



Когнитивный потенциал естественных языков агглютинативного типа в интеллектуальных технологиях

© 2023, Д.Ш. Сулейманов, Р.А. Гильмуллин✉, А.Р. Гатиатуллин, Н.А. Прокопьев

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

Аннотация

Описан подход к разработке новых интеллектуальных технологий на основе исследования когнитивного потенциала и семиотического моделирования лексико-грамматических структур агглютинативных языков (на примере татарского языка). Проведена классификация когнитивных лексико-грамматических структур, определяющих информационную технологичность языка. На конкретных примерах показаны возможности татарского языка для компактного представления знаний, оперирования с нечёткой информацией, раскрыты такие важные свойства, как регулярность, естественная и конструктивная сложность, рекурсия, глубокий морфологический эллипсис, функциональное разнообразие и семантическая многовалентность аффиксов. Предложенный подход, основанный на семиотических моделях агглютинативного языка и применении децентрализованной системы технологических средств вербализации и контекстного распознавания значений элементов языка, может позволить получать эффективные решения для создания интеллектуальных инструментов накопления и обработки знаний. Работы в этом направлении создают перспективу построения когнитивно-коммуникативной системы искусственного интеллекта на базе универсального языка общения систем искусственного интеллекта между собой и с человеком.

Ключевые слова: когнитивное моделирование, татарский язык, лексико-грамматические структуры, морфология, рекурсия, нечёткие команды, активность знаний, интеллектуальные системы.

Для цитирования: Сулейманов Д.Ш., Гильмуллин Р.А., Гатиатуллин А.Р., Прокопьев Н.А. Когнитивный потенциал естественных языков агглютинативного типа в интеллектуальных технологиях // *Онтология проектирования*. 2023. Т.13, №4(50). С.496-506. DOI:10.18287/2223-9537-2023-13-4-496-506.

Финансирование: работа поддержана Фондом научных исследований (проект № 23-21-10083).

Конфликт интересов: авторы заявляют об отсутствии конфликта интересов.

Введение

В исследовании естественных языков (ЕЯ) можно выделить три аспекта: когнитивный, коммуникативный и технологический [1]. Когнитивный аспект – это характеристика ЕЯ с точки зрения возможностей описания модели мира, представления знаний, организации и отображения процессов мышления. Коммуникативный аспект отражает потенциал ЕЯ для кодирования, приёма и передачи, семиотической обработки информации, организации диалога. Технологический аспект определяет формальный и концептуальный потенциалы ЕЯ 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. В основе искусственных языков и систем программирования лежат глубинные

структуры (ментальность) ЕЯ и, таким образом, эти системы реализуют описательный и вычислительный потенциал соответствующего ЕЯ.

Опыт авторов в области разработок и использования инфокоммуникационных технологий позволяет отметить, что современные средства накопления и обработки знаний неэффективны и не справляются с такими задачами, как поиск и отбор информации в распределённых базах данных, извлечение знаний, семантический анализ текстовой информации. Причиной их неинтеллектуальности является главным образом то, что создаются они с использованием относительно примитивных искусственных языков программирования, представляющих собой подмножество флективно-аналитических языков или искусственных структур, созданных на основе ЕЯ. Морфосинтаксические структуры этих ЕЯ больше ориентированы на реализацию коммуникативных, а не когнитивных функций.

Другая причина неинтеллектуальности систем обработки ЕЯ связана с организацией их моделей на основе формальных систем, практически игнорирующих семантику, в частности, порождающих грамматик [2], что создаёт фундаментальные проблемы: монотонность результатов логического вывода и пассивность инструментов логико-семантического анализа информации. В работе [3] такая организация моделей ЕЯ названа глобальным подходом к исследованию ЕЯ. Эти методы мало подходят для описания агглютинативных языков в целом и татарского языка в частности, поскольку не позволяют эффективно описать важные когнитивные особенности языков этого типа. Например, таких, как рекурсивно-параметрические конструкции, фрактальность лексико-грамматических структур, многовалентность и проактивность корневых и аффиксальных морфем и др.

Основной задачей фундаментальных исследований когнитивного потенциала ЕЯ агглютинативного типа для создания интеллектуальных технологий является разработка и использование семиотических моделей лексико-грамматических конструкций тюркских языков, как формальной основы интеллектуальных технологий, а также лексического корпуса, как лингвистической базы данных. Среди важных признаков интеллектуальности систем принятия решений выделяются такие свойства, как активность знаний, т.е. первичность анализа данных и вторичность принятия решения; возможность оперировать нечёткой информацией, семантически управляемой контекстом, и исполнять нечёткие команды.

