

УДК 811.512.145+81'37+81'367.625

ПРЕДСТАВЛЕНИЕ ГЛАГОЛОВ ФИЗИОЛОГИЧЕСКОГО ДЕЙСТВИЯ И СОСТОЯНИЯ В ЛЕКСИКОГРАФИЧЕСКОЙ БАЗЕ ДАНЫХ ТАТАРСКИХ ГЛАГОЛОВ

А.М. Галиева

Академия наук Республики Татарстан, г. Казань, 420111, Россия

Аннотация

В статье представлен опыт отображения татарских глаголов в лексикографической базе данных семантических классов, разрабатываемой в Институте прикладной семиотики Академии наук Республики Татарстан. Описаны общие принципы представления в ней глаголов, характер выделяемых семантических группировок, которые затем иллюстрируются примерами подклассов глаголов, чьей сферой денотации является физиология. Разработанная классификационная схема основана на параметрическом принципе и выполняется с учетом набора семантических, морфологических, деривационных и синтаксических признаков, релевантных грамматической и лексико-семантической системе татарского языка, что позволяет выделять подклассы внутри тематических классов, которые сами по себе являются конгломератами слов с разными структурно-семантическими свойствами. Предлагаемый комплексный подход позволяет выделять более частные и однородные семантические группировки внутри больших тематических классов.

Глаголы физиологического действия и состояния в татарском языке представляют собой обширный тематический класс единиц с разными семантическими, деривационными и синтаксическими характеристиками; внутри класса в настоящее время нами выделяются 28 подклассов, каждый из которых соответствует определенному набору признаков, в статье описано 4 таких подкласса.

Ключевые слова: татарский язык, глагол физиологического состояния, глагол физиологического действия, семантика, лексикографическая база данных, семантические классы лексики

Введение

Разработка лексикографических баз данных (БД) – машиночитаемых словарных массивов – для языков народов Российской Федерации является актуальной задачей, от решения которой во многом зависит успешность реализации проектов по автоматическому анализу текстов, машинному переводу, распознаванию речи и т. п.

В статье анализируется опыт представления татарских глаголов в лексикографической БД на примере глаголов физиологического состояния и действия. Глаголы данного тематического класса являются обширной гетерогенной группировкой единиц с разными свойствами, характеризующаясь заметным разнообразием значений, словообразовательных структур, синтаксического поведения и актантной деривации. Поэтому указанные глаголы – это хорошая иллюстрация

того, как семантика, морфологические особенности и сочетаемость татарских глаголов фиксируются в разрабатываемой БД.

При классификации лексики и создании БД в качестве источника языкового материала послужили данные имеющихся словарей татарского языка, в частности: Толкового словаря татарского языка под редакцией Ф.А. Ганиева (ТТАС), Татарско-русского словаря (ТРС), а также текстовые коллекции татарских корпусов: Татарского национального корпуса «Туган тел» (ТНК) и Письменного корпуса татарского языка (ПКТЯ). Перевод примеров на русский язык выполнен автором статьи.

Глаголы в лексикографических базах данных

Глагол принадлежит к числу наиболее сложных, семантически емких и грамматически содержательных частей речи, обладающих разветвленной системой значений и грамматических форм. «Семантика глагольного слова – не элементарна, а комплексна в том смысле, что она отображает не законченное, полное понятие о классе предметов, как это имеет место в предметных именах, а минимальные дискретные “кусочки действительности”, приближающиеся к элементарным ситуациям и событиям» [1, с. 137–138].

Для описания семантики глагольных лексем лингвисты предлагают различные форматы представления значения и различные классификационные схемы. Наиболее часто в литературе упоминается классификация английских глаголов Б. Левин, основанная на сходстве значений и синтаксических свойств глагольных лексем [2]. Семантические классы, которые были выделены этим исследователем, легли в основу многих классификаций и лексикографических БД, в том числе проекта VerbNet¹. VerbNet – иерархический тезаурус, не привязанный к определенной предметной области, в котором каждому классу дано синтаксическое описание, отражающее возможные поверхностные реализации структуры аргумента в типах конструкций, включающих переходные и непереходные глаголы, сочетания глагола с предлогами, набор залоговых преобразований [3].

Еще один известный ресурс, созданный для представления глагольной семантики, – PropBank², который позиционируется разработчиками как «банк пропозиций» английских глаголов. PropBank представляет собой корпус с ручной разметкой аргументов и семантических ролей, независимый от предметной области (события, состояния дел или действия, выражаемые существительными, в данном ресурсе не аннотируются) [4].

Глаголы фиксируются в лингвистических ресурсах разной структуры и назначения. Тезаурус WordNet³ – это иерархическая сеть лексикализованных понятий; его основными структурными единицами являются синонимические ряды – синсеты [5, 6]. Синонимические отношения в WordNet устанавливаются не между словами, а между их отдельными значениями. Разным частям речи приписываются различные типы отношений. Семантические отношения между глагольными синсетами определяются следующим образом:

¹ <https://verbs.colorado.edu/~mpalmer/projects/verbnet.html/>.

² <http://propbank.github.io/>.

³ <https://wordnet.princeton.edu/>.

- отношения следования (entailment): *идти – шагать*;
- отношения тропонимии: *сказать – шептать*;
- каузативные отношения: *есть – кормить*.

Проект FrameNet⁴ представляет собой практическую реализацию семантики фреймов – лингвистической концепции, предложенной Ч. Филлмором [7–9], в этом ресурсе также значительное место занимают глагольные лексемы.

