

УДК 81'322+81'37+811.512.14+81'367.625

БАЗОВЫЕ ПРИНЦИПЫ ПОСТРОЕНИЯ ТЕЗАУРУСА ТИПА WORDNET ДЛЯ ТАТАРСКОГО ЯЗЫКА*

А.М. Галиева, О.А. Невзорова

Аннотация

В статье рассматриваются принципы построения нового лексического ресурса – тезауруса типа wordnet для татарского языка на базе глагольной лексики. Дан обзор основных лексикографических ресурсов, привлекаемых для целей проекта. На примере глаголов звучания обсуждается общая методология разработки wordnet-ресурсов и рассматриваются ключевые проблемы, возникающие при построении синсетов для татарского языка.

Ключевые слова: wordnet, глагол, татарский язык, синонимы, синсет.

Введение

Тезаурусы типа wordnet созданы во многих странах и успешно используются при обработке естественного языка: при информационном поиске, разрешении многозначности, анализе тональности и др. Принципы построения и особенности представления лексической информации в тезаурусах подобного типа представлены в [1, с. 61–97]. Информация о существующих и разрабатываемых в настоящее время wordnet'ах представлена на сайте Всемирной WordNet-ассоциации (The Global WordNet Association), <http://globalwordnet.org/wordnets-in-the-world>.

Wordnet'ы для разных языков находятся на разной стадии разработки. На сегодняшний день отсутствует достоверная информация о проектах по созданию тезаурусов типа wordnet для тюркских языков, поэтому данная задача является очень актуальной. Проект турецкого wordnet'а был инициирован Лабораторией языка и речевых технологий (The Human Language and Speech Technologies Laboratory) в университете Сабанчи (Sabanci University) под руководством Кемаля Офлазера (Kemal Oflazer) [2], но, к сожалению, работа над проектом не получила продолжения.

В данной статье рассматривается подход к созданию wordnet-тезауруса для татарского языка на базе глагольной лексики. Глаголы образуют ядро лексико-семантической и грамматической системы любого языка и характеризуются концептуальным богатством, сложностью семантической структуры и синтагматики. Семантика глагола не только номинативна, но и реляционна, что требует оптимальной интеграции разных исследовательских приемов при ее описании с учетом как сигнификативных, так и синтагматических компонентов значения.

* Статья выполнена при финансовой поддержке РФНФ и Правительства Республики Татарстан (проект № 14-14-16031a(p)).

Актуальность и научная новизна проекта обусловлены необходимостью создания идеографических словарей татарского языка и комплексного исследования лексико-семантических структур татарского языка на материале глагольных лексем. В настоящее время в татарском языкознании отсутствуют специальные исследования комплексного характера, раскрывающие в полном объеме своеобразие семантической организации системы татарских глаголов (для всех лексико-семантических групп). Практическая значимость проекта связана с созданием из алфавитного массива глагольной лексики иерархической семантической классификации глаголов как сети синсетов, а также с многоаспектным структурно-семантическим описанием единиц полученных классов глагольных лексем. В рамках проекта предполагается систематизация массива татарских глаголов: 1) по доминирующим семантическим компонентам и общности значения (с определением количественных и качественных характеристик лексико-семантических категорий глагола на базе лексических парадигм); 2) по синтагматическим компонентам значения и лексической валентности (моделирование семантической совместимости между предикатом и его актантами, определение характера логических отношений между ними).

1. Используемые источники

При разработке wordnet'ов обычно используются источники, которые могут быть поделены на следующие типы:

- 1) принстонский WordNet (PWN) как первоначальный каркас (оттуда заимствуются тип лексикографической базы данных, идея представления концептов через синсеты, типы отношений между синсетами);
- 2) существующие таксономии языка (как на уровне слов, так и на уровне отдельных значений);
- 3) двуязычные словари (обычно используются для перевода с английского языка синсетов PWN);
- 4) моноязычные лексиконы [3, р. 65].

Дадим краткий обзор основных источников, используемых нами при создании wordnet-тезауруса для татарского языка на базе глагольной лексики.

