

Невзорова Ольга Авенировна, Салимов Фарид Ибрагимович, Хакимов Булат Эрнстович,
Гатиатуллин Айрат Рафизович, Гильмуллин Ринат Абрекович, Галиева Альфия Макаримовна,
Якубова Дильяра Джавдетовна, Аюпов Мадехур Масхутович

**СЕМАНТИКО-ГРАММАТИЧЕСКАЯ АННОТАЦИЯ В РУССКО-ТАТАРСКОЙ
ЛЕКСИКОГРАФИЧЕСКОЙ БАЗЕ ДАННЫХ**

В статье описаны модели и методы представления лингвистической информации в разработанной русско-татарской лексикографической базе данных, отличительными чертами которой являются ориентация на компьютерные системы обработки текстов, а также глубокая детализация и взаимная интеграция лингвистических сведений. Созданная лексикографическая база впервые использует семантико-грамматические аннотации татарских лексем, построенные с учетом типологических особенностей тюркских языков, и в перспективе может быть использована в многоязычных компьютерных приложениях.

Адрес статьи: www.gramota.net/materials/2/2012/7-1/37.html

Источник

Филологические науки. Вопросы теории и практики

Тамбов: Грамота, 2012. № 7 (18): в 2-х ч. Ч. I. С. 141-146. ISSN 1997-2911.

Адрес журнала: www.gramota.net/editions/2.html

Содержание данного номера журнала: www.gramota.net/materials/2/2012/7-1/

© Издательство "Грамота"

Информация о возможности публикации статей в журнале размещена на Интернет сайте издательства: www.gramota.net

Вопросы, связанные с публикациями научных материалов, редакция просит направлять на адрес: voprosy_phil@gramota.net

УДК 81'32; 811.512.1

Филологические науки

В статье описаны модели и методы представления лингвистической информации в разработанной русско-татарской лексикографической базе данных, отличительными чертами которой являются ориентация на компьютерные системы обработки текстов, а также глубокая детализация и взаимная интеграция лингвистических сведений. Созданная лексикографическая база впервые использует семантико-грамматические аннотации татарских лексем, построенные с учетом типологических особенностей тюркских языков, и в перспективе может быть использована в многоязычных компьютерных приложениях.

Ключевые слова и фразы: лексикографическая база данных; автоматическая обработка текста; лингвистическая аннотация; грамматика; семантика; поисковый запрос; тюркские языки.

Ольга Авенировна Невзорова, к.т.н., доцент

Фарид Ибрагимович Салимов, к.ф.-м.н., доцент

Булат Эрнстович Хакимов, к. филол. н.

Айрат Рафизович Гатиатуллин, к.т.н.

Ринат Абрекович Гильмуллин, к.ф.-м.н.

Альфия Макаримовна Галиева, к. филос. н., доцент

Диляра Джавдетовна Якубова, к. филол. н.

Мадехур Масхутович Аюпов

Казанский (Приволжский) федеральный университет

Научно-исследовательский институт "Прикладная семиотика" Академии наук Республики Татарстан

onevzoro@gmail.com; Farid.Salimov@ksu.ru; khakeem@yandex.ru; agat1972@mail.ru;

rinatgilmullin@gmail.com; agaliev@yandex.ru; suleymanovad@gmail.com; madehur@mail.ru

СЕМАНТИКО-ГРАММАТИЧЕСКАЯ АННОТАЦИЯ В РУССКО-ТАТАРСКОЙ ЛЕКСИКОГРАФИЧЕСКОЙ БАЗЕ ДАННЫХ[©]

Исследование выполнено при финансовой поддержке РГНФ в рамках научно-исследовательского проекта РГНФ («Разработка моделей грамматической и семантической аннотации для русско-татарской лексикографической базы данных»), проект № 11-14-16024а/В.

