стереотипы, когнитивные модели и механизмы, описывающие ситуации и процессы в модели мира.

В данной работе на примере татарского языка исследованы лексико-грамматических структуры, определяющие технологичность языка, представляющих определенный методологический и практический интерес для создания программных средств эффективной обработки естественно-языковой информации.

Как показывают исследования [2–4], татарский язык, как агглютинативный язык, обладающий регулярной морфологией и вместе с тем естественной сложностью, разрешаемой контекстно, является эффективным инструментом для создания интеллектуальных систем обработки информации. В силу минимальных показателей временных и емкостных оценочных функций для генерации и анализа цепочек татарских словоформ (за счет регулярности) достигается эффективность при накоплении и обработке информации на татарском языке. Компактность передачи смысла текста на поверхностном, лексическом, уровне объясняется также возможностями языка синтетически, т. е. словоформой, кодировать смысл которой для других языков (английский, русский) формируется аналитически, чаще всего несколькими предложениями. Агглютинативность языка, алгоритмические закономерности, минимальность исключений, наличие мощного мета-аппарата, достаточная жесткость синтаксиса позволяют ставить задачу о построении языка промежуточной трансляции, т. е. языка-посредника на базе татарского языка, и даже разработки новых операционных систем на основе новой идеологии.

Литература

1. Пospelов, Д. Прикладная семиотика / Д. Пospelов, Г. Осипов. 1999. Доступно: <http://raii.org/library/ainews/1999/1/OSPOS.ZIP>.
2. Suleymanov D.Sh. (2010). Natural Cognitive Mechanisms in the Tatar language [Text]. In the Collection of the Vienna Proceedings of the Twentieth European Meeting in Cybernetics and Systems Research. Ed. by Robert Trappel. Vienna, Austria, 6-9 April 2010. P. 210-213.
3. Gatiatullin Ayrat, Suleymanov Dzhavdet, Prokopyev Nikolai, Khakimov Bulat (2020). About Turkic Morpheme Portal. In *Proceedings of the Computational Models in Language and Speech Workshop*. Kazan, Russian, November 12-13, 2020. Pp. 226-243 (<http://ceur-ws.org/Vol-2780/>).
4. Формальные модели и программные инструменты компьютерной обработки татарского языка / Р. Р. Гатауллин, А. Р. Гатиатуллин, Р. А. Гильмуллин [и др.]. Научное издание / Под редакцией Д. Ш. Сулейманова, А. Ф. Хусаинова ; Академия наук РТ, Институт прикладной семиотики АН РТ. – Казань : Изд-во Академии наук РТ, 2019. – 260 с.
5. Heintz J. and Schonig C. (1989). Turcic Morphology as Regular Language. *Central Asianic Journal (CFJ)*, P. 1-24.

References

1. Pospelov, D. Prikladnaja semiotika / D. Pospelov, G. Osipov. 1999. Dostupno: <http://raii.org/library/ainews/1999/1/OSPOS.ZIP>.
2. Suleymanov D.Sh. (2010). Natural Cognitive Mechanisms in the Tatar language [Text]. In the Collection of the Vienna Proceedings of the Twentieth European Meeting in Cybernetics and Systems Research. Ed. by Robert Trappel. Vienna, Austria, 6-9 April 2010. P. 210-213.
3. Gatiatullin Ayrat, Suleymanov Dzhavdet, Prokopyev Nikolai, Khakimov Bulat (2020). About Turkic Morpheme Portal. In *Proceedings of the Computational Models in Language and Speech Workshop*. Kazan, Russian, November 12-13, 2020. Pp. 226-243 (<http://ceur-ws.org/Vol-2780/>).
4. Formal'nye modeli i programmnye instrumenty komp'yuternoj obrabotki tatarskogo jazyka / R. R. Gataullin, A. R. Gatiatullin, R. A. Gil'mullin [i dr.]. Nauchnoe izdanie / Pod redakciej D. Sh. Sulejmanova, A. F. Husainova ; Akademija nauk RT, Institut prikladnoj semiotiki AN RT. – Kazan' : Izd-vo Akademii nauk RT, 2019. – 260 s.
5. Heintz J. and Schonig C. (1989). Turcic Morphology as Regular Language. *Central Asianic Journal (CFJ)*, P. 1-24.