1. Печатный словарь синонимов татарского языка (Синонимнар сүзлеге), составленный Ш.С. Ханбиковой и Ф.С. Сафиуллиной, содержит около 25000 близких по значению слов разных частей речи, представленных в 4500 синонимических рядах. Слова в словаре даны без толкований, стилистических помет и иллюстративного материала, но после доминанты представлен вариант ее перевода на русский язык. По мнению составителей, в татарском языке глагол и прилагательное в наибольшей степени демонстрируют богатство синонимических рядов (СС, б. 7). Тем не менее словарь не лишен некоторых недочетов. В частности, в силу недостаточной проработанности теоретических аспектов синонимии, словарь в ряде случаев содержит описательные выражения. Приведем в качестве иллюстрации синсет *ЖИҺАД (война за распространение религии) – дин өчен сугыш* (СС, б. 40); здесь вместо синонима представлена дефиниция (в которой к тому же нарушено правило соразмерности дефиниента и дефиниции). Кроме того, материал словаря весьма ограничен и не отражает реального богатства татарского языка. Являясь стандартным словарем синонимов,

Синонимнар сүзлеге не учитывает семантических отношений между синсетамы, и его материалы могут быть использованы лишь как отправная точка для последующей аналитической работы: требуется значительное увеличение числа синсетов, корректировка и пополнение содержания представленных в словаре синонимических рядов, установление их иерархии. Отметим, что на данный момент общедоступные электронные словари синонимов татарского языка отсутствуют.

2. Толковые словари татарского языка (трехтомное издание 1977–1981 гг. (ТС1) и однотомное 2005 г. (ТС2)) отражают основной массив лексики татарского языка, включая лексику различных сфер употребления; в толковых словарях синонимы часто являются способом определения значения слова. Материалы толковых словарей являются основным источником при разработке лексикографической базы данных татарского wordnet'a.

3. Двухязычные русско-татарские словари также могут служить источником для извлечения синонимов, поскольку в них синонимы обычно представлены как способы перевода того или иного русского слова на татарский язык. Так, Русско-татарский словарь под редакцией Ф.А. Ганиева (РТС) содержит 47000 слов. В основу его словника был положен Словарь русского языка С.И. Ожегова (1977 г.); словарь содержит как общепотребительную лексику и словосочетания, раскрывающие значения слов, так и специальную терминологию. Разбивка на значения производится исходя из многозначности лексем русского языка, что позволяет дать значительное количество татарских синонимов при переводе; в словаре учтена стилистическая и эмоциональная окраска слов. Электронная версия РТС реализована фирмой АВВУУ.

4. Татарский национальный корпус «Туган тел» (ТК) является информационно-справочной системой, представляющей собой аннотированную коллекцию текстов на современном татарском языке. Объем корпуса на сентябрь 2013 г. составлял более 26 млн словоупотреблений. Корпус содержит тексты различных жанров (художественная литература, тексты СМИ, тексты официальных документов, учебная литература, научные публикации и др.). Каждый документ имеет метаописание (указываются авторы, их пол, выходные данные и дата создания текста, его жанровая характеристика и др.). Тексты, включенные в корпус, снабжены морфологической разметкой (информация о части речи и грамматических характеристиках словоформы). Морфологическая разметка текстов корпуса выполняется автоматически с использованием модуля двухуровневого морфологического анализа татарского языка, реализованного в программном инструментарии РС-КИММО. Для поддержки корпуса адаптирована поисковая система Восточноармянского национального корпуса (ЕАНС), позволяющая искать материал по лексеме, словоформе, а также по отдельным грамматическим характеристикам [4]. Татарский корпус «Туган тел» доступен по адресу http://web-corpora.net/TatarCorpus/search/?interface_language=ru.

Таким образом, для создания wordnet-тезауруса мы имеем значительный массив лексикографических источников и корпусную коллекцию текстов. Комплексный характер представления лингвистической информации в wordnet'ax требует привлечения различных источников. ТК содержит массив данных о реальном функционировании слов в языке, но лексико-семантическая информация

в корпусах представлена в имплицитной форме и предполагает применение специальных, зачастую чрезвычайно сложных и трудоемких аналитических процедур. Словари включают ограниченный лексический материал, их составители работают в рамках определенных теоретических установок, которые мы не всегда разделяем. Соответственно, при моделировании системы татарских глаголов и выстраивании сети синсетов лексикографический материал при необходимости нами дополняется, проверяется и уточняется посредством корпусных данных.