В предлагаемом подходе используются языковые инструменты обработки информации, разработанные с учётом свойства проактивности морфем агглютинативных языков. По функциональности такие инструменты являются агентами, подобно агентам в многоагентных системах. Эти инструменты менее автономны, привязаны к контексту и функционируют в тесной связи между собой в соответствии с условиями, сформированными координирующими модулями. Такие модули, разрабатываемые на основе семиотических моделей Поспелова-Полякова [4, 5], обеспечивают целостность и полноту использования доступной информации на каждом этапе функционирования системы обработки информации.

В данной работе на конкретных примерах показывается перспективность исследования в этих целях татарского языка как языка агглютинативного типа, обладающего такими важными свойствами, как регулярность, рекурсия, активность знаний и др.

1 Классификация когнитивных лексико-грамматических структур

Исследование возможностей ЕЯ для разработки языков ИИ входят в число базовых проблем в области построения интеллектуальных систем (ИС) [1]. Особенно важными являются исследования когнитивного потенциала ЕЯ исходя из того, что любой ЕЯ обладает необходимым лексико-грамматическим потенциалом для реализации базовых когнитивных функций, присутствие которых в системах интеллектуальной обработки информации в настоящее

время становится критически востребованной. Такие задачи, как: компьютерная обработка больших массивов ЕЯ-текстов; разработка онтологий в условиях нечёткости данных и динамичности пространства; представление и обработка разнородных и динамических графов знаний; ЕЯ-диалог с ИС; создание больших банков информации на основе ЕЯ; разработка языков-посредников в многоязычной информационной среде, базирующихся на развитых лингвистических моделях, - приобретают особую актуальность в связи с развитием глобальных компьютерных сетей, формированием больших объёмов распределённых данных, и необходимостью создания объяснительного ИИ.

Разработка нового поколения программного инструментария включает:

- исследование грамматических (морфологических, синтаксических, семантических, прагматических) конструкций в различных ЕЯ в целях создания на их базе языков ИИ нового поколения;
- разработку языка-посредника на основе подмножеств и конструкций языков с определёнными свойствами, позволяющими адекватно и сжато описывать контекст и быстро обрабатывать тексты на ЕЯ.

Для ИС обработки знаний важны следующие характеристики:

- время обработки;
- объём памяти для хранения информации;
- компактность хранения и передачи смысла;
- возможность кодирования и обработки нечеткой информации;
- активность знаний.

Первые три характеристики описывают эффективность, а последние – интеллектуальность систем. Татарский язык, являясь одним из тюркских языков, имеет богатую, сложную, регулярную морфологию, обладает потенциалом, позволяющим эффективно кодировать и компактно хранить информацию, а также реализовывать на уровне аффиксальных морфем такие явления, как рекурсия и нечёткость [6].

В объектно-предикативной модели мира именные группы, как правило, обозначают объект или объекты, в то время как действие, процессы, отношения между объектами и группами объектов описываются глагольной группой. Выделяются когнитивные механизмы, реализуемые в рамках именной и глагольной групп, соответственно. Кроме того, структура текста, которая определяется синтаксическими закономерностями языка, служит одним из когнитивных механизмов языка, управляющим в тексте такой важной характеристикой, как активность знаний, естественным образом реализуя логическую схему «анализ-действие».

2 Морфология татарского языка

Татарская морфология является регулярной, почти автоматной [6-8] и в то же время имеет естественную сложность, которая заключается в следующем.

- 1) Возможность аффиксального преобразования словоформы с одного типа в другой путём присоединения определённых аффиксальных морфем справа (праволинейная грамматика). Присоединяясь к словоформе, специальный аффикс-переключатель меняет («переключает») слово с одного типа на другой, например, превращает именную словоформу в глагольную или в форму прилагательного, и наоборот (*уй-уй+ла*: мысль-думай; *кара-кара+у-карау+чы*: наблюдай-наблюдение-наблюдатель). Здесь и далее знак «+» - это условный разделитель слова и аффикса-переключателя. В тексте слитное написание: *уйла* - думай.
- 2) Морфологическое (синтетическое) задание признаков модальности, настроения, эмоционально-личностного отношения к ситуации, объекту или процессу, описываемых данной

словоформой (*кояш-кояш+тай*: солнце-словно солнце; *алды-алды+мыни*: взял-неужели взял).

