

ПРИКЛАДНАЯ ЛИНГВИСТИКА

УДК 81'322.2

**МЕТАЯЗЫК ОПИСАНИЯ СТРУКТУРЫ
ТАТАРСКОЙ СЛОВОФОРМЫ ДЛЯ КОРПУСНОЙ
ГРАММАТИЧЕСКОЙ АННОТАЦИИ***А.М. Галиева, Б.Э. Хакимов, А.Р. Гатиатуллин***Аннотация**

В статье рассматриваются вопросы представления грамматической информации в электронном корпусе татарского языка, описываются основные проблемы, связанные с созданием системы грамматической аннотации, и предлагаются пути решения этих проблем. Установлено, что источником различного рода затруднений при создании метаязыка для описания структуры татарской словоформы являются отсутствие стандартов по разработке корпусных аннотаций и универсальной терминологии, многозначность и омонимия аффиксов, требование совместимости с другими электронными лингвистическими ресурсами и др.

Ключевые слова: лингвистический корпус, корпусная аннотация, татарский язык, грамматические категории, аффикс, словоформа.

Введение

В настоящее время активно разрабатываются проекты по созданию общедоступных электронных корпусов для разных (в том числе тюркских) языков. В этом направлении в научно-исследовательском институте «Прикладная семиотика» Академии наук Республики Татарстан совместно с Казанским федеральным университетом продолжается работа по созданию корпуса татарского языка «Туган тел» (ТТ). Данный корпус разрабатывается на технологической платформе Восточно-армянского национального корпуса (EANC) как моноязычный корпус с грамматической и семантической аннотацией. Татарский национальный корпус содержит тексты различных стилей и жанров современного татарского литературного языка.

Будучи исследовательским лингвистическим ресурсом, ТТ представляет собой совокупность концептуально-функциональных моделей различных уровней татарского языка [1]. Важнейшим компонентом корпуса является лингвистическая разметка (аннотация), позволяющая хранить информацию о текстах и представленных в них лингвистических объектах в формализованном виде и автоматически обрабатывать эту информацию при помощи поисковых запросов. На данном этапе тексты в ТТ снабжены грамматической аннотацией словоформ,

реализуемой при помощи специально разработанной системы условных обозначений (тегов).

В настоящей статье рассматриваются теоретические и практические вопросы создания системы корпусной грамматической аннотации для татарского языка, отвечающей требованиям теоретической обоснованности и прагматической ориентированности; показываются основные сложности, с которыми столкнулись разработчики ТТ, и предлагаются пути решения этих проблем.

1. Особенности грамматической аннотации в корпусе татарского языка

Представление в ТТ информации о грамматических категориях, формах и значениях татарского языка является отдельной научной проблемой. Тексты должны нести максимум лингвистической информации, «в том числе обязательно поморфемное глоссирование» [2]. Открытость корпусных данных для разных языков предполагает, что: а) «аннотация должна быть в высокой степени стандартизована для того, чтобы облегчить поиск сходных явлений в разных языках, описанных разными исследователями»; б) «пользователь должен иметь возможность выбрать для отображения только интересующие его компоненты (слои) информации» [2].

Как показывает опыт разработчиков ТТ, создание системы условных корпусных обозначений выходит за рамки чисто прикладной задачи, неизбежно вынуждая принимать решения по многочисленным теоретическим проблемам, для которых в течение долгих лет параллельно существуют альтернативные трактовки в зависимости от целей и аспектов исследования.

В лингвистике возможны два основных подхода к описанию грамматических категорий – от семантики к форме и от формы к семантике. В разработанной нами грамматической модели татарского языка осуществляется движение от аффиксов, которые выделяет морфологический анализатор [1, 3], к определению формы слова. При этом не учитываются коммуникативные аспекты, семантические параметры лексем и даже синтаксические факторы, которые чрезвычайно важны при интерпретации грамматических явлений человеком. Совмещение исходно-формальных данных, семантики, контекстуальной обусловленности языковых единиц и прагматических установок говорящего при описании языкового материала в данном случае входит в круг задач пользователя, интерпретирующего корпусные данные.

Движение от формы к значению было доминирующим в традиционной морфологии, основной объект которой был ограничен формами слов (то есть описанием того, из каких значимых компонентов могут состоять словоформы данного языка и по каким правилам происходит соединение морфем в составе слова). Несмотря на то что такой подход имеет совершенно очевидные недостатки и в настоящее время выглядит в некоторой степени как анахронизм, если реализуется человеком, тем не менее опора на форму «даёт возможность выявить и охарактеризовать значения, выражаемые именно в этом языке, во всей их сложности» [4, с. 64].