2. Методология построения тезауруса типа wordnet

Основными единицами структуры словарей типа wordnet являются синонимические ряды (синсеты), которые связаны между собой различными семантическими (парадигматическими и синтагматическими) отношениями.

Синонимические отношения, связывающие слова по общности значений, являются базовыми для всех тезаурусов типа wordnet. Синонимические отношения определяются не между словами, а между лексико-семантическими вариантами слова.

Несмотря на то что понятие синонимии является общепринятым, точные критерии синонимичности до сих пор являются предметом дискуссий. В рамках проектов PWN и EuroWordNet синонимия определяется через понятие взаимозаменяемости: так, в проекте EuroWordNet слова считаются семантически эквивалентными, когда они обозначают один и тот же ряд сущностей, независимо от морфолого-синтаксических, стилистических, диалектных различий, а также различий в прагматическом использовании слова. Кроме того, синонимы не могут быть связаны между собой другими типами семантических отношений [5, p. 18].

Разработчики wordnet-тезауруса RussNet для русского языка критерий взаимозаменяемости рассматривают как дополнительный по отношению к критерию семантической близости. Последний выявляется при дефиниционном анализе, для которого требуется установление идентичности словарных определений или взаимная отсылка в синонимических определениях [6].

В наших исследованиях критерий общности значения считается основополагающим. Таким образом, в проекте по разработке татарского wordnet-тезауруса синонимами считаются лексико-семантические варианты слов одной и той же части речи с одинаковым или близким значением, которые могут быть взаимозаменяемы в пределах контекста. Такое определение синонимов следует считать рабочим, поскольку оно не может претендовать на всесторонность охвата сущности синонимии в языке.

Семантические отношения между синсетами для глаголов, традиционно выделяемые в тезаурусах типа wordnet, описываются следующим образом:

- а) отношения следования (Entailment): *бару* 'идти' – *атлау* 'шагать';
- б) отношения тропонимии: *айту* 'сказать' – *пышылдау* 'шептать';
- в) отношения каузативности: *кыштырдау* (некаузативное значение 'шуршать', например *что-то шуршит*) – *кыштырдау* (каузативное значение 'шуршать', например *шуршать бумагами*); в татарском языке в подавляющей части

глаголов каузативность регулярно выражается морфологически, хотя имеется и лексическая каузативность (типа *ташу* ‘таскать’, *илту* ‘относить’).

Отношение тропонимии – это особый случай отношения следования. Глагольные иерархии, представляющие отношение тропонимии, обладают более узкой, но в то же время более кустистой структурой по сравнению с существительными, число уровней в иерархии при этом обычно не превышает четырех [1, с. 69].

При создании словарей типа wordnet обычно используются два основных подхода.

1. В качестве основы «Модели расширения» (Expand Model) используется база данных PWN, синсеты которого последовательно переводятся с помощью двуязычных электронных словарей.

2. В «Модели объединения» (Merge Model) отбор лексики и установление семантических отношений между словами производится на базе ресурсов того языка, для которого конструируется тезаурус wordnet, затем полученные синсеты выравниваются по базовым концептам PWN и других wordnet'ов [5, p. 52].

При применении «Модели расширения» можно быстро и без усилий получить ядро для wordnet'a, однако в этом случае лексический состав языка подгоняется под словарный состав и семантические структуры английского языка. Результатом может явиться значительное количество описательных переводов английских лексем, при этом многие монолексеммы языка, не имеющие эквивалентов или аналогов в английском языке, с большой вероятностью будут потеряны. Особенно это существенно для языков, значительно отличающихся по типу и структуре от английского языка. При использовании «Модели расширения» наибольшего успеха добиваются разработчики, которые не применяют автоматический перевод, а прибегают к помощи профессиональных переводчиков и обеспечивают контроль перевода лексикографической базы данных PWN по качеству и покрытию всех слов и значений. Одним из таких проектов можно назвать финский FinnWN; его уникальность в том, что на сегодняшний день это один из крупнейших имеющихся wordnet'ов для языков неиндоевропейского типа, полученный на основе перевода базы данных PWN с последующей коррекцией [7].