- 3) Контекстное разнообразие значений аффикса. В татарском языке порядка 90 словоизменительных аффиксальных морфем. Большинство из них имеют несколько значений, т.е. используются для кодирования в тексте нескольких, иногда до 20 и более, различных контекстных значений (смыслов). Например, аффикс *-дан/-тан*, обозначаемый как аффикс исходного падежа: *бакча+дан чыга* (выходит из сада), используется также для кодирования других контекстных смыслов: *корыч+тан каты* (крепче стали: сравнение), *кояш+тан аста* (ниже солнца: точка отсчёта), *укытучы+дан сора* (спроси у учителя: адресат) и т.п.

Именная группа, как правило, кодирует некую семантическую субъектно-объектную ситуацию, а глагольная группа – контекстные отношения над этой ситуацией. Возможность перехода от 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 кодирование в рамках одной словоформы некоторого значения, описываемого на флективно-аналитических языках (например, на английском) несколькими словосочетаниями или предложениями.

Сложность татарской морфологии конструктивная, т.к. аффиксальная морфема однозначно выделяется в последовательности аффиксов в словоформе, и её значение практически однозначно определяется в лексико-грамматическом контексте. Морфология татарского языка в большой степени регулярна, близка к автоматной, с небольшим количеством исключений из правил, не меняющих смысла, что обеспечивает минимизацию ёмкостных и временных функций при обработке текстов на татарском языке, упрощает анализ структуры и значения словоформы.

Важное свойство татарской морфологии - это фиксированная позиция, «жёсткое» размещение аффиксов в последовательности аффиксальных морфем. Это означает, что одна и та же схема сочетания морфем (морфотактика) присуща почти всем именным и глагольным группам. Это даёт возможность по одной и той же схеме (автоматно) образовывать словоформы с одинаковыми глубинными значениями аффиксов. Например: *кул, куллар, кулларым, кулларыма* – (рука, руки, мои руки, моим рукам); *бакча, бакчалар, бакчаларым, бакчаларыма* – (сад, сады, мои сады, моим садам).

Разные именные корневые морфемы *кул* (рука), *бакча* (сад) имеют одни и те же последовательности аффиксальных морфем с идентичными значениями. Приведённые парадигмы описываются следующими схемами: $X(\text{Имя сущ.})$, $X(\text{Имя сущ.})+лар(\text{афф.мн.})$, $X(\text{Имя сущ.})+лар(\text{афф.мн.})+ым$ (афф. притяж., 1-е л., ед.ч.), $X(\text{Имя сущ.})+лар(\text{афф.мн.})+ым$ (афф. притяж., 1-е л., ед.ч.) + *a* (афф. падежн., дат. падеж).

Позиции аффиксальных морфем, составляющих словоформу, неизменны относительно друг друга. Таким образом, можно определить следующие когнитивные механизмы татарского языка: регулярность и естественная сложность морфологии; фиксированность позиций соответствующих типов аффиксальных морфем. Эти механизмы универсальны, т.е. работают на всех словах, входящих в соответствующие классы лексем (части речи).

3 Морфологический эллипсис

Следующая возможность в татарской морфологии, которая может быть отнесена к когнитивному механизму (компактность), называется морфологическим эллипсисом [9] - это

возможность пропуска последовательности аффиксов при однородных именных словоформах с сохранением её в последней словоформе. В понятии морфологической эллипсис последовательность пропускаемых аффиксов может быть сколь угодно длинной (теоретически не ограничена). Т.е. возможен вывод любой последовательности аффиксов, общих для однородных членов, вправо за последовательность однородных членов и присоединение их к последнему справа однородному члену. Например:

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

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

4 Морфологическая рекурсия

Одним из важных и мощных когнитивных механизмов в татарском языке является механизм рекурсии [10] - это возможность циклического порождения нового значения путём последовательного применения одной и той же «формулы», т.е. повторного присоединения одного и того же аффикса. Такими свойствами обладают аффиксальные морфемы *-ДАГЫ*¹ (локатив2) и *-НЫКЫ* (притяжательность), которые можно назвать также аффиксами неопределённости, т.е. аффиксами, придающими неопределённость множеству сущностей, связанных с лексемой, к которой присоединяется аффикс рекурсии.