Корпусный словарь глагольных конструкций FrameBank – русский фреймнет-ориентированный ресурс, спроектированный с учетом «традиций отечественной лексической семантики и специфики русского языка, где информация о предложно-падежной реализации управления предикатов и поверхностно-синтаксических свойствах других конструкций имеет особую ценность»⁵. Он не является калькой англоязычного FrameNet: если центральный элемент в проекте FrameNet образуют фреймы – типовые ситуации с известными участниками и расписанными ролями, то русский FrameBank создается вокруг конструкций конкретных лексем [10]. Разработчики англоязычного ресурса основываются на идее, что сеть фреймов универсальна для всех языков, в то время как русский FrameBank описывает:

- структуру лексико-семантических групп и полисемии в русском языке;
- парадигматические отношения между значениями многозначных слов (на материале конструкций);
- лексико-семантические ограничения на слоты конструкций;
- грамматические особенности русского языка (порядок слов, особенности использования падежей, согласования и т. п.) [10].

Ядро проекта FrameBank образуют 2200 ключевых русских глаголов, а также ассоциированных с ними конструкций и корпусных примеров; каждый глагол отнесен к определенному семантическому классу. Все конструкции представлены как шаблоны, в которых указаны: а) морфосинтаксические характеристики элементов конкретной конструкции, б) синтаксический ранг участника, в) экспликация (роль) участника, г) основные семантические ограничения на заполнение слота [10, с. 328].

Таким образом, в последние десятилетия были разработаны и продолжают разрабатываться специализированные лексикографические ресурсы, которые дают возможность получить нужную информацию о семантике и распределении слов различных частей речи, моделях управления лексемы, вариативности лексических единиц тех или иных классов, предоставляют иллюстративный материал из реальных текстов и т. п. Поскольку семантическая структура глагола выступает сложным конгломератом онтологических и реляционных компонентов значения, которые могут получать свое формальное выражение на разных уровнях языковой структуры, при фиксации глаголов в электронных лексикографических ресурсах разработчики исходят из различных методологических установок и отдают приоритет различным аспектам семантической организации глагольной лексики, чем объясняется значительное разнообразие структурных и содержательных характеристик имеющихся глагольных БД, разработанных для материала разных языков [11].

⁴ <https://framenet.icsi.berkeley.edu/>.

⁵ <http://framebank.ru/>.

Параметрический принцип представления глагольной лексики в базе данных татарских глаголов

Разработанная нами классификационная сетка основана на параметрическом принципе и выполняется с учетом набора семантических, морфологических, деривационных и синтаксических признаков, обусловленных грамматической и лексико-семантической системой татарского языка. Подобный комплексный подход позволяет выделять более частные семантические группировки внутри тематических классов.

В основу классификации положены следующие признаки:

- тематический (онтологический) класс, обусловленный сферой денотации глагола;
- словообразовательные модели глагольных лексем внутри тематического класса/подкласса;
- возможности актантной деривации глаголов внутри подкласса;
- актантная структура глаголов, особенности глагольного управления и тематические роли актантов [12].

Такой интегрированный грамматико-синтаксический метод определения семантических классов соответствует основным тенденциям в мировой науке при описании семантики глагола и является попыткой создать своеобразный аналог проектов VerbNet и FrameBank на материале татарского языка.

Описывая значения глаголов, мы выделяем семантические теги двух типов:

- 1) тематические теги, определяющие лексико-семантический класс глагола;
- 2) строевые компоненты значения, не привязанные к конкретному тематическому классу, но фиксирующие регулярные семантические признаки глаголов [13].

При разграничении тематических классов нами использован как древесный, так и фасетный принцип классификации, что позволяет по мере необходимости снабжать глагол одновременно несколькими тегами.

В БД фиксируются как базовые преимущественно формы, не осложненные залоговыми аффиксами, возможность образования залоговых дериватов отмечается отдельно. Глаголы даются в форме императива 2-го лица единственного числа, то есть в грамматической форме, не осложненной аффиксами имени действия или инфинитива и представляющей собой основу глагола, которая служит базой для образования финитных и нефинитных форм. Такой способ представления является оптимальным с учетом предполагаемого использования БД для компьютерных приложений.

В результате проведенной классификации, как правило, в один подкласс попадают глаголы сходной семантики: синонимы, антонимы, а также часто слова, относящиеся к одному и тому же гиперониму (такой гипероним может оформляться лишь в качестве условной концептуальной структуры, не будучи вербализованным в языке). При этом лексические синонимы, характеризуемые неодинаковой синтаксической валентностью, включаются нами в разные подклассы. Так, глаголы питания *азыклан* 'питаться, кормиться' и *ризыклан* 'питаться, кормиться' включены в один подкласс, а *сыйлан* 'угощаться' – в другой, так как последний реализует другую структуру управления аргументами.

Анализ большого массива глагольной лексики свидетельствует о том, что между семантикой и словообразовательной структурой татарских глаголов

наблюдается значительная корреляция, которая носит системный характер и имеет свою специфику внутри отдельных семантических классов и подклассов. Так, отсубстантивные дериваты, образованные при помощи аффикса *-ла/-лэ* обычно являются переходными и каузативными, как, например, глаголы с инструментальной семантикой (пример 1):

(1) *бораула* ‘сверлить, бурить’ от *борау* ‘сверло’; *тырмала* ‘грести, боронить’ от *тырма* ‘грабли, борона’; *пычакла* ‘резать’ от *пычак* ‘нож’.

Но имеются отдельные классы (подклассы), где дериваты с аффиксом *-ла/-лэ* являются непереходными, например глаголы, обозначающие природные явления (пример 2), или выделяемый нами подкласс глаголов физиологического действия с общим значением ‘производить на свет детеныша, родить’ (пример 3):

(2) *буранла* ‘мести, выюжить’ от *буран* ‘метель, выюга’; *давылла* ‘бушевать (о выюге)’ от *давыл* ‘буря’; *болытла* ‘покрываться облаками, становиться пасмурным’ от *болыт* ‘облако, туча’;

(3) *балала* ‘родить детеныша (о животных)’ от *бала* ‘детеныш’; *бэбилэ* ‘родить ребенка (о женщине)’ от *бэби* ‘младенец, грудной ребенок’; *бозаула* ‘телиться’ от *бозау* ‘теленок’; *бэрэнлэ* ‘ягниться’ от *бэрэн* ‘ягненок, барашек’.