1. Введение

Электронные аннотированные лексикографические ресурсы являются ценным информационным источником и эффективным инструментом для научных исследований и различных лингвистических приложений (автоматическая обработка текста, машинный перевод, информационный поиск). К настоящему времени разработаны компьютерные лингвистические модели и ресурсы для тюркских языков [1; 4; 5; 13]. В то же время, для татарской и тюркской компьютерной лингвистики являются актуальными разработки специализированных лексикографических ресурсов, ориентированных на решение теоретических и прикладных задач компьютерной обработки текстов.

Разработанная русско-татарская лексикографическая база данных характеризуется большой степенью детализации лингвистической аннотации и моделирует как отношения внутри лексических систем русского и татарского языков, так и отношения между двумя лексическими системами. Основу ее составляет русско-татарский электронный словарь *ABBYU Lingvo X3*, созданный на базе русско-татарского словаря под редакцией Ф. А. Ганиева [8].

Поскольку языковая действительность не может быть исчерпывающе представлена в рамках одной модели, функциональные особенности лексикографической базы данных определяются кругом потенциальных задач, на которые она ориентирована. С одной стороны, лексикографическая база данных рассматривается как интерактивный исследовательский инструмент, с другой стороны, данный ресурс может использоваться в качестве интегрированного модуля в системах автоматической обработки текста для решения различных задач, таких как автоматическое аннотирование корпусов текстов, автоматизированный перевод и информационный поиск и др.

База данных состоит из взаимосвязанных русской (R-) и татарской (T-) компонент, которые имеют независимую структуру, обусловленную языковой спецификой, и объединяются на уровне лексических эквивалентов. Каждая языковая компонента представляется грамматической и семантической моделью [7].

Лексический состав грамматической модели R-компоненты строится на базовом множестве словаря А. А. Зализняка [3] с терминологическими расширениями из различных предметных областей. Схема грамматического описания R-компоненты представляется в форме таблицы, в которой объектом описания является словоформа, заданная набором определенных атрибутов. Семантическая часть R-компоненты задается набором семантических признаков лексемы, основу которого составляет семантическая классификация, принятая в Национальном корпусе русского языка [6].

[©] Невзорова О. А., Салимов Ф. И., Хакимов Б. Э., Гатиатуллин А. Р., Гильмуллин Р. А., Галиева А. М., Якубова Д. Д., Аюпов М. М., 2012

2. Грамматическая аннотация *T*-компоненты

Грамматический сегмент *T*-компоненты представляется двумя взаимосвязанными словарями: словарем основ и словарем окончаний. Структуру словаря основ можно видеть из примера фрагмента словаря основ, представленного в Таблице 1.

Таблица 1.

Фрагмент словаря основ

T	Semant	DictForm	MNT1	TextForm	MNT2
N	1	Абажур	3		
N	1	лампа калпагы	19		
N	2	Аббат	5		
N	3	Аббатлык	9	Аббатлыг	11
N	4	Аббревиатура	1		
V	52	Авансла	29	Авансл	31
V	53	Аванслан	37		

Все изменения, происходящие с основой согласно морфонологическим правилам татарского языка в процессе аффиксации, отображаются в словаре основ. Для этого наряду со столбцом *DictForm* (словарная форма) используется специальный столбец *TextForm*, содержащий (при наличии) модифицированные формы основ.

Связь лексем в *R*- и *T*-компонентах словаря осуществляется с помощью семантического кода, который содержится в поле *Semant*. Переводным соответствиям в обоих языках присваивается единый семантический код. Все лексемы татарского языка в грамматической модели базы данных подразделяются на четыре морфонологических типа в зависимости от характера присоединяемых аффиксов. Информация о морфонологических типах хранится в столбце *MT* таблицы основ. В Таблице 2 приведен перечень этих морфонологических типов.

Таблица 2.