Например, пусть задана лексема *ком* (песок). Присоединение аффикса *-дагы* порождает множество новых сущностей. Например: *комдагы* {таш-камень, эз-след, агач-дерево, кеше-человек, кош-птица, баз-яма, ... все что угодно, любой возможный объект на песке}. Присоединение аффикса *-НЫКЫ* вносит определённость в предыдущую лексему, “заставляя” словоформу-лексему *-комдагы*-на песке заполнить постпозицию одной из сущностей (путём выбора, контекстным или явным, одного объекта из множества). Например, эта сущность - *человек*. Одновременно присоединение рекурсивного аффикса *-НЫКЫ* порождает множество новых объектов, принадлежащих человеку или свойств, присущих человеку, образуя неопределённость, которая заполняется контекстно или явно при присоединении следующего рекурсивного аффикса. Для данного случая это: *ком-ДАГЫ* (на песке) (*кеше*-человек)-*НЫКЫ* (голова, рука, шаг, взгляд, часы и т.п., принадлежащие или присущие этому человеку, который на/в песке).

По такой формуле возможно образование словоформы практически неограниченной длины. Такие длинные последовательности морфем в нормальной речи (при коммуникации между людьми, в текстах) практически не используются. Это связано с проблемами глубины памяти человека, удобства общения между людьми. Подобное словоизменение является совершенно корректным для грамматики татарского языка, и словоформа, образованная присоединением любой последовательности, гипотетически всегда имеет смысл. Конкретное значение приобретается при «погружении» словоформы в определённый контекст.

Пример со следующей словоформой: *бакчаныкындагыныкыныкындагы*, которая однозначно раскладывается на составляющие: *бакча-сад* (имя сущ.) + *ныкы* (притяж.) + *ндагы* (локатив2) + *ныкы* (притяж.) + *ныкы* (притяж.) + *ндагы* (локатив2).

Значение данной словоформы таково: «*нечто/некто, которое находится на/в нечто/некто, которое само принадлежит нечто/некто, которое само находится на/в нечто/некто, которое принадлежит саду*».

¹ Здесь и далее заглавные буквы в аффиксальных морфемах обозначают вариативность символов в соответствующих позициях согласно закону сингармонизма.

Эксплицитно задавая параметры после каждой морфемы, можно получить контекстную определённость словоформы. В реальных случаях такие параметры задаются имплицитно, наполняясь конкретным значением в зависимости от контекста речи.

Пример. Пусть после каждого аффикса неопределённости стоят параметры: $бакча+ныкы(x0)+ндагы(x1)+ныкы(x2)+ныкы(x3)+ндагы(x4)+ныкы(x5)$, где x_i – контекстные объекты, т.е. объекты, которые либо приобретают конкретное значение из контекста, либо их задаёт пользователь ($i = 1, \dots, 4$). Тогда, придавая значения параметрам: $x0 = \text{«яблоны»}$, $x1 = \text{«ветка»}$, $x2 = \text{«птичка»}$, $x3 = \text{«крыло»}$, $x4 = \text{«перо»}$, можно получить следующее контекстное значение: *«нечто (значение $x5$, придаваемое параметру последним аффиксом, осталось неопределённым) на пере, что принадлежит крылу, что принадлежит птичке, что на ветке, что принадлежит яблоне»*. Пример показывает, что применение рекурсивных аффиксов приводит к сжатию информации и существенной экономии памяти.

5 Когнитивные механизмы при описании действий

Поверхностное лексическое описание предикатов (действий, отношений), как правило, осуществляется глагольными группами [11, 12]. Свойства регулярности и конструктивной сложности морфологии 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, можно рассмотреть следующие естественные когнитивные механизмы, проявляющиеся в глагольных словоформах.

5.1 Возможность рекурсивно задавать нечёткие команды и описывать нечёткие действия

Данное свойство кодируется глагольными аффиксами – *Гала и -итыр*, занимающими позицию залога, т.е. сразу же после глагольной основы. Пример с глаголом *кара* (смотри, 2-е лицо, ед.ч., повел. накл.).

караитыр (смотри/поглядывай время от времени); *кара* (смотри)+*итыр* (время от времени); *караитыргала* (смотри время от времени, время от времени – т.е. реже); *кара* (смотри)+*итыр* (время от времени)+*гала* (время от времени); *караитыргалаитыргалаитыр...* (смотри время от времени, время от времени, время от времени – и ещё реже...); *кара* (смотри, корень, 2-е лицо, ед.ч., повел.накл.)+*итыр* (время от времени-изредка)+*гала* (время от времени - ещё реже)+*итыр* (ещё реже)+*гала*(ещё реже)+*итыр* (ещё реже)...