В БД подкласс с инструментальными глаголами (пример 1) выделяется внутри глаголов трудовой деятельности. Лексемы, аналогичные экземплярам из списка в примере 2, составляют основу класса глаголов природных явлений. Глаголы, представленные в примере 3, выделяются в качестве самостоятельного подкласса глаголов физиологического действия и состояния.

Словообразовательная модель (тип основы, глаголообразующий формант, словообразовательное значение) во многом предсказывает актантную деривацию и синтаксическое поведение глагола. Так, глаголы из примера 1 могут образовать формы пассивного залога (примеры 4, 5), а глаголы из примеров 2 и 3 не образуют:

(4) *Алар беренче тапкыр бораулана торган скважиналарда гына түгел, э иске фондны торгызганда да файдаланыла* (Татар-информ – ТНК). ‘Они используются не только на вновь пробуренных скважинах, но и при восстановлении старого фонда’.

(5) *Мин килгэндэ, тырмаланмаган буразна эчендэ Нургали йоклап ята иде – уяттым* (М. Магдеев – ТНК). ‘Когда я пришел, на не выровненной бороной борозде спал Нургали – я разбудил его’.

Семантика глаголов физиологического действия и состояния

Под глаголами физиологического состояния и действия нами подразумевается обширный тематический класс глаголов, областью денотации которых является сфера тела – особенности функционирования и регуляции биологических систем человека и животных. Глаголы физиологического действия и состояния обозначают важнейшие жизненные процессы: в названный семантический класс мы включаем глаголы, обозначающие процессы дыхания, питания, выделения, а также состояния сна, бодрствования, болезни, усталости и т. п.

Глаголы физиологического действия и состояния традиционно выделяются при семантической классификации русских глаголов. Так, в «Словаре лексико-семантических групп русских глаголов» под редакцией Т.В. Матвеевой содержатся подполе глаголов физиологического действия (включает подклассы гла-

голов питания, питья, дыхания, а также остаточный подкласс глаголов проявления различных физиологических действий) и подполе глаголов физиологического состояния (включает подклассы глаголов собственно физиологического состояния и приведения в физиологическое состояние, болезни) [14].

В системе семантической разметки Национального корпуса русского языка выделяются глаголы физиологической сферы, которые обозначаются специальным тегом *t:physiol*⁶, этот тег мы позаимствовали для обозначения аналогичных татарских глаголов в БД.

Глаголам физиологической сферы посвящен ряд исследований [15–18]. В работах на материале тюркских языков глаголы физиологического состояния и действия также представлены [19, 20], хотя и не изучались подробно. В диссертации Р.К. Иштановой, посвященной семантике татарских глаголов, глаголы физиологического действия и состояния как отдельный класс не рассматриваются, будучи растворенными в других семантических группировках; в частности, исследователь выделяет классы глаголов принятия пищи, глаголов физиологических изменений живых существ, а также класс глаголов, выражающих звуки, издаваемые при физиологических процессах [21].

Ввиду принципиальной антропоцентричности языка основная масса глаголов физиологической сферы связана с особенностями жизнедеятельности человека и содержит компоненты значения, обусловленные социокультурными факторами. Соответственно, физиологический компонент значения оказывается часто нагружен большим количеством самых разных дополнительных сем и культурных коннотаций. При этом имеются глаголы, в прямом значении называющие физиологические проявления у животных, например: *утла* ‘пасться, щипая траву (о травоядных животных)’, *кушә* ‘жевать жвачку (о жвачных животных)’, *чуку* ‘клевать (о птицах)’.

Фактором, осложняющим описание, является смежность многих глаголов рассматриваемого тематического класса с глаголами других классов. Так, ряд татарских глаголов физиологического действия и состояния являются смежными с глаголами восприятия (*акай* ‘выпучиться, вытарашиться, выкатиться (о глазах)’, *сукырай* ‘ослепнуть’), с глаголами звучания (*гырла* ‘храпеть’, *ыңгыраш* ‘стонать’) или других семантических классов. Поэтому в нашей классификации тег *t:physiol* может быть использован в качестве основного или дополнительного в зависимости от лексического наполнения выделяемых подклассов.

Подклассы глаголов физиологического действия и состояния в базе данных

На текущий момент нами выделено 28 подклассов внутри класса глаголов физиологического состояния и действия. Далее представим их более подробно на примере единиц четырех наиболее характерных подклассов.

Глаголы, обозначающие ощущения в какой-либо части тела. Первый подкласс включает в свой состав 9 базовых глаголов с общим значением ‘испытывать определенные ощущения в какой-либо части тела или органе’ (табл. 1).

⁶ <http://ruscorpora.ru/corpora-sem.html/>.

Табл. 1

Глаголы, обозначающие ощущения в какой-либо части тела

Глаголы	Русский перевод (основное значение)	Тематические теги
<i>авырт</i>	‘болеть’ (о боли)	t:physiol, t:perc
<i>сызла</i>	‘сильно болеть, ныть’	t:physiol, t:perc
<i>эрне</i>	‘болеть щемящей болью, ныть’	t:physiol, t:perc
<i>ачыт</i>	‘болеть саднящей, обжигающей болью’	t:physiol, t:perc
<i>кычыт</i>	‘чесаться, зудеть’	t:physiol, t:perc
<i>кызыш</i>	‘испытывать чувство лихорадки’	t:physiol, t:perc
<i>кымыржы</i>	‘свербеть, зудеть, першить’	t:physiol, t:perc
<i>чымырда</i>	‘ощущать мурашки на теле’	t:physiol, t:perc
<i>чемердэ</i>	‘ощущать мурашки на теле’	t:physiol, t:perc