Описание морфонологических типов татарских лексем

№	Обозначение	Описание
1.	Тип <i>N</i>	Основы, к которым присоединяются аффиксы имени существительного
2.	Тип <i>V</i>	Основы, к которым присоединяются глагольные аффиксы
3.	Тип <i>A</i>	Основы, к которым присоединяются аффиксы степени сравнения и аффиксы имени существительного
4.	Тип <i>S</i>	Основы, к которым присоединяются только модальные аффиксы (-дыр, -мы, -ми)

Всем основам (как в столбце *DictForm*, так и в столбце *TextForm*) словаря основ присваиваются морфонологические типы. Морфонологический тип основы определяется типом алломорфов аффиксальных морфем, которые могут присоединяться к данной основе. Морфонологический тип основы определяется такими признаками как ряд гласных (сингармонизм), характер последнего символа основы, изменения, происходящие с основой при присоединении алломорфов [9]. Всего в соответствующем словаре базы данных представлено 84 морфонологических типа основ. В словаре основ морфонологические типы хранятся в столбце *MNT1* для словарной формы основы и в столбце *MNT2* для текстовой формы лексемы.

Словарь окончаний содержит возможные наборы окончаний татарских словоформ для каждого морфонологического типа. Окончания представляют собой цепочки алломорфов, образованные по морфонологическим правилам татарского языка. Теоретически эти цепочки словоизменяемых аффиксов в агглютинативном татарском языке могут иметь бесконечную длину. Однако в базе данных установлено ограничение до пяти алломорфов, что является обоснованным со статистической точки зрения.

В унифицированном представлении татарских аффиксальных морфем чередующиеся в алломорфах звуки обозначаются заглавными буквами, неизменяющиеся – строчными буквами. Для обозначения морфонологических категорий, выражаемых соответствующими морфемами, используется система аннотации, реализованная в программном инструментарии *PC KIMMO* [Там же].

Грамматическая модель *T*-компоненты содержит также словообразовательную составляющую. В структуре словаря основ представлен деривационный способ словообразования путем добавления специальных полей описания.

3. Семантическая аннотация *T*-компоненты

Семантическая аннотация имеет существенные отличия от грамматической, которая строится на строго определенном, фиксированном количестве грамматических категорий. Количество семантических признаков зависит от степени семантической генерализации, которая определяется целями и задачами классификации. К значительным трудностям при семантическом описании лексики приводит отсутствие во многих случаях ясной границы между таксонами, необходимость оперировать огромными множествами признаков

и свойств, сложности разграничения компонентов значения в лексической единице и часто невозможность однозначного определения их характера.

При семантическом аннотировании необходимо сохранить баланс между степенью детализированности разметки и ее простотой и прозрачностью, удобством для разработчика и пользователя. Важным преимуществом корпусно-ориентированной аннотации, которую трудно реализовать в печатных идеографических словарях, является реальная возможность учитывать пересекаемость классов слов, когда лексема наделяется различными, не привязанными друг к другу пометами.

Принятый подход к семантической разметке лексикографической базы данных использует семантическую разметку Национального корпуса русского языка [6] с рядом важных дополнений, учитывающих специфику лексической и словообразовательной системы татарского языка. При семантической разметке лексем в Национальном корпусе русского языка указываются лексико-грамматические разряды слов, их собственно лексико-семантические и деривационные характеристики. При семантической разметке лексем татарского языка также описываются лексико-семантические характеристики и разряд лексем с учетом их словообразовательных особенностей.

По аналогии с Национальным корпусом русского языка, при аннотировании татарских слов выделяются основное (основные) и дополнительное (дополнительные) значения. Разграничение основного и дополнительных значений не всегда является простой задачей и осуществляется при помощи семного анализа с опорой на дефиниции, закрепленные в словарях. Например, для лексемы *айкальми* в строке «Семантика основная» даются пометы *r:abstr, t:weather*, выражающие значение «сильное волнение (моря, водной поверхности)», а в строке «семантика дополнительная» - пометы *r:abstr, t:psych, sc:hum*, соответствующие переносному значению лексемы «переполох (людской массы)». Важен и порядок представления помет: от более общего (часть речи, разряд) к более конкретному, в результате чего иерархическая структура, положенная в основу классификации, проявляет себя на всех уровнях представления информации.

Для предметных имен определяется таксономия, мереология, топология и оценочность. Столь разностороннее рассмотрение предопределяет тот факт, что одна лексема может быть описана с помощью значительного количества (до 10) помет [12]. Для непредметных имен описывается таксономия, мереология и оценка, а для имен собственных осуществляется только тематическая классификация.