Насколько редко требуется *смотреть/поглядывать* - определяется из контекста ситуации.

5.2 Возможность одной словоформой задавать объектно-предикативные сценарии, ролевые ситуации

Реализация данного свойства обеспечивается рядом специальных глагольных аффиксов, занимающих также залоговую позицию: *-н, -Бли, -т, -Дыр, -Тыр*.

Изменения ролевой ситуации при присоединении соответствующих аффиксов можно рассмотреть на примере с глагольной словоформой *кара* (смотри).

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

Для словоформы *кара* (смотри) ролевая ситуация следующая: S воздействие на O_k .

Присоединение аффиксов *-н, -Бли, -т, -Дыр* приводит к изменениям.

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

Ролевая ситуация: S воздействие S .

2) -*Би*: *караш* – *кара+ш* (помогай смотреть-ухаживать/ смотри-ухаживай 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 определённые аффиксы, можно создавать всё новые и новые ролевые ситуации и описывать процессы на лексическом уровне. Например, добавление аффикса -*Бл* 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 , т.е. воздействие на самого же инициатора-субъекта действия (бросить) через 4 промежуточных объекта-агента (или группы объектов-агентов).

6 Активность знаний

Известно, что английские предложения строятся по схеме *S-V-O* (*subject-verb-object: субъект-глагол-объект*) [13-15], а татарские - по схеме: *S-O-V*. Т.е. англичанин, если говорит, например, о намерении сходить в кино, сначала скажет, пойдёт или не пойдёт, 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*”). Здесь действие управляет ситуацией. После того, как высказано однозначно намерение субъекта *We’ll go* – *мы пойдём*, дальнейшая информация становится пассивной, практически не влияет на выбор способа действия или усложняет его. В тексте на татарском языке сначала даётся информация и её анализ, и только после этого определяется действие. («*Без дуслар белән иртәгә буласы “Аттила” киносына барабыз/бармыйбыз*» – Мы с друзьями на завтрашний фильм «Аттила» пойдём/не пойдём).

В ИС 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 структуре английского языка. Т.е. в тюркских языках сначала анализируется, обрабатывается информация, а затем осуществляется действие: подбирается соответствующая адекватная модель представления знаний или выбираются соответствующие алгоритмы и схемы реализации, эффективность которых во многом определяется корректностью и полнотой анализа информации. Это можно назвать событийным программированием.

Такие возможности татарского языка позволяют ставить задачу разработки программ накопления и извлечения знаний в глобальных компьютерных сетях.

Заключение

Как показывают исследования [6, 16-18], тюркские языки, и в частности татарский язык, как агглютинативный язык, обладающий регулярной морфологией и контекстно разрешаемой конструктивной сложностью, являются эффективным инструментом для создания ИС обработки информации. В силу малых значений показателей временных и ёмкостных оценочных функций для генерации и анализа татарских словоформ (за счёт регулярности) достигается эффективность при накоплении и обработке информации на татарском языке. Компактность передачи смысла текста на поверхностном, лексическом уровне объясняется возможностями языка синтетически, т.е. словоформой, кодировать смысл, который для других языков (английский, русский) формируется аналитически (несколькими предложениями). Агглютинативность языка, алгоритмические закономерности, минимальность исключений, жёсткость синтаксиса позволяют ставить задачу о построении языка промежуточной трансляции, т.е. языка-посредника на базе татарского языка.

Исследование когнитивных лексико-грамматических структур описания, запоминания, хранения, обработки и передачи информации ЕЯ способствует созданию ИС и технологий обработки информации. Представляют интерес ЕЯ агглютинативного типа, обладающие естественными когнитивными технологическими механизмами, которые позволяют на поверхностном уровне эксплицитно вербализовать, фиксировать и отображать стереотипы, когнитивные модели и механизмы, описывающие ситуации и процессы в модели мира [17, 18]. На примере татарского языка показано, что ЕЯ агглютинативного типа обладают когнитивным потенциалом, позволяющим описывать сложные информационные системы обработки знаний. Дальнейшее развитие и применение на практике предлагаемого подхода и принципов исследования лексико-грамматического потенциала ЕЯ позволит создать новые технологии и когнитивные модели обработки знаний и принятия решений в ИС.