В один подкласс данные глаголы отнесены на основании ряда общих признаков:

1) базовое значение ‘испытывать определенные ощущения в какой-либо части тела или органе’; для большей части глаголов это определенные болевые ощущения, поэтому перечисленные глаголы имеют общий компонент значения ‘испытывать ощущения’, что сближает их с глаголами восприятия, чем обусловлен дополнительный тег t:perc, который является основным для глаголов восприятия;

2) все глаголы являются непереходными и выражают состояние;

3) все глаголы имеют каузативные дериваты типа *авырттыр* ‘каузировать боль’, *сызлат* ‘каузировать сильную боль’, *эрнет* ‘каузировать щемящую боль’, *кычыттыр* ‘каузировать зуд’ и не имеют форм взаимно-совместного и пассивного залогов;

4) стандартным синтаксическим субъектом при всех этих глаголах являются слова, называющие части тела: *Баш авырта*. ‘Голова болит’; *Теш сызлый*. ‘Зуб болит’; *Кулым кычыта*. ‘У меня чешется рука’.

Кроме того, многие глаголы из указанного подкласса имеют переносные значения, связанные с психологическим состоянием:

Һәр икесең йөрәге үзенчә сызлый (Г. Апсалимов – ТНК). ‘У обоих болит сердце, у каждого по-своему’.

Бигрәк сагынам, бәгырьләрем әрни шул балалар өчен (Л. Кашфиева – ТНК). ‘Очень скучаю, душа болит за этих детей’.

Аның белән бәхәсләшәсе, әйткәләшеп аласы килә, тел кычыта башлый (Р. Миннуллин – ТНК). ‘Хочется с ним поспорить, пререкаться, язык начинает чесаться’.

Глаголы, обозначающие состояние болезни. В отдельный подкласс нами сгруппированы глаголы, имеющие базовое значение ‘быть больным, переносить какое-либо заболевание’ (табл. 2).

В один подкласс эти глаголы включены на основании следующих признаков:

1) глаголы имеют общее значение ‘быть больным, переносить какое-либо заболевание’;

2) глаголы являются непереходными;

3) узуально глаголы не имеют форм пассива, рефлексива, взаимно-совместного залога;

Табл. 2

Глаголы, обозначающие состояние болезни

Глаголы	Русский перевод (основное значение)	Тематические теги
<i>авыр</i>	‘болеть, хворать’	t:physiol:disease
<i>чирлэ</i>	‘болеть, хворать’	t:physiol:disease
<i>сырхаула</i>	‘болеть, хворать’	t:physiol:disease
<i>хастала</i>	‘болеть, хворать’	t:physiol:disease

4) синтаксическим субъектом при данных глаголах являются слова, называющие живых существ.

Часть глаголов второй группы представляют собой дериваты, образованные от слов, обозначающих болезнь или больного:

чирлэ – от *чир* ‘болезнь’;

сырхаула – от *сырхау* ‘больной’;

хастала – от *хаста* ‘больной’.

Глагол *авыр(-у)* образован от прилагательного *авыр* ‘тяжелый’, который может обозначать состояние больного человека.

Глаголы *авыр*, *чирлэ*, *сырхаула*, *хастала* часто управляют называющими недугом существительными, которые вводятся при помощи послелога *белэн* (значение средства):

16 яшькэ кадарге 878 бала церебраль паралич белэн авырый (Р. Миннуллин – ТНК). ‘878 детей до шестнадцати лет болеют церебральным параличом’.

Йөз төрле чир белэн хасталанып яткан кешене терелтү өчен аның бары тик бер генэ авыруын дэвалау житәме? (М. Юнус – ПКТЯ). ‘Чтобы вылечить человека, болеющего сотней разных болезней, достаточно ли вылечить лишь одну его болезнь?’

Русия инде күптәннән кан агучанлык – гемофилия авыруы белэн сырхалый (З. Мансуров – ТНК). ‘Россия давно уже больна болезнью кровоизлияния – гемофилией’.

Глагол *самсыра* ‘прихварывать’, несмотря на семантическую близость с глаголами из табл. 2, выведен за пределы этого подкласса, так как не реализует валентность с послелогом *белэн*, определяющую для глаголов подкласса.

Глаголы, обозначающие характерные движения части лица человека или морды животного. Данные глаголы используются при описании внешности и поведения человека и животных; они позволяют передать соответствующее выражение человеческого лица или морды животного и являются важным средством для передачи эмоционального состояния через внешние проявления. Перечень выявленных в ходе классификации базовых глаголов третьей группы представлен в табл. 3.

В один подкласс перечисленные глаголы объединены по следующим признакам:

1) глаголы имеют общее значение ‘совершать характерные движения той или иной частью лица/морды’;

2) глаголы являются непереходными;

3) синтаксическим субъектом при данных глаголах являются слова, обозначающие части тела (например: глаза, рот);

Табл. 3

Глаголы, обозначающие характерные движения частей лица человека или морды животного

Глаголы	Русский перевод (основное значение)	Тематические теги
<i>акай</i>	‘выпучиться, вытаращиться, выкатиться (о глазах)’	t:physiol
<i>тасырай</i>	‘вытаращиться, выпучиться (о глазах)’	t:physiol
<i>тожрай</i>	‘выпучиться, вытаращиться, выкатиться (о глазах)’	t:physiol
<i>чекерэй</i>	‘вытаращиться, выкатываться (о глазах)’	t:physiol
<i>ыржай</i>	‘скалиться, щериться’	t:physiol
<i>челэй</i>	‘таращиться, выпучиться (о глазах)’	t:physiol
<i>чапылда</i>	‘чавкать (о губах и рте)’	t:physiol, t:sound
<i>турсай</i>	‘надуться (о губах и рте)’	t:physiol

4) глаголы имеют каузативные корреляты, обозначающие движение субъектом – человеком или животным – соответствующей частью лица/морды: *кузларне акайт* ‘вытаращить, выпучить глаза’, *кузларне тасырайт* ‘вытаращить, выпучить глаза’, *тешларне ыржайт* ‘скалить, щерить зубы’, *авызны чапылда* ‘чавкать ртом’; соответственно, данные каузативные глаголы в качестве прямого объекта управляют словами, обозначающими подразумеваемые части тела.