При разметке глаголов указывается каузативность, для фазовых и служебных глаголов используются специальные обозначения.

В таксономии глаголов указываются различные семантические признаки:

t:move – движение (*чабу* – бежать);

t:move:body – изменение положения тела, части тела (*утыру* – сесть, садиться);

t:put – помещение объекта (*төяү* – грузить, *кую* – класть, ставить);

t:impact – физическое воздействие (*сугу* – ударить);

t:impact:creat – создание физического объекта (*төзү* – строить);

t:impact:destr – уничтожение (*яндыру* – сжигать) и т.д.

Как уже отмечалось, в качестве основы была принята система лексико-семантической разметки Национального корпуса русского языка, с рядом корректив и дополнений, учитывающих специфику татарского языка, в зависимости от количественного и качественного состава слов различных лексико-семантических групп. Так, при описании глаголов физического воздействия (*t:impact*) нами используются различные конкретизирующие пометы, отсутствующие в НКРЯ, например:

t:impact:tool – воздействие при помощи орудия/инструмента (*борулау* - сверлить).

Посессивная сфера (*t:poss*) конкретизируется за счет использования помет, указывающих на направление и характер отношения обладания:

t:poss:acquire – приобретение (*табу* - находить, *кораллану* - вооружаться);

t:poss:deprive – лишение (*югалту* - терять, *коралсызландыру* - обезоружить).

При описании семантики производного глагола могут быть использованы пометы, представленные при описании семантики производящих слов, так как принцип системности предполагает использование одних и тех же помет для слов разных грамматических классов со сходным значением. В следующих примерах для глаголов с семантикой изменения в качестве уточняющих используются пометы, являющиеся основными при описании значений производящих прилагательных:

t:change:size – изменение размера (*зураю* – увеличиваться в размерах);

t:change:color – изменение цвета (*саргаю* – желтеть);

t:change:humq – изменение свойства человека (*явызлану* – озлобиться).

Мы считаем, что при аннотировании всех знаменательных частей речи татарского языка в лексикографической базе данных следует по возможности придерживаться единых принципов и системы обозначений.

В настоящее время для семантического аннотирования лексикографической базы татарского языка используется около 180 помет.

4. Функции поиска в русско-татарской лексикографической базе данных

Лексикографическая база данных поддерживает различные типы поисковых запросов. На проектирование потенциальных функций поиска существенно влияет принятая система аннотаций, а также структура базы данных. В процессе разработки функций поиска была решена нетривиальная задача трансформации основных исследовательских, прикладных и справочных запросов в стандартизированные логические формы.

Реализация поисковых функций в лексикографической базе данных учитывает существующий опыт в данной области [2; 6]. При этом, учитывая разноструктурность языков, для *R*- и *T*-компонент были выделены различные группы типовых поисковых запросов. Рассмотрим наиболее специфические запросы для *T*-компоненты. Типовыми запросами являются запросы о грамматических признаках заданной лексемы (словоформы) в виде основы и цепочки морфем или набора грамматических категорий (прямые запросы). Другой разновидностью являются обратные запросы, которые при задании грамматических признаков позволяют получать либо соответствующие выборки словоформ, либо варианты набора грамматических признаков по заданному набору аффиксальных морфем. Реализация множества таких запросов к базе данных позволяет моделировать работу морфологического анализатора или генератора словоформ.

Грамматические условия запроса задаются логической формулой (используется конъюнктивная нормальная форма). Реализация запросов требует уточнения модели представления информации о грамматических характеристиках в базе данных.

Приведем примеры прямого и обратного запроса к лексикографической базе данных.