Список источников

- [1] *Сулейманов Д.Ш.* К вопросу исследования технологического аспекта естественных языков // Обработка текста и когнитивные технологии: Труды XI Междунар. науч. конф. (7–14 сентября 2009, Констанца). Казань: Изд-во Казанского университета., 2010. С.232-245.
- [2] *Chomsky N.* Syntactic Structures. The Hague: Mouton, 1957. 117 p.
- [3] *Поспелов Д.А., Осипов Г.С.* Прикладная семиотика. <http://raii.org/library/ainews/1999/1/ OSPOS.ZIP>.
- [4] *Pospelov D.* Situational Control: Theory and Practice. Columbus OH: Batelle Memorial Institute, 1986.
- [5] *Поляков В.Н.* Проблемы представления, приобретения и использования знаний в свете обработки естественного языка // Когнитивно-семиотические аспекты моделирования в гуманитарной сфере. Казань: Изд-во Академии наук РТ, 2017. С.145-163.
- [6] *Suleymanov D.S.* Natural Cognitive Mechanisms in the Tatar language // 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.
- [7] *Гузев В.Г.* О некоторых экзотических особенностях тюркских языков («тюркские чудеса») // Актуальные проблемы мировой политики. СПб.: Изд-во С.-Петербург. ун-та, 2020. Вып.10. С.231-245.
- [8] *Щербак А.М.* О морфологической структуре слов в тюркских языках // Морфологическая структура слова в языках различных типов. 1963. С.267-270.
- [9] *Лекант П.А.* Эллипсис как проблема синтаксиса и фразеологии // Ученые записки МОПИ им. Н.К. Крупской. 1966. Т.160, Вып.11. С.210-224.
- [10] *Лодатко Е.А.* Рекурсивные лингвистические структуры // Теоретические и прикладные проблемы русской филологии: Научно-метод. сборник. Славянск: СГПУ, 2004. Вып. XII. С.86-95.
- [11] *Fillmore C.J.* Frame Semantics and the Nature of Language // In Annals of the New York Academy of Sciences: Conference on the Origin and Development of Language and Speech. 1976, Vol. 280. P.20-32.
- [12] *Осипов Г.С., Смирнов И.В., Тихомиров И.А.* Реляционно-ситуационный метод поиска и анализа текстов и его приложения // Искусственный интеллект и принятие решений. 2008. № 2. С.3-10.
- [13] *Givón T.* Syntax: An introduction. John Benjamins Publishing Co., 2001. 500 p.

- [14] **Тестелец Я.Г.** Введение в общий синтаксис. М.: Издательство РГГУ, 2001. 805 с.
- [15] **Гринберг Дж.** Некоторые грамматические универсалии, преимущественно касающиеся порядка значимых элементов // Новое в лингвистике. Языковые универсалии. 1970. Вып. 5. С. 114-162.
- [16] **Gatiatullin A., Suleymanov D., Prokopyev N., Khakimov B.** About Turkic Morpheme Portal // In Proceedings of the Computational Models in Language and Speech Workshop (CMLS 2020). Kazan, November 12-13, 2020. P.226-243.
- [17] **Oflazer K., Sarachlar M.** Turkish Natural Language Processing // Theory and Applications of Natural Language Processing. Series ed. J. Hirshberg, E. Hovy, M. Jhonson. Springer International Publishing AG, part of Springer Nature 2018. 357 p.
- [18] **Tantug A.G., Adali E.** Machine Translation Between Turkic Languages // Turkish Natural Language Processing. Theory and Applications of Natural Language Processing. Series ed. J. Hirshberg, E. Hovy, M. Jhonson. Springer International Publishing AG, part of Springer Nature 2018. P.237-255.
-

Сведения об авторах



Судейманов Джавдет Шевкетович, 1955 г. рождения. Окончил механико-математический факультет Казанского государственного университета в 1977 г., к.т.н. (1985), д.т.н. (2000). Научный руководитель Института прикладной семиотики Академии наук Республики Татарстан (АН РТ), академик АН РТ, профессор. Заслуженный деятель науки РТ, член Российской ассоциации искусственного интеллекта (РАИИ). В списке научных трудов более 300 работ в области прикладной семиотики, компьютерной и когнитивной лингвистики, искусственного интеллекта, электронной и социальной педагогики. Author ID (РИНЦ): 9142;

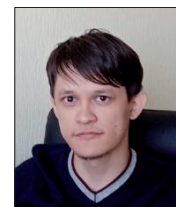
Author ID (Scopus): 6603474810; Researcher ID (WoS): B-4793-2014. dvdt.slt@gmail.com