Более трудоемкая «Модель объединения» предполагает создание модели семантической системы языка по имеющимся лексикографическим источникам и корпусным данным, то есть в этом случае обеспечивается сохранение лингвостатистических особенностей.

В проекте по разработке татарского wordnet-тезауруса каждый синсет включает глаголы разных типов:

1) синтетические глаголы: *булышу* ‘помогать’, *шыбырдау* ‘шуршать, барабанить’ (о дожде или падающей сверху воде);

2) аналитические глаголы, образованные от полнозначного знаменательного слова и вспомогательного глагола: *ярдәм итү* ‘помогать’, *ярдәм кылу* ‘помогать’;

3) аналитические глаголы, образованные от звукоподражательного или образоподражательного слова: *шыбыр-шабыр итү* ‘шуршать, барабанить’ (о дожде или падающей сверху воде), *жәм-жәм итү* ‘мерцать, блестеть’;

4) аналитические глаголы, образованные от полнозначного знаменательного слова и полнозначного глагола в функции вспомогательного: *ашыйсы килү* ‘проголодаться’.

На первом этапе проекта нами была подготовлена выборка из 5156 наиболее употребительных татарских синтетических и 350 аналитических глаголов, которые в настоящее время классифицируются по целому ряду семантических и грамматических признаков с учетом парадигматических и синтагматических отношений для выявления синонимов; одновременно с этим осуществляются процессы определения внутриязыковых семантических отношений между синсетам и выстраивания семантической иерархии синсетов.

3. Особенности создания фрагмента тезауруса на примере глаголов звучания

Разработка wordnet-тезауруса глагольной лексики для татарского языка предполагает:

- 1) выделение слов отдельных лексико-семантических групп (ЛСГ) глаголов на базе имеющихся лексикографических ресурсов и корпусных данных;
- 2) выделение синонимических рядов внутри ЛСГ;
- 3) построение иерархии синсетов.

При выделении глагольных семантических классов учитываются их тематические показатели, синтагматические свойства и компоненты структуры лексического значения. Слова одной лексико-семантической группы обладают совокупностью взаимозависимых свойств: грамматических, словообразовательных, синтаксических, коммуникативно-дискурсивных. Эти свойства в значительной степени мотивированы семантически, поэтому семантические классы мы стремимся строить таким образом, чтобы свойства отдельных членов класса (хотя бы в тенденции) устанавливались на основе функции принадлежности классу [8, с. 57].

Рассмотрим некоторые особенности разработки фрагмента тезауруса на примере глаголов звучания и выделим ключевые проблемы, возникающие при описании глагольной семантики.

Языковой материал показывает, что в татарском языке большая часть глаголов звучания имеет звукоподражательный характер и отличается яркой денотативно ориентированной семантикой, поскольку данные глаголы тесно привязаны к миру вещей, производящих звуки, к их объективным физическим или иным характеристикам. Звукоподражательный характер глагола приводит к тому, что значительное число таких глаголов в толковых словарях татарского языка дается исключительно путем буквального воспроизведения, но не описания характера звуков, обозначаемых глаголом: *даңгырдау* – ‘производить звуки, напоминающие «даңгыр»’ (ТС2, б. 139); *дыңгырдау* – ‘производить звуки, напоминающие «дыңгыр»’ (ТС2, б. 152) и т. п.

В таких случаях требуется предварительная аналитическая работа по раскрытию значения лексемы. Так, звукоподражание может существенно отличаться даже внутри синонимического ряда:

{*безләу, бызлау, бызылдау, жуылдау, тызылдау*} – ‘жужжать’ (о насекомых).

