

УДК 811.111'282.2(73)

ОПЫТ МЕТАЯЗЫКОВОГО ОПИСАНИЯ СЕМАНТИЧЕСКИХ КОНСТАНТ РУССКОЙ ЛЕКСИКИ

А. А. Кретов, Е. Н. Подтележникова

Воронежский государственный университет

Поступила в редакцию 5 июня 2016 г.

Аннотация: в статье представлен семантический метаязык, предназначенный для компьютерной классификации лексической семантики. В ходе работы были выделены следующие семантические константы: «натурфакты», «растения», «животные», «человек», «артефакты», «ментефакты», «социум» и «сверхъестественное». При создании метаязыковых дефиниций использовался строгий синтаксис: первое место занимает семантическая константа, затем – переменные. Описываемый метаязык ориентирован скорее на компьютер, чем на человека; он предполагает обобщение и формализацию лексического значения.

Ключевые слова: семантика, метаязык, семантические константы и переменные, компьютерный анализ лексики.

Abstract: the paper presents semantic metalanguage which can be used for computer analysis of vocabulary. Approximately four thousand Russian words are divided into seven groups of constants: “naturfacts”, “plants”, “animals”, “man”, “artifacts”, “mentefacts”, “society” and “supernatural”. To make metalanguage definitions clear we use strict metalanguage syntax. The first position is usually taken by semantic constant, then variables follow: place, time, form, part, substance, etc. The metalanguage described is a special type of definitions – not for a man but for computer analysis: it presupposes generalization and formalization of lexical meanings to use metalanguage definitions for automatic text processing systems.

Key words: semantics, metalanguage, semantic constants and variables, computer analysis of words.

В самой грубой и простой версии исследование эволюции лексико-семантической системы можно представить как движение означающих слов в системе координат их означаемых.

Как известно, система означаемых лексико-семантической системы значительно более устойчива, чем система означающих, а сопоставительные исследования отдельных фрагментов лексико-семантической системы дают основание для вывода о гомологической инвариантности лексико-семантической системы человеческого Языка относительно всего множества конкретных языков.

Инвариант существует в виде множества наблюдаемых *вариантов*. Гомологическую инвариантность обеспечивает наличие системы трансформаций одного варианта в другой. Эта система трансформаций подобна наличию закономерных фонетических соответствий в компаративистике, посредством которых рефлекс одного родственного языка может быть трансформирован в рефлекс другого родственного языка, а они оба – в архетип языка-предка.

Если мы хотим создать историческую лексикологию русского языка, нам необходима система означае-

мых русского языка. Такой системой означаемых обычно предлагают считать синопсис идеографического словаря.

Тут и начинаются сложности. Словарь может строиться дедуктивно – как логическая схема, набрасываемая на внеязыковую действительность и затем соотносимая со значениями слов, или индуктивно – как интерпретация языковых значений с их пошаговым укрупнением и обобщением. При индуктивном построении тезауруса с каждым последующим шагом обобщения возрастает произвол исследователя в интерпретации материала. Поэтому, вероятно, не будет ошибкой сказать, что первичное членение синопсиса идеографического словаря всегда основано на той или иной концепции исследователя, декларируется она или нет, как, например, в тезаурусе Роже [1] или «Лексической основе русского языка» [2] (ср. [3, с. 126]).

Мы приняли за основу «древо Порфирия», основанное на категориях Аристотеля и весьма близкое к «наивной картине мира», отраженной в языке.

Теоретической базой данного исследования является положение представителей Московской семантической школы о том, что семантика конкретного естественного языка должна строиться на основе

ограниченного подъязыка самого естественного языка [4]. При этом основной единицей метаязыка мы считаем существительное как наименее маркированную часть речи [5] и лексические константы представляем в виде существительных, а не выводим из глагольных, как это делается в модели «Смысл ↔ Текст» [6]. Кроме того, набор семантических констант ближе к набору корней (корневых слов) данного языка и не имеет ничего общего с семантическими примитивами А. Вежбицкой [7; 8] (которые при ближайшем рассмотрении оказываются некоторым отдаленным и малоинформативным подобием грамматических переменных).