а) При прямом запросе на вход подается словоформа, например, *урмангаракмы* 'ближе к лесу?'.
Морфологический анализ производится перебором всевозможных разбиений словоформы на корневую и аффиксальную части и при наличии такого разбиения на выход поступает разбиение словоформы в виде лемма+окончание с набором грамматических помет и их расшифровкой:

Лемма	Алломорфная структура окончания	Морфемная структура окончания	Комментарий
<i>урман</i>	<i>га+рак+мы</i>	<i>ГА+РАК+МЫ</i>	н.п., ср.ст., вопр.форм.2

б) При обратном запросе на вход подается набор грамматических признаков, например: падеж направительный, принадлежность (лицо первое, число единственное) или падеж направительный, принадлежность (лицо первое, число множественное). Этой комбинации признаков соответствует логическая формула в виде конъюнктивной нормальной формы (КНФ): $CASE_DIR(ГА) \& (IPOSS_Sing(ЫМ) \vee IPOSS_PLU(ЫБЫЗ))$.

Требуется выдать набор словоформ, имеющих заданные грамматические признаки. Ниже приведен фрагмент результата запроса:

Словоформы

аббатлыг+ым+a
абсолютлыг+ым+a
өрег+ем+ә
аббатлыг+ым+a+мы
абсолютлыг+ым+a+мы
өрег+ем+ә+ме
аббатлыг+ым+a+мыни
абсолютлыг+ым+a+мыни
өрег+ем+ә+мени
абажур+ыбыз+га

Обсудим некоторые нетривиальные вопросы представления данных и полученные решения:

1. В грамматической модели для определенных грамматических категорий целесообразно выделить грамматические значения, которые приписываются словоформе по умолчанию (при отсутствии явных указаний). При этом желательно однозначно толковать соответствующую конъюнкцию грамматических признаков.

Особенность поиска по грамматическим признакам в татарском языке (и вообще, в тюркских языках) заключается в том, что не все грамматические категории выражаются аффиксальными морфемами, и в некоторых случаях для определения того, что словоформа имеет определенное грамматическое значение, необходимо убедиться, что в структуре словоформы отсутствует противопоставленный по значению аффикс. В других случаях отсутствие соответствующего аффикса не обладает самостоятельным грамматическим значением, а лишь указывает на отсутствие грамматической категории (например, категория принадлежности в отличие от нулевой формы в единственном числе).

В более формализованном виде морфологические параметры словоформ в татарском языке с точки зрения внутренней структуры можно подразделить на простые и сложные [13]. Простые параметры представлены единственным значением (возможно наличие, либо отсутствие параметра). Сложные параметры представлены двумя и более альтернативными значениями и могут быть обязательными или факультативными. Факультативные параметры отличаются от обязательных тем, что могут отсутствовать в словоформе, все простые параметры являются факультативными. В результате анализа данных параметров для частей речи татарского языка было принято решение размечать по умолчанию грамматические признаки, соответствующие безаффиксальным формам сложных обязательных параметров.

2. Для эффективного использования запросов с отрицанием необходимо точное описание случаев возможного применения отрицания. С точки зрения пользователя-лингвиста, релевантным может быть запрос с отрицанием любого из грамматических параметров. Однако учет вышеуказанных типов признаков позволяет сократить случаи использования отрицания и избежать неоправданного усложнения модели. Так, невозможно отрицание обязательных сложных параметров.

С другой стороны, указание при разметке в качестве самостоятельного значения нулевых форм сложных факультативных параметров (как в случае с описанной выше категорией принадлежности) также позволяет избежать в этих случаях запроса с отрицанием. Таким образом, сфера применения запросов с отрицанием ограничивается вариантами сложных обязательных параметров и простыми факультативными параметрами.

3. Особого подхода требуют так называемые «рекурсивные» аффиксы, для правильной интерпретации которых алгоритм отбора должен учитывать кратность вхождения грамматического признака в словоформе. Подобными аффиксами в татарском языке являются атрибутивизатор *-ДАГЫ* и посессивный аффикс *-НЫКЫ*. Возможно многократное присоединение этих аффиксов к именной основе с возникновением «циклической» словоформы, которая теоретически может иметь неограниченную длину (*урман+нар+ым+дагы+лар+ым+дагы...* «...те, что в тех, что в моих лесах»).