В корпусе татарского языка грамматическое значение определяется исключительно по типизированным формальным средствам – аффиксам, характерным

для всего грамматического класса словоформ (например, по аффиксам падежей и принадлежности у существительных, аффиксу сравнительной степени у прилагательных). Таким образом, принципиальными для данного ресурса являются системно-языковые аспекты.

Система грамматической аннотации ТТ ориентирована на представление всех реально существующих грамматических форм слов, которые не всегда отражаются в описательных исследованиях по татарской грамматике или имеют различные альтернативные трактовки. Для формального представления татарской агглютинативной морфологии используется модель, в которой словоформа строится на базе последовательного присоединения к основе регулярных словообразовательных и словоизменяющих аффиксов. Таким образом, для разметки словоформы необходимо проанализировать структуру её аффиксальной цепочки [3]. На данном этапе морфологический анализатор выделяет только те грамматические категории, которые выражаются при помощи аффиксов, то есть объектом анализа являются только синтетические формы слова.

С точки зрения удобства пользователя и решаемых исследовательских задач грамматическая аннотация в корпусе татарского языка должна соответствовать следующим требованиям: 1) простота; 2) релевантность грамматической и семантической системе татарского языка; 3) понятность для пользователя – средне-статистического лингвиста-тюрколога; 4) универсальность (понятность для лингвистов, не являющихся специалистами в области татарского языка, например типологов); 5) совместимость с размеченными корпусами и лексикографическими базами данных для других языков. Одновременное удовлетворение всех этих требований представляет довольно сложную задачу, так как данные установки зачастую взаимно противоречивы. В частности, попытки ориентации как на тюркологов – носителей языка, так и на специалистов в области общей типологии обнаруживают существенные различия в подходах к обозначению и описанию отдельных категорий. В то же время требование совместимости с аналогичными ресурсами для других языков предполагает применение общих стандартов.

При разработке системы корпусных обозначений для грамматических категорий татарского языка нами были изучены системы обозначений в словарях разного типа и грамматиках тюркских языков, система грамматической аннотации в НКРЯ, работы по общей морфологии и другие исследования. Особо следует выделить «Лейпцигские правила глоссирования» [5], которые были разработаны в отделе лингвистики Института эволюционной антропологии имени Макса Планка и в отделе лингвистики Лейпцигского университета. Данную систему правил можно считать своеобразным общепризнанным стандартом у лингвистов мира, в первую очередь типологов. Обозначение результатов поморфемного анализа в ТТ приближено к данным правилам. Привлекались и другие источники. В частности, была изучена система категорий базы данных Verbum, которая отражает состав и структуру элементарных глагольных значений, выявляемых путём сопоставления форм глаголов на материале большого количества языков [6].

2. Метаязык корпусной грамматической аннотации

Перечислим основные проблемы, с которыми столкнулись авторы ТТ при разработке системы грамматической аннотации.

Во-первых, это **отсутствие универсального метаязыка и значительная вариативность терминологии**. Несмотря на то что большая часть специалистов убеждена в необходимости пользоваться при письменной фиксации текстов строкой поморфемного глоссирования, единого унифицированного формата представления лингвистических данных в настоящее время не существует. Различия касаются инвентаря грамматических глосс, количества и состава необходимых слоёв репрезентации. «Это связано как с объективными научно-содержательными проблемами (неизоморфность грамматической структуры различных языков, различия в степени прозрачности морфонологических процессов и т. п.), так и с организационными (отсутствие единого координирующего центра или стандарта)» [2]. Упомянутые выше «Лейпцигские правила глоссирования» и универсальная типологическая терминология на практике мало используются современными татарскими лингвистами – носителями языка. В свою очередь, татарская лингвистика, теоретически и методологически опирающаяся на достижения отечественной тюркологической науки, имеет свою исторически сложившуюся уникальную терминологию, которая во многих случаях оказывается более релевантной для отображения грамматической структуры татарского языка.