Наиболее близким аналогом нашей работы является опыт Ю. Н. Караулова [9], предложившего особый метаязык (на базе русского), ориентированный на компьютерный анализ и приспособленный для него.

При этом мы постарались учесть, как опыт Ю. Н. Караулова, так и основанный на четко алгоритмизированном методе словарной идентификации опыт семантической классификации лексики Э. В. Кузнецовой [10; 11] и ее последователей во главе с Л. Г. Бабенко [12–14].

В формулировке И. М. Кобозевой недостаток метода Ю. Н. Караулова состоит в «игнорировании таких особенностей лексического значения, как структурированность». Как следствие – не различаются периферийные компоненты значения и центральные компоненты, по которым и следует относить слово к семантическому классу [3, с. 134]. Причина этого – в игнорировании синтаксиса словарных толкований, посредством которого и осуществляется структурирование компонентов лексического значения на центральные и периферийные. В нашем метаязыке синтаксис четкий и жесткий.

Применение метода Э. В. Кузнецовой показало, что лексическими средствами могут обозначаться как лексические, так и грамматические значения. Соответственно, применение метода ступенчатой идентификации приводит к созданию как лексических (например, глаголов перемещения или глаголов физического воздействия на объект), так и грамматических объединений слов (например, ЛСГ глаголов начальной фазы бытия, ЛСГ глаголов существования и ЛСГ глаголов прекращения бытия) [11, с. 146–149].

Так, в частности, таксон 2.1.1.1. «Глаголы каузации начала существования» с базовым глаголом *породить* включает в себя глаголы *колоситься/выколоситься* с субъектом (*растением*) и объектом – его частью (колосом), *отрастить/отращивать* (волосы, усы, бороду, ногти – о человеке), *плодить/расплодить* (например, домашних животных – о человеке), *плодоносить* (о плодовых деревьях), *родить/рождать*

и *рожать* (разг.) – (о человеке) [11, с. 63]. Интегральными для этих слов оказываются грамматические значения: «каузировать, начать, быть», а дифференциальными – лексические: *растение* (злак или дерево), *животное, человек*, тогда как, на наш взгляд, при работе с лексической семантикой должно быть наоборот.

Именно опыт Свердловско-Екатеринбургской семантической школы привел нас к осознанию фундаментальной грамматичности глагола и лексичности существительного. Как следствие – наш метаязык ориентирован на существительные и содержит два типа значений: константы и переменные.

Лексический материал настоящего исследования получен в результате обработки русской части семи малых романско-русских словарей объемом 10±2 тысячи слов [15–21]. На момент начала исследования тот материал был обработан и доступен.

В данном случае входной язык словарей не имеет существенного значения, так как в выходной (толкующей) части русский язык выступает в качестве *метаязыка*, описывая семантику (в толковом словаре – русских слов, в двуязычных – иностранных с помощью метаслов, т.е. русских слов в метаязыковой функции).

Почему были взяты двуязычные малые словари? Необходимо было определить круг значений, составляющий семантическое ядро языка. Известно, что 10–15 тысяч слов – словарь образованного человека. Следовательно, в этом корпусе заведомо содержится семантическое ядро языка, описываемое русскими метасловами, и, как следствие, – ядро русского метаязыка.

Из словарей были удалены метаслова, встретившиеся только в одном словаре, метаслова с частотой менее 4, несубстантивные метаслова и метасловоформы, не являющиеся леммами существительных, что в итоге дало 3767 метаслов-существительных, подлежащих описанию с помощью метаязыка. Затем общее количество слов было увеличено до 3932 в результате разрешения многозначности и снятия омонимии: многозначное слово обычно рассматривается в своем первом значении, а омонимы приводятся с индексами (*крошка*¹ – ‘мельчайшая часть, крохотный кусочек чего-нибудь, преим. хлеба’; *крошка*² – ‘действие по гл. крошить’).