В нашей модели в таких словоформах выделяются рекурсивные группы морфем, где каждой морфеме *-ДАГЫ* или *-НЫКЫ* могут быть сопоставлены собственные морфемы множественности и принадлежности: *урман+нар+ым+(дагы+лар+ым)¹+(дагы+лар+ым)²+...+(дагы+лар+ым)ⁿ+ны+дыр*. Таким образом, падежные и модальные морфемы присоединяются только к последнему элементу этой последовательности. В то же время, учитывая характер реального функционирования данного явления в языке, можно задать ограничение на максимальное число повторений рекурсивной группы.

Заключение

Предложенные модели грамматической и семантической аннотации татарской лексики используются для аннотирования русско-татарской лексикографической базы данных. В настоящее время аннотирована экспериментальная выборка татарской лексики, реализованы функции прямого и обратного поиска к лексикографической базе данных.

В число первоочередных задач входит расширение базовой выборки лексем, уточнение и усовершенствование концептуальной модели лексикографического ресурса, расширение словообразовательной информации за счет других способов словообразования, усиление горизонтальной интеграции на разных языковых уровнях.

Семантическое аннотирование *T*-компоненты выполнено на основе классификации, принятой в Национальном корпусе русского языка. Произведена адаптация данной системы для описания татарской лексики. Размеченный ресурс позволяет осуществлять сопоставительные исследования семантики русских и татарских лексем, что является важным инструментом для лингвистических задач.

Предложенный подход к организации информационного поиска в русско-татарской лексикографической базе данных учитывает особенности агглютинативной морфологии тюркских языков. По мере развития лексикографической базы данных ее функциональность будет дополнена возможностями поиска по семантическим и словообразовательным признакам.

Список литературы

1. Бухараев Р. Г., Сафиуллина Ф. С., Сулейманов Д. Ш. и др. К концепции Машинного фонда Республики Татарстан // Интеллект. Язык. Компьютер. Казань: Изд-во Казан. гос. ун-та, 1995. Вып. 2. С. 20-35.
2. Восточно-армянский национальный корпус [Электронный ресурс]. URL: <http://eanc.net> (дата обращения: 15.09.2012).
3. Зализняк А. А. Грамматический словарь русского языка. Словоизменение. М.: Русский язык, 1977. 880 с.
4. Казанский лингвографический фонд [Электронный ресурс]. URL: <http://klf.ksu.ru> (дата обращения: 12.09.2012).
5. Машинный фонд башкирского языка [Электронный ресурс]. URL: <http://mfbl.ru> (дата обращения: 15.09.2012).
6. Национальный корпус русского языка [Электронный ресурс]. URL: <http://ruscorpora.ru> (дата обращения: 15.09.2012).
7. Невзорова О. А., Хакимов Б. Э., Салимов Ф. И., Гатиатуллин А. Р., Гильмуллин Р. А. Проект создания двуязычного лексикографического ресурса для задач автоматической обработки текстов // Диалектология, история и грамматическая структура тюркских языков. Казань, 2011. С. 169-172.
8. Русско-татарский словарь / под ред. Ф. А. Ганиева. М.: ИНСАН, 1997. 718 с.
9. Сулейманов Д. Ш., Гильмуллин Р. А. Двухуровневое описание морфологии татарского языка // Тезисы Международной научной конференции «Языковая семантика и образ мира». Казань: Изд-во Казан. гос. ун-та, 1997. Кн. 2. С. 65-67.
10. Сулейманов Д. Ш., Невзорова О. А., Гатиатуллин А. Р. и др. Основные компоненты прикладной грамматической модели татарского языка // Труды Международной научной конференции «Диалог-2007». М., 2007. С. 525-530.
11. Сулейманова Д. Д. Особенности семантической классификации имен существительных с конкретным значением в русском и татарском языках // Труды Казанской школы-семинара по компьютерной и когнитивной лингвистике TEL'2012. Казань: Фэн, 2012. С. 168-174.
12. Хакимов Б. Э., Гильмуллин Р. А. К разработке морфологического стандарта для систем автоматической обработки текстов на татарском языке // Системный анализ и семиотическое моделирование: материалы всеросс. конф. с международ. участием (SASM-2011). Казань, 2011. С. 209-214.
13. Turk Dil Kurumu [Электронный ресурс]. URL: <http://tdk.org.tr> (дата обращения: 10.09.2012).