Как уже упоминалось выше, одной из задач корпуса татарского языка является выработка такого стандарта (инструментария описания), который был бы понятен широкому кругу не только специалистов-тюркологов, но и типологов и который позволил бы максимально полно описать грамматическую систему татарского языка. Подобная стандартизация сильно ограничена объективно существующим разнообразием форм и категорий естественных языков, в которых практически нет совпадающих категорий (если не иметь в виду близкородственные языки). Грамматические категории всегда реализуются в системе языка, и их содержание и форма определяются взаимодействием большого количества факторов, вследствие чего категория оказывается несводимой к аналогичной категории другого языка. Так, совершенно очевидно, что генитив в русском и татарском языках – это совершенно разные грамматические категории имён существительных, системы времён и залогов русского глагола имеют состав и структуру, значительно отличающиеся от состава и структуры аналогичных категорий татарского языка.

В лингвистике имеются различные подходы, специфические «логики описания» грамматических категорий и граммем не только в разных языках, но и внутри одного языка. В ряде случаев татарские грамматисты вообще не дают названия категориям, а используют в монографиях и грамматиках такие формулировки, как «форма на *-асы*», «форма на *-макчы*», «форма на *-ырлык*», «деепричастие на *-ып*» и т. д. [7, 8]. Подобное обозначение некоторых грамматических форм вполне приемлемо для описательных исследований; более того, не всегда формируя полноценные грамматические категории, такие формы совершенно обоснованно могут не иметь утверждённой терминологической единицы. Однако в электронных размеченных лингвистических ресурсах должно быть отражено как можно больше языковых явлений, а для этого данные явления должны

иметь свой уникальный «ярлык», который должен быть не просто условным, но и теоретически обоснованным и по возможности интуитивно понятным.

Тем не менее некоторые шаги для унификации представления лингвистических данных всё же делаются. Одно из самых востребованных направлений – это стандартизация морфологического глоссирования. В разрабатываемом корпусе татарского языка в качестве одного из базовых нормативов были приняты «Лейпцигские правила глоссирования», отсюда заимствуются основные сокращения, принятые для обозначения тех или иных грамматических категорий, а также правила совмещения категорий при описании словоформ.

Во-вторых, затруднение вызывает **многозначность грамматических форм и категорий**. В идеале условное обозначение (тег) грамматической категории должно адекватно отражать её основное значение. Но в языке практически нет однозначных типовых категорий (инвариантов).

Как мы уже отмечали, в ТТ грамматическая категория определяется по аффиксу. Несмотря на то что татарский язык – агглютинативный (а для агглютинативных языков характерна тенденция к однозначности аффиксов), значительная часть татарских аффиксов всё же многозначны. Так, к основным значениям *направительного* падежа (DIR в нашей системе разметки) можно отнести значения места, времени, объекта, цели; к значениям *исходного* падежа (ABL в нашей системе) – значения пространства, времени, объекта, причины, меры и степени действия [8, с. 53–55].

О «несводимости граммемы к называющей её этикетке» говорит В.А. Плунгян: «Неудовлетворённость, которую многие современные лингвисты ощущают от концепции инварианта, состоит не в том, что инвариант вводится в описание, а в том, что он остаётся единственной реальностью описания, то есть имя граммемы начинает подменять саму граммему во всех ситуациях, а не только в тех, ради которых имя, собственно, и было придумано» [9, с. 230]. Это замечание для корпуса тем более актуально, что результатом поискового запроса будет не совокупность примеров, отфильтрованных лингвистом согласно критерию соответствия заранее данным теоретическим предпосылкам (так обычно подбирается материал для учебных пособий), а все фрагменты текстов, автоматически выбранных по чисто формальным показателям, с заранее заданной разметкой – грамматической или семантической.

Поскольку, как отмечает В.А. Плунгян, «в семантическом отношении граммема есть некий аморфный конгломерат слабо связанных друг с другом семантических элементов» [9, с. 231], следует ориентироваться на ядерные смыслы, наиболее значимые для современного носителя.

В ряде случаев многозначность граммем отражается в названиях грамматических категорий татарского языка, ср.: *местно-временной падеж* (*урын-вакыт килеше*), выражающий обстоятельственные значения места и времени; *взаимно-совместный залог* (*уртаклык юнәлеше*), выражающий значения ассоциатива, ассистива и реципрока. Требование удобства системы аннотации исключает такие громоздкие формулировки, как *местно-временной падеж*, поэтому приходится фиксировать в помете ядерное значение, которое не всегда легко выделить (например, *локатив* (LOC) для местно-временного падежа).