Табл. 1

Примеры каузативных пар глаголов

‘издавать звук, названный звукоподражательной основой’	‘каузировать издавание звука, названного звукоподражательной основой’
<i>даңылдау</i>	<i>даңылдату</i>
<i>даңгырдау</i>	<i>даңгырдату</i>
<i>дыңылдау</i>	<i>дыңылдату</i>
<i>дыңгырдау</i>	<i>дыңгырдату</i>
<i>деңгердәу</i>	<i>деңгердәту</i>

Фонетическая близость глаголов звучания часто является показателем того, что они могут быть включены в один синсет:

{*даңылдау, даңгырдау, дыңылдау, дыңгырдау, деңгердәу*} – ‘грохотать, тарыхтеть, громко стучать’ (следует иметь в виду, что русские словарные соответствия передают звучание этих глаголов очень неточно).

В подобных синсетах выделение доминирующего глагола осуществляется не по семантическим или стилистическим, а по чисто квантитативным характеристикам на основе корпусных данных: наиболее частотный глагол представляется как доминанта.

Татарские глаголы, обозначающие звуки, издаваемые неживыми предметами (исключая явления природы), регулярно имеют каузативные корреляты, то есть между синсетами существуют каузативные отношения. В табл. 1 представлены примеры каузативных пар татарских глаголов.

Показательно, что мотивирующими (первичными) в татарском языке являются именно некаузативные формы, а каузативные – производными. Это указывает на значимость для татарского языкового сознания того факта, что определенные типы звуков издаются отдельными классами неживых предметов, такая способность является важным свойством мира вещей (*яфрак кыштырды* означает, что *шуршат* именно *листья*, а не *человек листьями*). Но необходимая для производства звуков активность каузатора регулярно фиксируется путем употребления понудительного залога: неодушевленные предметы не издают звуки самопроизвольно, но лишь под внешним воздействием, что маркируется грамматически.

В значительной части синсетов с глаголами звучания представлены лексемы, которые могут быть рассмотрены в качестве фонетических вариантов одного глагола. Тем не менее они часто по-разному отражены в лексикографических источниках, что может быть связано как с недостаточной системностью дефиниций словарей, так и с отличиями в значении слов:

- *кылтырау* – ‘издавать звуки, напоминающие «кылтыр-кылтыр»’ (ТС2, б. 306), *келтерәу* – ‘издавать ритмичные мерные звуки’ (ТС2, б. 246);

- *кылтырау* – ‘трястись с шумом, дрожать с шумом’ (ТРС, с. 189), *келтерәу* – ‘дрезжать, трястись с легким звуком’ (ТРС, с. 153).

По данным ТК, сочетаемостные свойства этих глаголов также совпадают не полностью: так, *тегу машинасы* (швейная машинка) *келтери* (стрекошет), но не *кылтырый*.

В связи с этим для каждой группы таких глаголов вопрос о том, считать ли их вариантами или синонимами, должен решаться индивидуально. При близости значений такие лексемы включаются нами в один синсет:

{*быкырдау, быгырдау*} – ‘бурлить, клокотать’ (о жидкости).

В этот же синсет должен быть включен глагол *гөбердәу* с аналогичным значением: *су быкырдап/быгырдап/гөбердәп кайнап чыга* (вода, клокоча, закипает), поскольку в данном контексте приведенные выше глаголы могут заменять друг друга.

Но сходство фонетического облика двух глаголов не всегда является признаком того, что они имеют одинаковую семантику. Рассмотрим глаголы *шыгырдау* ‘скрипеть’ и *шыкырдау* ‘скрежетать’, которые по своей фонетической структуре аналогичны глаголам *быгырдау* и *быкырдау*: согласные в середине корня отличаются лишь признаком звонкости/глухости (г/к). Тем не менее рассматриваемые глаголы имеют разную семантику и сочетаемость:

1) глагол *шыгырдау* имеет широкую сочетаемость; согласно корпусным данным источником звучания при нем могут быть объекты разных типов: *кар* (снег), *ишек* (дверь), *идән* (пол), *көе сиртмәсе* (ворот колодца), *агач* (древесина), *карават* (кровать), *арба* (телега) и т. п.;

2) глагол *шыкырдау* имеет ограниченную сочетаемость: с существительным *теш/тешләр* (зубы), встречаются также сочетания с деепричастными формами: *шыкырдап кату*, *шыкырдап туңу* ‘замерзнуть, стать твердым от мороза’.