Классификация полученных метаслов позволила выделить 8 классов *констант*: «натурфакт», «растения», «животные», «человек», «артефакт», «ментефакт», «сверхъестественное», «социум». В метаязыковом толковании константа занимает первое место, следующие позиции занимают переменные (выделенные и описанные ранее [15]) или истолкованные ранее метаслова. Таким образом, синтаксис метаязыка строгий и жесткий.

После определения принадлежности слов к классу констант работа велась в рамках каждого класса. Как и ожидалось, легче всего формулировались метаязыковые толкования для слов конкретной лексики, относящихся к классам «натурфакты», «растения», «животные».

В табл. 1 представлены толкования отдельных **натурфактов**, удовлетворяющие основным требованиям к метаязыку. Синтаксис метаязыка предполагает заполнение первого места в дефиниции семантической константой высшего уровня, в данном случае это натурфакт. Следующее место заполняется переменной, как в случае со словами *вода* и *камень*, или уже истолкованным метасловом, например, *штиль* определяется через *ветер*, а *озеро*, *океан*, *раковина* – через *воду*.

«Расшифровать» полученные метаязыковые толкования можно следующим образом: *ветер* – движение воздуха; *штиль* – отсутствие ветра; *пена* – результат движения воды; *вода* – жидкий натурфакт; *озеро* и *океан* – виды водного пространства; *раковина* – контейнер для животного, живущего в воде; *камень* – твердый натурфакт.

На примере натурфактов удобно показать принципиальное отличие нашего подхода от подхода Ю. Н. Караулова, у которого «слово *вода* вошло в поля ВОДОЁМ, ЖИДКИЙ, ПИТЬ» [16, с. 98]. У нас *вода* толкуется как «жидкий натурфакт», разумеется, не единственный, но самый распространенный и типичный, поэтому даваемый без уточнений, обязательных у всех остальных жидких натурфактов. Поступая так, мы воспроизводим логику естественного метаязыка,

в котором у слова *вода* есть множество конкретизаторов – вплоть до «околоплодные воды» и «туалетная вода».

Водоем – «водное пространство», а *пить* – «совершать с водой действие, являющееся условием существования живых существ». Глаголы не имеют собственной лексической семантики (если не считать за таковую *бытийность*) и определяются через семантические классы, связываемые глаголом: в данном случае – *животное/человек* (в позиции «кто?») и *натурфакт (вода)* (в позиции «что?»). Например, «каузировать *воду* быть внутри (о *человеке/животном*)». *Внутри* – семантическая переменная «пространство:вид7». *Каузировать* – производная семантическая переменная «действие:существование:начало».

В табл. 2 представлены метаязыковые толкования слов, принадлежащих к классу «**животное**». *Дельфин* – это вид животного, обитающего в воде, *жаба* обитает и в воде, и на суше. *Свинья*, *поросенок*, *боров* и *кошка* – это животные, которые используются человеком в определенных целях, причем *свинья* называет вид животных и поэтому никак не маркируется (по аналогии с *дельфином* и *жабой*), *поросенок* – это детеныш свиньи, а *боров* – самец свиньи. *Кошка* также используется человеком, но не как сельскохозяйственное, а как домашнее животное.

При толковании **растений** выяснилось, что большинство из них используется в качестве пищи (*кукуруза* – растение:вид:зерно:пища) или воспринимается человеком в первую очередь как цветок (*роза* – растение:вид:цветок).

Т а б л и ц а 1

Метаязыковое толкование натурфактов

ВЕТЕР	натурфакт	воздух	движение	
ШТИЛЬ	натурфакт	ветер	нет	
ПЕНА	натурфакт	вода	движение	результат
ВОДА	натурфакт	жидкое		
ОЗЕРО	натурфакт	жидкое	пространство	мало
ОКЕАН	натурфакт	жидкое	пространство	много
РАКОВИНА	натурфакт	контейнер	животное	вода
КАМЕНЬ	натурфакт	твердое		

Т а б л и ц а 2

Метаязыковое толкование слов, принадлежащих к классу «животное»

ДЕЛЬФИН	животное	вид	вода			
ЖАБА	животное	вид	вода	земля		
СВИНЬЯ	животное	вид	земля	человек	функция	
ПОРОСЁНОК	животное	вид	земля	возраст	человек	функция
КОШКА	животное	вид	земля	человек	функция	отношение

Перейдем к описанию наиболее обширного класса «человек» (табл. 3). Как всегда, начнем с самого простого и очевидного – это тело человека. Воспользуемся общепризнанным делением тела человека на три части – голову, туловище и ноги. Тогда *голова* – верхняя часть тела, *туловище* – нижняя, а *ноги* мы описываем как ‘конечность:низ:пара’, чтобы противопоставить их *рукам* как верхним конечностям.