Поскольку многие глаголы настоящей группы часто не имеют в русском языке прямых эквивалентов с такой же актантной структурой, остановимся на них более подробно.

Базовые татарские глаголы в качестве синтаксических (и логических) субъектов имеют слова, обозначающие части лица; так, например, предполагается, что глаза вытаращиваются непроизвольно, без участия их обладателя (в русском языке используются рефлексивные дериваты):

Гөлнараның күзе акайды (С. Зыялы – ТНК). ‘У Гульнары вытаращились глаза’.

Альбертның ачудан күзләре чекерәйде (З Махмуди – ТНК). ‘У Альберта от злости выкатились глаза’.

Аяклары языла башлады, башы артка китте, иреннәре әкрән генә ачылып, тешләр ыржайды, ике теш казнасы телен кысты (Р. Батулла – ТНК). ‘Ноги разошлись, голова откинулась назад, слегка приоткрылись губы и ощерились зубы, челюстями прижало зубы’.

Ряд глаголов данной группы также могут употребляться при субъектах, обозначающих человека или животное:

Варис, бу җавап белән килешмичә, нидер әйтергә иткән иде, Вил Садриевич аңа башын борып бер акайды, теге шунда ук тынды (Х. Камалов – ТНК). ‘Варис, не соглашаясь с ответом, что-то хотел сказать, но Виль Садриевич, повернув голову, так на него выкатил глаза, что тот сразу замолчал’.

Терекөмеш йоткан кебек тожрап-биеп торган Динис: «Барысы да аңлашылды», – дип, түбәнчелек белән баш иде (М. Хасанов – ТНК). ‘Динис, который уставился, пританцовывая, словно проглотил живого серебра, смиренно произнес: «Всё понятно»’.

Кинәт ул ыржайды, янаулы ырады, якын килгән этләргә тешләтмәскә, үзе аларны тешләргә чамалый башлады (Р Батулла – ПКТЯ). ‘Вдруг он оскалился,

грозно зарычал, прикидывая, как не дать покусать приближающимся собакам, а покусать их самих’.

Примечательно, что пять из восьми глаголов группы передают значение ‘выпучиться, вытаращиться, выкатиться’ и в качестве субъекта имеют слова, обозначающие глаза (см. примеры выше) или части глаз:

Гөл апаның бәбәкләре акайды (Н. Гиматдинова – ТНК). ‘У тети Гуль выпучились зрачки’.

Три глагола из данного подкласса связаны с зоной рта: *ыржай* ‘скалиться, щериться’, *чапылда* ‘чавкать (о губах и рте)’, *турсай* ‘надуться (о губах и рте)’.

Глаголы этой группы указывают на разнообразные характерные и часто выразительные движения мышц лица/морды животного, что является важной формой внешнего проявления тех или иных психических или физических состояний.

Каузативные корреляты позволяют обозначать оперирование субъектом соответствующей частью лица/морды животного:

Жәмил күзләрен акайтты (З. Махмуди – ТНК). ‘Джамиль вытаращил глаза’.

Бурсык ыслап теиен ыржайтты (Н. Каштанов – ТНК). ‘Барсуку, зашипев, оскалил зубы’.

Гарьлегеннән жән ачулары купкан, жән ачулары чыккан Халәп агай, фикердәш эзләгәндәй, күзләрен тожрайтып, янә як-ягына каранып алды (М. Хасанов – ТНК). ‘Дядя Халяп, очень злой от перенесенной обиды, словно в поисках сторонника вытаращил глаза и огляделся по сторонам’.

При переводе таких глаголов на русский язык используются соответствующие русские каузативные глаголы.

Глаголы еды и питья. Татарские глаголы физиологического действия, касающиеся сферы питания человека и животных, образуют компактную, но достаточно гетерогенную группировку, внутри которой могут быть сформированы отдельные подклассы по несовпадающим тематическим или синтаксическим признакам.

Глаголы еды и питья включают часть глаголов подполя питания, характеризующихся общим значением ‘глота, поглощать пищу или жидкость’ (табл. 4).

Табл. 4

Базовые глаголы еды и питья

Глаголы	Русский перевод	Тематические теги
<i>аша</i>	‘есть, кушать’	t:physiol
<i>им</i>	‘сосать’	t:physiol
<i>кимер</i>	‘грызть’	t:physiol
<i>күшә</i>	‘жевать жвачку (о жвачных животных)’	t:physiol
<i>эч</i>	‘пить’	t:physiol
<i>чөмер</i>	‘пить, хлебать’	t:physiol
<i>тык</i>	‘жрать, лопать’	t:physiol
<i>чемче</i>	‘щипать траву’	t:physiol
<i>уртла</i>	‘пить большими глотками’	t:physiol
<i>утла</i>	‘пасть, щипая траву’	t:physiol
<i>йотымла</i>	‘прихлебывать, пить глотками’	t:physiol

В один подкласс данные глаголы входят по ряду признаков:

- 1) все глаголы имеют базовую сему 'есть' или 'пить';
- 2) все глаголы являются переходными и выражают действие;
- 3) все глаголы имеют каузативные дериваты типа *ашат* 'кормить', *имез* 'кормить грудью', *эчер* 'поить', *утлат* 'пасти, давая возможность щипать траву' и не имеют форм взаимно-совместного залога;
- 4) стандартным синтаксическим субъектом при этих глаголах являются слова, обозначающие живых существ: *Бала имэ*. 'Ребенок сосет грудь'; *Атлар утлап йөри*. 'Лошади щиплют траву'.