Оба этих глагола обозначают зону звуков, испускаемых при трении поверхностей предметов или их частей. Но при этом *шыгырдау* предполагает некоторую длительность и гомогенность звучания трущихся объектов, а *шыкырдау* – серию коротких неравномерных звуков, обычно издаваемых небольшими по размеру твердыми объектами. В подавляющем количестве контекстов они не являются взаимно заменимыми: такие сочетания, как **кар шыкырдый*, **идән шыкырдый*, **көе сиртмәсе шыкырдый* и т. п., в татарском языке невозможны. Сочетания *теше шыгырдый* и *теше шыкырдый* имеют аналогичные отличия в семантике, что и русские сочетания *скрипеть зубами* и *скрежетать зубами* (с учетом некаузативности татарских глаголов в приведенных выше сочетаниях).

Таким образом, глаголы *шыгырдау* и *шыкырдау*, несмотря на сходство их фонетического облика, должны быть включены в разные синсеты:

• {*шыгырдау, шыгыр-шыгыр килү, шыгыр-шыгыр итү, шыгыр-шагыр килү, шыгыр-шагыр итү*} – ‘скрипеть’;

• {*шыкырдау, шыкырт-шыкырт итү, шакыр-шыкыр килү*} – ‘скрежетать’.

Важным критерием для включения лексемы в синсет является ее представленность в ТК. Так, например, аналитические глаголы *шыгыр-шогыр килү, шыгыр-шогыр итү* зафиксированы в словаре Ш.С. Ханбиковой и Ф.С. Сафиуллиной (СС., б. 116), но отсутствуют в ТК, поэтому нами не включены в синсет {*шыгырдау*}.

Многозначные глаголы включаются в разные синсеты, исходя из их лексико-семантических вариантов:

• {*быкырдау, быгырдау, гөбердәу*} – ‘бурлить, клокотать’ (о жидкости);

• {*быкырдау, быгырдау, гөбердәу, чурлау*} – ‘урчать’ (о желудке).

В проекте EuroWordNet выделяются базовые концепты (Base Concepts), которые являются значимыми в структуре wordnet-тезауруса. Значимость концепта определяется как частотностью его употребления, так и соотносительностью с большим числом других концептов, а также тем, что он может функционировать в качестве якоря при привязке других концептов и выстраивании иерархии [9, p. 12]. Базовыми обычно являются концепты верхних уровней.

Как показывает языковой материал, одной из характерных особенностей глагольной лексики татарского языка является наличие богатого инвентаря концептов нижних уровней (например, для глаголов звучания характерно наличие большого количества слов, описывающих отдельные особенности звуков, издаваемых неживыми предметами) при лакунах с обозначениями гиперонимов. Так, в татарском языке для глагола *звучать* ('издавать звук') нет соответствующего эквивалента, тем не менее сема 'издавать звук' существует в связанном виде в значении большого количества глаголов звучания с конкретным значением. Глагол *яңгырау*, который в двуязычных словарях обычно описывается как татарский аналог для глагола *звучать* (см., например, (РТС, с. 181)), в действительности обозначает процесс распространения звука в пространстве, но не испускания звука (в татарском языке невозможны такие сочетания, как *скрипка звучит*, *гитара звучит*, но можно говорить *скрипка тавышы яңгырый* (раздаются звуки скрипки)). Тем не менее невербализованный татарский концепт 'звучать' совершенно необходим при выделении и структурировании ЛСГ глаголов звучания.

Таким образом, рассмотрев особенности создания фрагмента тезауруса, можно выделить следующие ключевые задачи, решение которых потребуется при построении синсетов и их иерархической сети:

- уточнять значение лексемы в случаях, когда дефиниция в толковых словарях является неполной;
- определять степень близости значений слов, имеющих значительное сходство в фонетическом облике, для включения/невключения их в один синсет;
- включать в синсеты аналитические формы;
- учитывать сочетаемостные свойства глаголов;
- отображать каузативные пары;
- отображать отношения следования;
- конструировать гиперонимы (при отсутствии вербализованной формы);
- учитывать представленность глагольной формы в ТК.