В-третьих, разработчики ТТ столкнулись с такой проблемой, как **синкретичность (неатомарность) грамматических значений, отсутствие во многих случаях «чистых» категорий**. Так, синтетические формы прошедшего времени в татарском языке связаны с категорией эвиденциальности (субъектной модальности), формы будущего – с модальными значениями уверенности/неуверенности говорящего и обязательности/потенциальности действия. Аффикс *ДЫ*¹ выражает и прошедшее время, и эвиденциальность, аффикс *Ыр* имеет одновременно значение будущего времени и потенциальности. При этом в языке при концептуализации времени в системе грамматических категорий представлены разные типы модальности, разная «логика» моделирования времени: если прошедшее время выражает субъектную модальность, то будущее – оттенок предположительности, когда «говорящий оценивает ситуацию на фоне своих ожиданий относительно её наступления/ненаступления» (Evaluation/Expectation по терминологии в базе данных Verbum) [6].

Одним из возможных путей решения этой проблемы является вариант истолкования категории времени и категории наклонения как одной грамматической категории. Так, А.П. Володин и В.С. Храковский, анализируя систему тюркских глагольных форм, приходят к следующему выводу: «Нет двух равноположенных категорий времени и наклонения, а есть одна категория наклонения, в рамках которой временная отнесённость выступает как способ обозначения реальности» [10, с. 50].

Если эмпирические факты свидетельствуют о том, что грамматические значения существуют лишь в «связанном» виде, то возникает вопрос: что именно фиксировать в системе обозначений? При семантическом подходе возможно выявление элементарных значений («семантических атомов» [6]) путём межъязыкового сопоставления. Однако в ТТ отправными точками классификации являются аффиксы, которые объединяют комплексы значений, при этом интерпретация зависит от контекста.

Например, парадигматическим значением прошедшего категорического времени в татарском языке является «выражение очевидного, целостного, однократного действия в прошлом» [8, с. 106], при этом в различных контекстах может идти речь о временной локализации действия, о динамике его развёртывания, законченности/незаконченности, длительности, повторяемости и т. д.

Прошедшее категорическое в живой речи обычно противопоставлено прошедшему результативному не по аспектуальным параметрам, а по вовлечённости субъекта высказывания: оно выражает действие очевидное, несомненное, свидетелем которого являлся сам субъект (субъективная модальность), осознающий это (ср. нормативное *яңгыр яуды* ‘шёл дождь’). При прошедшем неочевидном субъект может наблюдать результат действия в прошлом (*яңгыр яуган* ‘прошёл дождь’), но не само действие (русский перевод не отражает специфику противопоставленных форм, так как в татарском языке субъект, говоря о дожде, одновременно даёт добавочную информацию о себе). В татарском

¹ Здесь и далее группа алломорфов, объединяемых одним грамматическим значением, обозначается следующим образом: чередующиеся графемы обозначены прописными буквами, инвариантные – строчными; представителем группы алломорфов выступает вариант, состоящий из звонких согласных и гласных заднего ряда.

языкознании не выделяется специализированная грамматическая категория эвиденциальности, тем не менее значение очевидности маркируется в связанном виде со значением прошедшего времени и является отличительным признаком прошедшего категорического времени.

Четвёртой проблемой является **омонимичность аффиксов**. Омонимичными по отношению друг к другу могут быть, во-первых, словообразовательные и грамматические аффиксы, а во-вторых – аффиксы, выражающие разные грамматические значения.

Одним из аспектов этой проблемы является отсутствие чётких границ между собственно грамматикой и словообразованием, вынужденность границ между ними. Это касается и таких традиционно «проблемных» для тюркологии языковых явлений, как категории залога или имени действия у глагола, атрибутивные формы существительных (на *лы*, на *сыз*), и отдельных аффиксов, которые в зависимости от контекста выступают то как словообразовательные, то как грамматические. Приведём только один пример с аффиксом *лык*: *экология айлыгы* ‘месячник, посвящённый экологии’ (словообразовательное значение), но *бер айлык эш хагы* ‘заработная плата за один месяц’ (грамматическое значение).

В ряде случаев аффиксы, имеющие одинаковый план выражения, передают разные грамматические значения в зависимости от контекста или от смысла высказывания. В частности, серьёзной проблемой для разработчиков является автоматический анализ значения аффикса *Дыр* в конце слова. Так, форма *укы-тучыдыр* может иметь 2 основных значения: 1) ‘он учитель’ (категория сказуемости); 2) ‘он, вероятно, учитель’ (предположительная модальность). Сложность заключается в том, что такие формы могут использоваться в совпадающих контекстах, и для их правильного распознавания нужно владеть сведениями о стилистических особенностях текста, времени его написания, а также об описываемой в предложении ситуации, фонде знаний говорящего и т. п.