Четвертый класс включает глаголы, обозначающие процесс еды и питья как у человека и животных (например: *аша*, *эч*, *кимер*), так и только у животных (*утла*, *кушэ*).

В качестве прямого объекта при глаголах еды и питья выступают слова, обозначающие пищу или жидкости:

Касэсендэге салкын чайне уртлады (А. Халим – ТНК). 'Пил большими глотками холодный чай из пиалы'.

Минем ашыгыр эфирем юк, сыраны эз-мэз генэ йотымлап, шуларга карап утырам (А. Еники – ТНК). 'Мне спешить некуда, сажу смотрю на них, прихлебывая пиво'.

Шулчак үлэн утлап йөргән бер фил, күрмичә, ялгыш аның өстенә баскан (Р. Валиев – ТНК). 'И тогда слон, щиплющий, траву, наступил на него'.

Глаголы данной группы могут использоваться в качестве непереходных:

Ичмасам, туйганчы ашый да алмадык... (Р. Галиуллин – ТНК). 'Даже не смогли наесться досыта'.

Бер көтүдә егерме-утыз сыер утлый (Р. Батулла – ТНК). 'В одном стаде пасется двадцать-тридцать коров'.

Жәйләү тулы жьылкы, куй, кәжә көтүе утлый (Р. Батулла – ПКТЯ). 'На летнем пастбище щиплют траву стада коней, овец, коз'.

Некоторые глаголы данного класса имеют рефлексивные корреляты (например: *тык – тыгын*, *чемче – чемчен*), которые соответствующим образом фиксируются в БД.

К глаголам еды и питья тесно примыкает еще один подкласс, связанный со сферой питания – непереходные глаголы, обозначающие процесс питания. Последний подкласс в текущем варианте БД включает 4 базовых глагола: *азыклан* 'питаться, кормиться', *ризыклан* 'питаться, кормиться', *авызлан* 'отведать, пробовать пищу', *туклан* 'питаться'. Приведенные глаголы образованы при помощи аффикса *-лан* и являются возвратными с неопредельным значением, все они имеют каузативные дериваты.

Два последних подкласса, взятые вместе, образуют более широкий семантический класс глаголов питания, выражая значимые физиологические аспекты, связанные с питанием и утолением чувства голода.

Смежными с глаголами питания являются также некоторые другие подклассы глаголов физиологического действия и состояния, например глаголы, обозначающие голод и жажду, физиологические проявления чрезмерного приема той или иной пищи, и некоторые другие.

Ввиду антропоцентричности языковой картины мира основная масса глаголов физиологической сферы связана с особенностями жизнедеятельности человека и содержит культурно обусловленные компоненты значения, однако имеются глаголы, называющие в прямом значении физиологические проявления у животных, главным образом имеющих хозяйственное значение.

Описанные примеры выделенных подклассов и отдельных лексических единиц дают общее представление о структуре и принципах построения и заполнения БД семантических классов татарских глаголов. Разрабатываемая БД позволяет получить список семантических классов и подклассов с их лексическим наполнением, а также содержит информацию о важнейших аспектах семантики, актантной деривации и синтагматики отдельных глаголов.

Заключение

Классификационная сетка, используемая при разработке и заполнении БД, основана на параметрическом принципе и выполняется с учетом набора семантических, морфологических, деривационных и синтаксических признаков, релевантных грамматической и лексико-семантической системе татарского языка. Регулярная морфологическая структура производных глаголов во многом определяет базовое значение конкретных дериватов, относимых к одному семантическому классу (подклассу). В будущем для более детального описания подклассов планируется ввести еще один параметр – функционирование глаголов в составе тех или иных грамматикализованных конвербных конструкций, что внесет существенные уточнения в свойства ситуаций, описываемых глагольными лексемами.

Использованный при разработке БД комплексный подход позволяет вычленять более частные и однородные семантические группировки внутри гетерогенных тематических классов. Так, в настоящее время нами выделено 28 подклассов внутри класса глаголов физиологического состояния и действия.

При определении тематических классов нами использован как древесный, так и фасетный принцип классификации, в результате глагол при необходимости можно снабжать одновременно несколькими тегами.

Описываемый проект нацелен на представление структуры и реального лексического наполнения семантических классов татарских глаголов. Лингвистическая информация, содержащаяся в БД татарских глаголов, может быть использована при разработке различных компьютерных приложений, а также может служить эмпирической базой в теоретических исследованиях для объяснения морфосинтаксического поведения слов и проведения последовательного композиционного анализа глагольных предикатов.

Источники

- ТТАС – Татар теленең аңлатмалы сүзлеге / Ред. Ф.Ә. Ганиев. – Казан: Матбугат йорты, 2005. – 848 б.
- ТРС – Татарско-русский словарь / Под ред. Ф.А. Ганиева. – Казань: Тат. кн. изд-во, 2004. – 488 с.

ТНК – Татарский национальный корпус «Туган тел». – URL: <http://tugantel.tatar/>, свободный.

ПКТЯ – Письменный корпус татарского языка. – URL: <http://www.corpus.tatar/>, свободный.