Лицо – это часть головы, находящаяся снаружи впереди; его части – *глаза*, *губы*, *нос*. Внутри головы находится *мозг* (вверху) и *ротовая полость* (внизу); *череп* – контейнер для мозга. Аналогично «разбираются» на составляющие туловище и ноги. *Кровь*, а также *пот* и *слюна* описываются как жидкие субстанции; *волосы* – через переменную *много*, а *кожа* – через метаслово *покров*.

Кроме частей тела к классу «человек» относятся его свойства (*миловидность*) и состояния (*тошнота*, *отдых*).

Из группы *артефактов* наиболее проработанной является часть слов, обозначающих одежду (табл. 4). Здесь используется метаслово *контейнер*, потому что любой предмет одежды облакает или охватывает определенную часть тела, даже, как ни странно, *шарф* и *ремень*. Исключение составляют различного вида *застежки*, которые не являются контейнерами, а

представляют собой части одежды с функцией скрепления.

Для толкования таких слов, как *куртка*, *платье*, *брюки*, используется понятие *талии*, разделяющей тело человека на верхнюю и нижнюю часть. Тогда ‘туловище:верх’ – это всё, что выше талии, а ‘туловище:низ’ – всё, что ниже талии. Исходя из этого, *брюки* трактуются как контейнер для нижней части туловища и ног, а *платье* – как предмет одежды, облакающий все туловище и ноги. Для толкования слов, описывающих верхнюю и нижнюю одежду, вводятся соответствующие метаслова (см. *трусы*, *пальто*).

Еще один выделенный нами класс семантических констант относится к классу «*сверхъестественное*», куда входят такие единицы, как *крест* – ‘сверхъестественное:артефакт:форма’, *бог* – ‘сверхъестественное:вид’ и др.

В ходе работы возникла необходимость создания нового класса семантических констант – «*социум*». Здесь можно выделить несколько групп: социально значимые действия (*казнь*, *клятва*, *надзор*, *восстание*); места, где оказываются социальные услуги (*аптека*, *больница*, *школа*, *метро*); общественные объединения и образования (*государство*, *буржуазия*) и т.д.

Наиболее сложным является класс семантических констант «*ментефакты*». Нам предстояло выяснить,

Т а б л и ц а 3

Метаязыковое толкование слов, принадлежащих к классу «человек»

ТЕЛО	человек	форма			
ГОЛОВА	человек	форма	часть	верх	
ТУЛОВИЩЕ	человек	форма	часть	середина	
НОГА	человек	форма	конечность	низ	пара
РУКА	человек	форма	конечность	верх	пара
ЛИЦО	голова	часть	снаружи	часть	перед
ГЛАЗ	лицо	часть	верх	пара	
ГУБА	лицо	часть	низ	пара	
МОЗГ	голова	часть	внутри	часть	верх
РОТ	голова	часть	внутри	часть	низ
ЧЕРЕП	голова	часть	контейнер	мозг	

Т а б л и ц а 4

Метаязыковое толкование артефактов

ШАРФ	артефакт	контейнер	шея	форма	
ШАПКА	артефакт	контейнер	голова		
БРЮКИ	артефакт	контейнер	туловище	низ	ноги
ПЛАТЬЕ	артефакт	контейнер	туловище	ноги	
ТРУСЫ	артефакт	контейнер	туловище	низ	нижнее
ПАЛЬТО	артефакт	контейнер	туловище	ноги	верхнее

какие из 3767 слов относятся к этому классу. На начальном этапе работы в группу ментефактов попали все слова, не имеющие отношения к другим, описанным выше классам: *интрига, клятва, спор, затруднение, предел, итог, жребий, образец* и др. После создания класса «социум» многие из этих слов были истолкованы как социальные действия, например, *заявление* – ‘социум:действие:текст’; *издевательство* – ‘социум:действие:отношение:плохо’; *ласка* ‘социум:действие:отношение:хорошо’. В результате остались «чистые» ментефакты: *вероятность, затруднение, замысел, мнение, загадка, двусмысленность, энэт, миф* и т.д.