С ситуацией несколько иного плана мы сталкиваемся при классификации глагольных форм на *ГАН*, именуемых иногда в типологических работах «перфектом». При помощи данного аффикса выражаются как значение прошедшего неочевидного (некатегорического, результативного) времени, так и значение причастия прошедшего времени (*үткән заман сыйфат фигыль*). Интерпретация этих значений контекстуально обусловлена и зависит от синтаксической функции словоформы. Помимо этого, объединяясь с некоторыми другими аффиксами, данная форма образует вторичные глагольные формы с разными значениями [8, с. 201–211]. Возможны две стратегии обозначения этой формы в корпусной аннотации: объединение всех случаев под одной пометой либо присвоение отдельных помет каждому значению. Первый подход предполагает поиск адекватного объединяющего термина, покрывающего все случаи, второй требует контекстного различения значений. Одной из перспективных задач разработки ТТ является поиск решений для каждого подобного случая.

В-пятых, перед разработчиками встаёт **вопрос о совместимости с другими размеченными корпусами и электронными базами данных**. Эта проблема особенно актуальна для разного рода параллельных корпусов, но следует сказать, что и многие одноязычные корпусы в процессе применения интегрируются с другими лингвистическими ресурсами. Как мы уже отмечали выше, используемые

пометы должны быть понятными широкому кругу пользователей корпуса и в максимально возможной степени соответствующими системам разметки в других лингвистических проектах. Это предъявляет особые требования к выбору обозначений. Например, несмотря на то, что для взаимно-совместного залога (*уртаклык юнәлеше* в татарских грамматиках) базовым является значение ассоциатива, широкая употребительность термина *реципрок* определила то, что в грамматической разметке корпуса используется именно он (тег RECP).

Наконец, определённую трудность представляет собой **выбор языка описания**. Система условных обозначений корпуса татарского языка в принципе может быть реализована на базе латинской, английской, русской или татарской терминологии. Как мы уже отмечали, для разных категорий пользователей предпочтительными являются несовпадающие терминологические инструменты и вследствие этого разные графические системы и алфавиты. В некоторых случаях в рамках одного ресурса могут одновременно использоваться элементы нескольких систем.

Например, в основу метаязыка грамматических помет в НКРЯ положена система сокращённых помет на основе латинского алфавита, но следует оговорить, что большая часть помет представлена без явных примет языка описания. При семантической аннотации в НКРЯ в качестве базового используется английский язык (*t:tool* – инструменты и приспособления, *t:stuff* – вещества и материалы), тем не менее в ряде грамматических помет сохраняется латинский диграф *ae* (*praet* – прошедшее время, *praes* – настоящее время), что нехарактерно для новых языков.

По мнению разработчиков ТТ, в современных условиях научной интеграции предпочтительной выглядит международная система кодировки грамматических категорий на базе английского языка. При этом корпусная аннотация может сопровождаться специальными элементами пользовательского интерфейса, где части помет могут соответствовать более привычные термины или даже несколько терминологически альтернативных списков грамматических признаков. В таком случае пользователь имеет возможность выбрать более удобную для себя терминологическую систему вывода лингвистической информации, однако в программном инструментарии ТТ используется унифицированная система аннотации.

Заключение

Как показывает проведённая работа, разработка метаязыка для описания структуры словоформы в корпусе татарского языка не является чисто прикладной задачей, а требует рассмотрения широкого круга теоретических вопросов. В основе концепции корпусной грамматической аннотации для татарского языка лежит представление словоформы в виде цепочки аффиксов, каждый из которых обозначается уникальной пометой.

Продуктивным представляется закрепление в пометах основных, типичных значений, но в ряде случаев выделение таких значений связано со значительными трудностями и требует привлечения большого количества корпусных данных и их тщательной обработки. Из этого следует, что предлагаемая система аннотации в дальнейшем будет корректироваться с учётом статистических данных,

полученных при помощи самого корпуса. Предполагается, что инвентарь условных корпусных обозначений в дальнейшем будет дополняться новыми пометами, в том числе и за счёт создания возможностей для контекстного разграничения омонимичных и многозначных аффиксов.

Проект осуществляется в рамках и при финансовой поддержке Программы фундаментальных исследований РАН «Корпусная лингвистика» 2012–2014 гг. (<http://www.corpling-ran.ru/>).