**SEMANTIC-GRAMMATICAL ANNOTATION
IN RUSSIAN-TATAR LEXICOGRAPHICAL DATABASE**

Ol'ga Avenirovna Nevzorova, Ph. D. in Technical Sciences, Associate Professor

Farid Ibragimovich Salimov, Ph. D. in Physical and Mathematical Sciences, Associate Professor

Bulat Ernstovich Khakimov, Ph. D. in Philology

Airat Rafizovich Gatiatullin, Ph. D. in Technical Sciences

Rinat Abrekovich Gil'mullin, Ph. D. in Physical and Mathematical Sciences

Al'fiya Makarimovna Galieva, Ph. D. in Philosophy, Associate Professor

Dilyara Dzhavdetovna Yakubova, Ph. D. in Philology

Madekhur Maskhutovich Ayupov

Kazan' (Volga Region) Federal University

Research Institute "Applied Semiotics" of Academy of Sciences of the Republic of Tatarstan

*onevzoro@gmail.com; Farid.Salimov@ksu.ru; khakeem@yandex.ru; agat1972@mail.ru; rinatgilmullin@gmail.com;
agalieva@yandex.ru; suleymanovad@gmail.com; madehur@mail.ru*

The authors describe the models and methods of linguistic information representation in the developed Russian-Tatar lexicographical database, which distinctive features are the orientation at text processing computer systems, as well as the profound detailed elaboration and mutual integration of linguistic data, and tell that the created lexicographical base for the first time uses the semantic-grammatical annotations of the Tatar lexemes, formed taking into account the typological features of the Turkic languages, and has the potential to be used in multilingual computer applications.

Key words and phrases: lexicographical database; automatic text processing; linguistic annotation; grammar; semantics; retrieval request; Turkic languages.

УДК 81'255.2:6

Филологические науки

В статье излагаются основные аспекты реферативного перевода как сложного вида иноязычной речевой деятельности. Особое внимание уделено этапу аналитической обработки оригинала, заключающейся в осмыслении, понимании и свертывании воспринимаемой информации. Затрагивается вопрос об оптимальном уровне понимания исходного текста в соотношении с общими требованиями к квалификации переводчика-референта.

Ключевые слова и фразы: реферативный перевод; референтская деятельность; переводчик-референт; смысловое восприятие; понимание.

Наталья Михайловна Нестерова, д. филол. н., профессор

Антон Юрьевич Наугольных, к. пед. н.

Кафедра иностранных языков, лингвистики и межкультурной коммуникации

Пермский национальный исследовательский политехнический университет

nest-nat@yandex.ru; crispian@yandex.ru

Евгения Андреевна Наугольных, к. филол. н.

Кафедра иностранных языков

Пермская государственная фармацевтическая академия

ru_lina_jane@mail.ru

**«СКАЗАТЬ МАЛО, НО ХОРОШО»:
ДЕЯТЕЛЬНОСТНАЯ ХАРАКТЕРИСТИКА РЕФЕРАТИВНОГО ПЕРЕВОДА[©]**

Сектор технического перевода является одновременно и самым крупным, и наиболее быстро растущим сектором отечественного рынка переводческих услуг. По разным подсчетам, доля технических текстов в общем объеме переводимых материалов составляет от 60 до 90% [10], таким образом, все основные тенденции рынка переводов сегодня так или иначе обусловлены состоянием спроса на перевод технической документации.

Растет интерес участников профессиональной коммуникации к сокращенным видам перевода, в числе которых особое место отводится реферативному переводу как вторичному тексту, претендующему на замещение оригинала с точки зрения объема ключевой информации.

Вынесенная в заглавие аллюзия на высказывание К. Пруткина как нельзя точнее передает суть референтской деятельности. Вместе с тем нельзя не признать, что все серьезные исследования этого вида языкового преобразования остались в XX столетии. Методологической базой научных изысканий в данном направлении