Создание метаязыковых толкований ментефактов вызвало серьезные затруднения, и было решено обратиться к метаязыку словарей для того, чтобы проанализировать и структурировать их дефиниции. Однако лексикографическая информация не всегда позволяет выделить метасемы: при описании абстрактной лексики чаще всего используется синонимичес-

кий способ: *смысл* – ‘внутреннее, логическое *содержание* (слова, речи, явления), постигаемое разумом, *значение*’; *значение* – ‘*смысл*, то, что данный предмет (слово, жест, знак) значит’.

Только переработав лексикографические данные и подкрепив их дополнительным исследованием, удалось сформулировать метаязыковые толкования ментефактов (табл. 5). Так, *знание* – это результат обработки информации, *тайна* – отсутствие результата поиска информации, *смысл* – цель работы с информацией и т.д.

Таким образом, в результате проведенной работы были сделаны следующие выводы:

1. Описанный метаязык представляет собой особый тип дефиниций, ориентированных на компьютерный анализ лексики и как следствие – скорее на компьютер, чем на человека. Это предполагает снятие неоднозначности (асимметрии) метаслов – как в виде омонимии-полисемии, так и в виде синонимии.

Т а б л и ц а 5

Метаязыковое толкование ментефактов

ЗНАНИЕ	ментефакт	информация	результат	
ТАЙНА	ментефакт	информация	результат	нет
СМЫСЛ	ментефакт	информация	цель	
ПАМЯТЬ	ментефакт	контейнер	знание	
МНЕНИЕ	ментефакт	оценка		
ПРАВДА	ментефакт	информация	хорошо	

2. Лексические константы целесообразно представлять в виде существительных.

3. В отличие от компонентного анализа создание метаязыковых толкований предполагает не детализацию и различение, а обобщение и формализацию лексических значений слов с целью их использования при создании систем автоматической обработки текста. Представленный метаязык – это *метаязык обобщения*. Его назначение – не *отличать ветер от ветерка, бури, шквала, вихря* и т.п., а *объединять* их в единый семантический класс. Глубина членения в классификации – вопрос открытый, но в любом случае – не до отдельного ЛСВ, а до группы слов: синонимов или эквонимов.

4. Предъявляются жесткие требования к синтаксису метаязыка для обеспечения правильного результата использования метаязыковых толкований.

5. Легче всего поддается метаязыковому описанию предметная лексика, представленная классами «натурфакты», «растения», «животные». Ментефакты являются наиболее сложным классом слов, требующим применения специальных, т.е. отражающих их специфику, методов анализа.

ЛИТЕРАТУРА

1. The Penguin Roget's thesaurus of English words and phrases / New edition completely revised, updated and abridged by Susan M. Lloyd, Penguin books, 1986. – 776 p.
2. Морковкин В. В. Лексическая основа русского языка : комплексный учебный словарь / В. В. Морковкин [и др.] ; под ред. В. В. Морковкина. – М. : Рус. яз., 1984. – 1168 с.
3. Кобозева И. М. Лингвистическая семантика : учебник / И. М. Кобозева. – Изд. 2-е. – М. : Едиториал УРСС, 2004. – 352 с.
4. Апресян Ю. Д. О московской семантической школе / Ю. Д. Апресян // Вопросы языкознания. – 2005. – № 1. – С. 3–30.
5. Руделев В. Г. Слово в словаре / В. Г. Руделев // Слово II : сб. науч. работ. – Тамбов, 1997. – С. 104–115.
6. Мельчук И. А. Русский язык в модели «Смысл ↔ Текст» / И. А. Мельчук. – М. ; Вена : Языки рус. культуры, 1995. – 714 с.
7. Wierzbicka A. Semantic primitives. – Frankfurt : Athenäum, 1972.