Литература

1. *Уфимцева А.А.* Лексическое значение: принцип семиологического описания лексики. – М.: Наука, 1986. – 239 с.
2. *Levin B.* English Verb Classes and Alternations: A Preliminary Investigation. – Chicago: Univ. of Chicago, 1993. – 348 p.
3. *Kipper K., Korhonen A., Ryant N., Palmer M.* Extending VerbNet with novel verb classes // Proc. 5th Int. Conf. on Language Resources and Evaluation. – URL: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.380.5541&rep=rep1&type=pdf>, свободный.
4. *Palmer M., Kingsbury P., Gildea D.* The Proposition Bank: An annotated corpus of semantic role // Comput. Linguist. – 2005. – V. 31, No 1. – P. 71–106. – doi: 10.1162/0891201053630264.
5. *Fellbaum C.* WordNet. An Electronic Lexical Database. – Cambridge: Mass: MIT Press, 1998. – 423 p.
6. *Mille G.A.* WordNet: A lexical database for English // Commun. ACM. – 1995. – V. 38, No 11. – P. 39–41.
7. *Fillmore C.* Frame semantics and the nature of language // Ann. N. Y. Acad. Sci. – 1976. – V. 280, No 1. – P. 20–32. – doi: 10.1111/j.1749-6632.1976.tb25467.x.
8. *Baker C.F., Fillmore C.J., Lowe J.B.* The Berkeley FrameNet Project // Proc. 17th Int. Conf. on Computational linguistics. – 1998. – V. 1: Association for Computational Linguistics. – P. 86–90.
9. *Fillmore C.J., Baker C.F., Sato H.* FrameNet as a “net” // Proc. LREC. – Lisbon, 2004. – V. 4. – P. 1091–1094.
10. *Кашкин Е.В., Ляшевская О.Н.* Семантические роли и сеть конструкций в системе FrameBank // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной междунар. конф. «Диалог», 2013 г. – М.: Изд-во РГГУ, 2013. – Вып. 12. – С. 325–343.
11. *Галиева А.М.* Глагольная лексика в лексикографических базах данных: обзор основных ресурсов // Электронные библиотеки. – 2016. – Т. 19, № 2. – С. 80–97.
12. *Galieva A., Vavilova Z., Gatiatullin A.* Semantic classification of Tatar verbs: Selecting relevant parameters // Proc. XVIII EURALEX Int. Congress: Lexicography in Global Contexts. – Ljubljana: Ljubljana Univ. Press, 2018. – P. 811–818.
13. *Galieva A., Nevzorova O.* Semantic annotation of verbs for the Tatar corpus // Proc. XVII EURALEX Int. Congress: Lexicography and Linguistic Diversity. – Tbilisi: Tbilisi State Univ., 2016. – P. 340–347.
14. Лексико-семантические группы русских глаголов: Учеб. словарь-справочник / Под общ. ред. Т.В. Матвеевой. – Свердловск: Изд-во Урал. ун-та, 1988. – 153 с.
15. *Магфурова С.О.* Лексико-грамматическое функционирование глаголов со значением физиологического действия и состояния: Автореф. дис. ... канд. филол. наук. – Северодвинск, 2002. – 19 с.
16. *Нелина И.А.* Функциональная семантика глаголов физиологического состояния в устной речи диалектоносителей // Вестн. Волгогр. гос. ун-та. Сер. 2: Языкознание. – 2013. – № 2. – С. 183–187.

17. *Кабанова С.А.* Глаголы со значением болезненного состояния индивида в сербском и русском языках // Вестн. Воронеж. гос. ун-та. Сер. Лингвистика и межкультурная коммуникация. – 2012. – № 2. – С. 162–167.
18. *Лельхова Ф.М.* Глаголы физиологического состояния в хантыйском языке (на материале сынского диалекта) // Вестн. угроведения. – 2011. – № 1. – С. 30–37.
19. *Усманова М.И.* Функционально-семантическая классификация глаголов башкирского языка: Дис. ... д-ра филол. наук. – Уфа, 2002. – 415 с.
20. *Рассадин В.И.* Тофаларский язык и его место в системе тюркских языков. – Элиста: Изд-во Калмыц. ун-та, 2014. – 218 с.
21. *Иитанова Р.К.* Глаголы татарского языка в семантическом аспекте: Дис. ... канд. филол. наук. – Казань, 2002. – 284 с.

Поступила в редакцию
21.08.18

Галиева Альфия Макаримовна, кандидат философских наук, ведущий научный сотрудник
Института прикладной семиотики

Академия наук Республики Татарстан
ул. Баумана, д. 20, г. Казань, 420111, Россия
E-mail: amgalieva@gmail.com

ISSN 2541-7738 (Print)
ISSN 2500-2171 (Online)

UCHENYE ZAPISKI KAZANSKOGO UNIVERSITETA. SERIYA GUMANITARNYE NAUKI
(Proceedings of Kazan University. Humanities Series)

2018, vol. 160, no. 5, pp. 1219–1234

**Representing Tatar Verbs Denoting Physiological Actions
and States in the Lexicographic Database**

A.M. Galieva

Tatarstan Academy of Sciences, Kazan, 420111 Russia
E-mail: amgalieva@gmail.com

Received August 21, 2018

Abstract

This paper describes how Tatar verbs denoting physiological actions and states are represented in lexicographic databases. The general principles of representing verbs in the database and the basic classification parameters have been given.

The thematic (ontological) classes of verbs essentially are heterogeneous aggregations of lexemes of various semantic, structural, and syntactic features. We suggest the way to distinguish more particular and homogeneous semantic ensembles within large thematic classes. Our classification of Tatar verbs is based on the following parameters of verbs: thematic features linked with the verb's thematic class; derivational features related to the verb's derivation pattern; grammatical features associated with the valency changing operations of voice affixes (possibility of producing grammatical voice derivatives and particular meanings of voice forms); syntactic features related to the allowable predicate-argument structure and thematic roles of arguments.

A number of examples demonstrating that Tatar verbs denoting physiological actions and states form a broad thematic class of items with different semantic, derivational, syntactic characteristics have been provided; within this class we select 28 subclasses, a detailed description of the lexical content of four such subclasses has been performed. Many of the analyzed lexemes lack direct equivalents in the Russian language.