Summary

A.M. Galieva, B.E. Khakimov, A.R. Gatiatullin. A Metalanguage for Describing the Structure of Tatar Word Forms for Corpus Grammatical Annotations.

The article discusses the issues of presentation of grammatical information in the Electronic Corpus of the Tatar language, describes the main problems occurring when developing the system of grammatical annotation, and proposes solutions to these problems. Creating a system of corpus conventional symbols goes beyond a purely applied problem, inevitably making it necessary to solve numerous theoretical problems, which have for many years had alternative interpretations depending on research goals and aspects. It is determined that various problems arising when creating a metalanguage for the description of the structure of Tatar word forms are due to the lack of standards on the development of corpus annotations and universal terminology, ambiguity and homonymy of affixes, the requirement of compatibility with other electronic linguistic resources, etc.

Keywords: linguistic corpus, corpus annotation, Tatar language, grammatical category, affix, word form.

Источники

- ТТ – Татарский национальный корпус «Туган тел». – URL: <http://web-corpora.net/TatarCorpus/search/>, свободный.
- EANC – Восточно-армянский национальный корпус. – URL: <http://www.eanc.net/>, свободный.
- НКРЯ – Национальный корпус русского языка. – URL: <http://www.ruscorpora.ru>, свободный.

Литература

1. Невзорова О.А., Салимов Ф.И., Хакимов Б.Э., Гатиатуллин А.Р., Гильмуллин Р.А., Галиева А.М., Якубова Д.Д., Аюпов М.М. Семантико-грамматическая аннотация в русско-татарской лексикографической базе данных // Филологические науки. Вопросы теории и практики. – Тамбов: Грамота, 2012. – № 7–1 (18). – С. 141–146.
2. Кибрик А.Е., Архипов А.В., Даниэль М.А., Кодзасов С.В., Майерс Т., Нахимовский А.Д. Технологии обработки языковых данных в документировании малых языков. – URL: <http://www.dialog-21.ru/digests/dialog2007/materials/html/35.htm>, свободный.
3. Сулейманов Д.Ш., Хакимов Б.Э., Гильмуллин Р.А. Корпус татарского языка: концептуальные и лингвистические аспекты // Вестн. ТГГПУ. – 2011. – № 4 (26). – С. 211–216.
4. Бондарко А.В. Категоризация в системе грамматики. – М.: Языки славян. культур, 2011. – 483 с.

5. The Leipzig Glossing Rules. – URL: <http://www.eva.mpg.de/lingua/resources/glossing-rules.php>, свободный.
6. Плунгян В.А. Классификация элементарных глагольных значений, используемых в БД «Verbum». – URL: <http://www.mccme.ru/ling/verbum.html>, свободный.
7. Татар грамматикасы: 3 т. – М.: ИНСАН, Казан: Фикер, 2002. – Т. 2. – 448 б.
8. Татарская грамматика: в 3 т. – Казань: Тат. кн. изд-во, 1993. – Т. 2: Морфология. – 397 с.
9. Плунгян В.А. Общая морфология: Введение в проблематику. – М.: Едиториал УРСС, 2003. – 384 с.
10. Володин А.П., Храковский В.С. Об основаниях выделения грамматических категорий // Проблемы лингвистической типологии и структуры языка. – Л.: Наука, 1977. – С. 42–54.

Поступила в редакцию
06.06.13

Галиева Альфия Макаримовна – кандидат философских наук, доцент кафедры общей лингвистики и лингвокультурологии, Казанский (Приволжский) федеральный университет; ведущий научный сотрудник, НИИ «Прикладная семиотика» Академии наук Республики Татарстан, г. Казань, Россия.

E-mail: agalieva@yandex.ru

Хакимов Булат Эрнстович – кандидат филологических наук, доцент кафедры математической лингвистики и информационных систем в филологии Казанский (Приволжский) федеральный университет; ведущий научный сотрудник НИИ «Прикладная семиотика» Академии наук Республики Татарстан, г. Казань, Россия.

E-mail: khakeem@yandex.ru

Гатиатуллин Айрат Рафизович – кандидат технических наук, начальник информационно-компьютерного отдела, Академия наук Республики Татарстан; доцент кафедры математической лингвистики и информационных систем в филологии, Казанский (Приволжский) федеральный университет, г. Казань, Россия.

E-mail: agat1972@mail.ru