Заключение

Проект по разработке wordnet-тезауруса татарских глаголов позволяет объединить опыт традиционной татарской лексикографии и современных информационных технологий. Применение корпусных технологий дает возможность создавать ресурс, адекватно отражающий распределение слов татарского языка и их лексико-семантических вариантов в реальном контекстуальном окружении. Одна из важнейших задач проекта – отображение лингвоспецифичности семантической системы татарских глаголов в ресурсе, отвечающем нуждам современной компьютерной лингвистики.

Summary

A.M. Galieva, O.A. Nevzorova. Basic Principles of Formation of a WordNet-Like Thesaurus for the Tatar Language.

The paper describes the basic principles of creation of a WordNet-like thesaurus for the Tatar language on the basis of verbal vocabulary. A review of the main lexicographical resources involved for the project is given. Using verbs of sound emission as an example, the authors discuss general methods for the development of WordNet-like resources and examine the key problems arising during the construction of synsets for the Tatar language.

Keywords: WordNet, verb, Tatar language, synonyms, synset.

Источники

- СС – Ханбикова Ш.С., Сафиуллина Ф.С. Синонимнар сүзлеге: 25 меңгә яқын синонимик берәмлек. – Казан: Хәтер, 1999. – 256 б.
- ТС1 – Татар теленең аңлатмалы сүзлеге: 3 т. – Казан: Тат. кит. нәшр., 1977–1981.
- ТС2 – Татар теленең аңлатмалы сүзлеге / Баш ред. Ф.Ә. Ганиев. – Казан: Матбугат йорты, 2005. – 848 б.
- РТС – Русско-татарский словарь / Под ред. Ф.А. Ганиева. – М.: ИНСАН, 1997. – 720 с.
- ТК – Татарский национальный корпус «Туган тел». – URL: http://web-corpora.net/TatarCorpus/search/?interface_language=ru, свободный.
- ТРС – Татарско-русский словарь / Под ред. Ф.А. Ганиева. – Казань: Тат. кн. изд-во, 2004. – 488 с.

Литература

1. Лукашевич Н.В. Тезаурусы в задачах информационного поиска. – М.: Изд-во Моск. ун-та, 2011. – 511 с.
2. Bilgin O., Çetinoglu Ö., Oflazer K. Building a Wordnet for Turkish // Romanian Journal of Information Science and Technology. – 2004. – V. 7, No 1–2. – P. 163–172.
3. Farreres X., Rigau G., Rodriguez H. Using WordNet for Building WordNets // Proc. COLING-ACL Workshop “Usage of Wordnet in Natural Language Processing Systems”. – Montreal, Canada, 1998. – P. 65–72.
4. 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 // Филол. науки. Вопросы теории и практики. – Тамбов: Грамота, 2012. – № 7, ч. 1. – С. 141–146.
5. EuroWordNet General Document. Version 3. / Ed by P. Vossen. – URL: <http://vossen.info/docs/2002/EWNGeneral.pdf>, свободный.
6. Азарова И.В., Митрофанова О.А., Синопальникова А.А. Компьютерный тезаурус русского языка типа wordnet. – URL: <http://www.dialog-21.ru/Archive/2003/Azarova.htm>, свободный.
7. Linden K., Niemi J. Is it possible to create a very large wordnet in 100 days? An evaluation // Lang Resources & Evaluation. – 2014. – V. 48, No 2. – P. 191–201.
8. Галиева А.М. Глаголы звучания в татарском языке: семантика и диатеза // Проблемы филологии народов Поволжья: Материалы Всерос. науч.-практ. конф. (11–13 апр. 2013 г.). – М.: Экон-информ, 2013. – Вып. 7. – С. 57–62.

9. *Vossen P., Bloksma L., Rodriguez H., Climent S., Calzolari N., Roventini A., Bertagna F., Alonge A., Peters W.* The EuroWordNet Base Concepts and Top Ontology. Version 2. – 1998. – 50 p.

Поступила в редакцию
05.06.14

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

E-mail: *amgalieva@gmail.com*

Невзорова Ольга Авенировна – кандидат технических наук, заместитель директора по науке, НИИ «Прикладная семиотика» АН РТ; доцент кафедры «Информационные системы», Казанский (Приволжский) федеральный университет, г. Казань, Россия.

E-mail: *onevzoro@gmail.com*