Linguistic information represented in the database has both applied and theoretical dimensions. So, it can be used to develop computer applications and as an empirical basis for explicating the morphosyntactic behavior of words and for carrying out a consistent compositional analysis of verbal predicates.

Keywords: Tatar language, verb of physiological state, verb of physiological action, semantics, lexicographic database, semantic classes of vocabulary

References

1. Ufimtseva A.A. *Leksicheskoe znachenie: printsip semiologicheskogo poisanija leksiki* [Lexical Meaning: The Principle of the Semiological Description of Vocabulary]. Moscow, Nauka, 1986. 239 p. (In Russian)
2. Levin B. *English Verb Classes and Alternations: A Preliminary Investigation*. Chicago, Univ. of Chicago, 1993. 348 p.
3. Kipper K., Korhonen A., Ryant N., Palmer M. Extending VerbNet with novel verb classes. *Proc. 5th Int. Conf. on Language Resources and Evaluation*. Available at: <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.380.5541&rep=rep1&type=pdf>.
4. Palmer M., Kingsbury P., Gildea D. The Proposition Bank: An annotated corpus of semantic role. *Computational Linguistics*, 2005, vol. 31, no. 1, pp. 71–106. doi: 10.1162/0891201053630264.
5. Fellbaum C. *WordNet. An Electronic Lexical Database*. Cambridge, Mass, MIT Press, 1998. 423 p.
6. Mille G.A. WordNet: A lexical database for English. *Communications of the ACM*, 1995, vol. 38, no. 11, pp. 39–41.
7. Fillmore C. Frame semantics and the nature of language. *Annals of the New York Academy of Sciences*, 1976, vol. 280, no. 1, pp. 20–32. doi: 10.1111/j.1749-6632.1976.tb25467.x.
8. Baker C.F., Fillmore C.J., Lowe J.B. The Berkeley FrameNet Project. *Proc. 17th Int. Conf. on Computational Linguistics*. Vol. 1: Association for Computational Linguistics, 1998, pp. 86–90.
9. Fillmore C.J., Baker C.F., Sato H. FrameNet as a “net”. *Proc. LREC*. Vol. 4. Lisbon, 2004, pp. 1091–1094.
10. Kashkin E.V., Lyashevskaya O.N. Semantic roles and structural network in FrameBank system. *Komp'yuternaya lingvistika i intellektual'nye tekhnologii: Po materialam ezhegodnoi mezhdunar. konf. "Dialog", 2013 g.* [Computer Linguistics and Intelligent Technologies: Proc. Annu. Int. Conf. "Dialog", 2013]. Moscow, RGGU, 2013, no. 12, pp. 325–343. (In Russian)
11. Galieva A.M. Verbal vocabulary in lexicographical data bases: Reviewing main resources. *Elektronnye Biblioteki*, 2016, vol. 19, no. 2, pp. 80–97. (In Russian)
12. Galieva A., Vavilova Z., Gatiatullin A. Semantic classification of Tatar verbs: Selecting relevant parameters. *Proc. XVIII EURALEX Int. Congress: Lexicography in Global Contexts*. Ljubljana, Ljubljana Univ. Press, 2018, pp. 811–818.
13. Galieva A., Nevzorova O. Semantic annotation of verbs for the Tatar corpus. *Proc. XVII EURALEX Int. Congress: Lexicography and Linguistic Diversity*. Tbilisi, Tbilisi State Univ., 2016, pp. 340–347.
14. *Leksiko-semanticheskie gruppy russkikh glagolov* [Lexical and Semantic Groups of Russian Verbs]. Matveeva T.V. (Ed.). Sverdlovsk, Izd. Ural. Univ., 1988. 153 p. (In Russian)
15. Magfurova S.O. The lexical and grammatical functioning of verbs with the meaning of physiological action and condition. *Extended Abstract of Cand. Philol. Sci. Diss.* Severodvinsk, 2002. 19 p. (In Russian)
16. Nelina I.A. Functional semantics of physiological state verbs in the speech of dialect speakers. *Vestnik Volgogradskogo Gosudarstvennogo Universiteta. Seriya 2: Yazykoznanie*, 2013, no. 2, pp. 183–187. (In Russian)
17. Kabanova S.A. Verbs denoting a disease state of an individual in the Serbian and Russian languages. *Vestnik Voronezhskogo Gosudarstvennogo Universiteta. Seriya Lingvistika i Mezhhul'turnaya Kommunikatsiya*, 2012, no. 2, pp. 162–167. (In Russian)
18. Lelkhova F.M. Verbs of a physiological condition in hanty language (based on synja dialect). *Vestnik Ugrovedeniya*, 2011, no. 1, pp. 30–37. (In Russian)

19. Usmanova M.I. Functional and semantic classification of verbs of the Bashkir language. *Doct. Philol. Sci. Diss.* Ufa, 2002. 415 p. (In Russian)
 20. Rassadin V.I. *Tofalarskii yazyk i ego mesto v sisteme tyurkskikh yazykov* [Tofa Language and Its Place in the System of Turkic Languages]. Elista, Izd. Kalmyts. Univ., 2014. 218 p. (In Russian)
 21. Ishtanova R.K. Verbs of the Tatar language in the semantic aspect. *Cand. Philol. Sci. Diss.* Kazan, 2002. 284 p. (In Russian)
-

⟨ **Для цитирования:** Галиева А.М. Представление глаголов физиологического действия и состояния в лексикографической базе данных татарских глаголов // Учен. зап. Казан. ун-та. Сер. Гуманит. науки. – 2018. – Т. 160, кн. 5. – С. 1219–1234. ⟩

⟨ **For citation:** Galieva A.M. Representing Tatar verbs denoting physiological actions and states in the lexicographic database. *Uchenye Zapiski Kazanskogo Universiteta. Seriya Gumanitarnye Nauki*, 2018, vol. 160, no. 5, pp. 1219–1234. (In Russian) ⟩