8. Вежбицка А. Семантические примитивы / А. Вежбицка // Семиотика / под ред. Ю. С. Степанова. – М., 1983. – С. 225–252.
9. Караулов Ю. Н. Русский семантический словарь. Опыт автоматического построения тезауруса : от понятия к слову / Ю. Н. Караулов [и др.]. – М. : Наука, 1982. – 566 с.
10. Кузнецова Э. В. Ступенчатая идентификация как средство описания семантических связей слов / Э. В. Кузнецова // Вопросы металингвистики. – Л. : Изд-во ЛГУ, 1973. – С. 84–95.
11. Лексико-семантические группы русских глаголов / под ред. Э. В. Кузнецовой. – Иркутск : Изд-во Иркут. ун-та, 1989. – 180 с.
12. Большой толковый словарь русских существительных : идеографическое описание. Синонимы. Антонимы / под ред. Л. Г. Бабенко. – М. : Аст-Пресс Книга, 2005. – 864 с.
13. Большой толковый словарь русских глаголов : идеографическое описание. Синонимы. Антонимы. Английские эквиваленты / под ред. Л. Г. Бабенко. – М. : Аст-Пресс Книга, 2007. – 576 с.
14. Словарь-тезаурус прилагательных русского языка / под общ. ред. Л. Г. Бабенко. – Екатеринбург : Изд-во Урал. ун-та, 2012. – 840 с.
15. Подосинов А. В. Латинско-русский словарь : около 13 000 слов / А. В. Подосинов, Г. Г. Козлова, А. А. Глухов. – М. : Флинта : Наука, 1998. – 376 с.
16. Красова Г. А. Итальянско-русский словарь : около 12 000 слов / Г. А. Красова, Г. Дзаппи. – М. : Рус. яз., 1996. – 352 с.
17. Испанско-русский и русско-испанский словарь для школьников. – Киев : Логос, 1998. – 768 с.
18. Шалагина И. Н. Карманный португальско-русский словарь : 9000 слов / И. Н. Шалагина. – М. : Рус. яз., 1987. – 352 с.
19. Выгодская К. С. Краткий французско-русский и русско-французский словарь / К. С. Выгодская, О. Л. Долгополова. – М. : Советская энциклопедия, 1970. – С. 9–310.
20. Андрианов Б. А. Карманный румынско-русский словарь : около 8000 слов / Б. А. Андрианов. – М. : Сов. энцикл., 1964. – 316 с.
21. Szmidt Dorota. Dictionarirus-catalá y catalá-rus / Szmidt Dorota, Zgustova Monika (rus-catalá), Bank Svetlana (catalá-rus). – Barcelona : Enciclopèdiacatalana, 1999. – 585 p.
22. Кретов А. А. Константы и переменные лексико-семантической системы / А. А. Кретов // Язык – когниция – коммуникация : тезисы Междунар. науч. конф. (3–6 ноября 2010 г.) / редкол. З. А. Харитончик (отв. ред.) [и др.]. – Минск : МГЛУ, 2010. – С. 46–47.
23. Караулов Ю. Н. Общая и русская идеография / Ю. Н. Караулов. – М. : Наука, 1976. – 356 с.

Воронежский государственный университет
Кретов А. А., профессор кафедры теоретической
и прикладной лингвистики
E-mail: a_a_kretov@rambler.ru

*Подтележникова Е. Н., доцент кафедры теорети-
ческой и прикладной лингвистики*
E-mail: podtelezchnikova@yandex.ru
Тел.: 8-915-589-27-34

Voronezh State University
Kretov A. A., Professor of the Theoretical and Applied
Linguistics Department
E-mail: a_a_kretov@rambler.ru

*Podtelezchnikova E. N., Associate Professor of the
Theoretical and Applied Linguistics Department*
E-mail: podtelezchnikova@yandex.ru
Tel.: 8-915-589-27-34