Гильмуллин Ринат Абрекович, 1975 г. рождения. Окончил Казанский государственный университет в 1997 г., к.ф.-м.н. (2010). Директор Института прикладной семиотики АН РТ. В списке научных трудов около 50 работ. ORCID: 0000-0002-8520-8921; Author ID (РИНЦ): 179405; Author ID (Scopus): 57079235700. rinatgilmullin@gmail.com ✉



Гатиатуллин Айрат Рафизович, 1972 г. рождения. Окончил Казанский государственный университет в 1994 г., к.т.н. (2002). Ведущий научный сотрудник Института прикладной семиотики АН РТ. В списке научных трудов более 60 работ. ORCID: 0000-0003-3063-8147; Author ID (РИНЦ): 161758; Author ID (Scopus): 56500678000. ayrat.gatiatullin@gmail.com

Прокопьев Николай Аркадиевич, 1992 г. рождения. Окончил Институт вычислительной математики и информационных технологий Казанского Федерального университета в 2015 году. Научный сотрудник Института прикладной семиотики АН РТ. В списке научных трудов около 40 работ. ORCID: 0000-0003-0066-7465; Author ID (РИНЦ): 999214; Author ID (Scopus): 57190803409; Researcher ID (WoS): S-3829-2016. nikolai.prokopyev@gmail.com



Поступила в редакцию 13.11.2023, после рецензирования 23.11.2023. Принята к публикации 1.12.2023.



Cognitive potential of agglutinative languages in intelligent technologies

© 2023, D.S. Suleymanov, R.A. Gilmullin ✉, A.R. Gatiatullin, N.A. Prokopyev

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

Abstract

This article describes an approach to development of new intelligent technologies based on the study of cognitive potential and semiotic modeling of lexical and grammatical structures of agglutinative languages (using the Tatar language as an example). A classification of cognitive lexical grammatical structures that determine the information technological effectiveness of a language is proposed. Specific examples show the relevant capabilities of the Tatar language for compact representation of knowledge, operating with fuzzy information, revealing such important properties as regularity, natural and constructive complexity, recursion, deep morphological ellipsis, functional diversity and semantic multivalence of affixes. The proposed approach, based on semiotic models of agglutinative language and the use of technological means of verbalization and contextual recognition of language meanings, can provide effective solutions for creating intelligent tools for accumulating and processing knowledge. Work in this direction creates the prospect of building a cognitive-communicative artificial intelligence system based on a universal language of communication between artificial intelligence systems as well as with humans.

Keywords: cognitive modeling, Tatar language, lexical grammatical structures, morphology, recursion, fuzzy commands, knowledge activity, intelligent systems.

For citation: Suleymanov DS, Gilmullin RA, Gatiatullin AR, Prokopyev NA. Cognitive potential of agglutinative languages in intelligent technologies [In Russian]. *Ontology of designing*. 2023; 13(4): 496-506. DOI:10.18287/2223-9537-2023-13-4-496-506.

Funding: This work was supported by the Scientific Research Foundation (project 23-21-10083).

Conflict of interest: The authors declare no conflict of interest.

References

- [1] **Suleimanov DS.** On the question of researching the technological aspect of natural languages [In Russian]. In: Text processing and cognitive technologies: Proceedings of the XI Intern. scientific conf. (Constanța, 7–14 September 2009). Kazan: Kazan Publishing House of the State Univ., 2010: 232–245.
- [2] **Chomsky N.** Syntactic Structures. The Hague: Mouton, 1957. 117 p.
- [3] **Pospelov DA, Osipov GS.** Applied semiotics [In Russian]. <http://raii.org/library/ainews/1999/1/OSPOS.ZIP>.
- [4] **Pospelov DA.** Situational Control: Theory and Practice. Batelle Memorial Institute, Columbus, OH, 1986.
- [5] **Polyakov V.** Problems of representation, acquisition and use of knowledge in natural language processing. In: Cognitive-semiotic aspects of modeling in the humanities. Publishing House of the Academy of Sciences of the Republic of Tatarstan, Kazan, 2017: 145-163.
- [6] **Suleymanov DS.** Natural Cognitive Mechanisms in the Tatar language. In: 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: 210-213.
- [7] **Guzev VG.** About some exotic features of Turkic languages («Turkic miracles») [In Russian]. Digest of World Politics. Annual Review. St Petersburg State University, 2020; 10: 231-245.
- [8] **Scherbak AM.** On the morphological structure of words in Turkic languages [In Russian]. Morphological structure of words in languages of different types. 1963: 267-270.
- [9] **Lekant PA.** Ellipsis as a problem of syntax and phraseology [In Russian]. Scientific notes of MOPI named after N.K. Krupskaya. M., 1966; 160(11): 210-224.
- [10] **Lodatkо EA.** Recursive linguistic structures [In Russian]. Theoretical and applied problems of Russian philology: Scientific and methodological collection. Slavyansk: SGPU, 2004; 12: 86-95.
- [11] **Fillmore CJ.** Frame Semantics and the Nature of Language. In: Annals of the New York Academy of Sciences:

- Conference on the Origin and Development of Language and Speech. 1976; 280: 20-32.
- [12] **Osipov GS, Smirnov IV, Tikhomirov IA.** Relational-situational method of searching and analyzing texts and its applications [In Russian]. Artificial Intelligence and Decision Making. 2008; 2: 3-10.
- [13] **Givon T.** Syntax: An introduction. John Benjamins Publishing Co., 2001. 500 p.
- [14] **Testelefs YG.** Introduction to general syntax [In Russian]. M.: RGSU, 2001.
- [15] **Grinberg J.** Some Grammar Universals, primarily concerning the order of Significant Elements [In Russian]. New in linguistics. Language universals. M.: Progress, 1970; 5: 114-162.
- [16] **Gatiatullin A, Suleymanov D, Prokopyev N, Khakimov B.** About Turkic Morpheme Portal. In: Proceedings of the Computational Models in Language and Speech Workshop (CMLS 2020). Kazan, Russian, November 12-13, 2020: 226-243.
- [17] **Oflazer K, Sarachlar M.** Turkish Natural Language Processing Theory and Applications of Natural Language Processing. Series ed. J. Hirshberg, E. Hovy, M. Jhonson. Springer International Publishing AG, part of Springer Nature 2018. 357 p.
- [18] **Tantug AG, Adali E.** Machine Translation between Turkic Languages. Turkish Natural Language Processing. Theory and Applications of Natural Language Processing. Series ed. J. Hirshberg, E. Hovy, M. Jhonson. Springer International Publishing AG, part of Springer Nature 2018: 237-255.
-

About the authors

Dzhavdet Shevketovich Suleymanov (b. 1955) graduated from the Faculty of Mechanics and Mathematics of Kazan State University in 1977, PhD (1985), Doctor of Technical Sciences (2000). Scientific director of the Institute of Applied Semiotics of the Tatarstan Academy of Sciences, academician of the Tatarstan Academy of Sciences, professor. Honored Scientist of the Republic of Tatarstan, member of the Russian Association of Artificial Intelligence (RAAI). The list of scientific works includes more than 300 works in the field of applied semiotics, computer and cognitive linguistics, artificial intelligence, electronic and social pedagogy. Author ID (RSCI): 9142; Author ID (Scopus): 6603474810; Researcher ID (WoS): B-4793-2014. dvdt.slt@gmail.com.

Rinat Abrekovich Gilmullin (b. 1975) graduated from Kazan State University in 1997, PhD (2010). Director of the Institute of Applied Semiotics of the Tatarstan Academy of Sciences. The list of scientific works includes about 50 works. ORCID: 0000-0002-8520-8921; Author ID (RSCI): 179405; Author ID (Scopus): 57079235700. rinatgilmullin@gmail.com ✉.

Ayrat Rafizovich Gatiatullin (b. 1972) graduated from Kazan State University in 1994, PhD (2002). Leading researcher at the Institute of Applied Semiotics of the Tatarstan Academy of Sciences. The list of scientific works includes more than 60 works. ORCID: 0000-0003-3063-8147; Author ID (RSCI): 161758; Author ID (Scopus): 56500678000. ayrat.gatiatullin@gmail.com.

Nikolai Arkadievich Prokopyev (b. 1992) graduated from the Institute of Computational Mathematics and Information Technologies of Kazan Federal University in 2015. Researcher at the Institute of Applied Semiotics of the Tatarstan Academy of Sciences. The list of scientific works includes about 40 works. ORCID: 0000-0003-0066-7465; Author ID (RSCI): 999214; Author ID (Scopus): 57190803409; Researcher ID (WoS): S-3829-2016. nikolai.prokopyev@gmail.com.

Received November 13, 2023, Revised November 23, 2023. Accepted December 01, 2023.
