

К ПРОБЛЕМЕ СОЗДАНИЯ МЕТАЯЗЫКА СЕМАНТИЧЕСКОГО ОПИСАНИЯ ЛЕКСИКИ

А. А. Кретов, Е. Н. Подтележникова

Воронежский государственный университет

Поступила в редакцию 5 июня 2014 г.

Аннотация: в статье проводится обзор литературы, посвященной разработке метаязыка семантического описания лексики. Несмотря на наличие большого количества работ, существующие методики не подходят для создания универсального метаязыка по двум причинам: субъективизм авторов и смешение понятий грамматической и лексической семантик. В статье также описаны требования к метаязыку и предложены принципы его создания.

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

Abstract: the article offers a review of methods of metalanguage creation. In spite of a wide variety of different methods they can not serve as a basis for metalanguage because of authors' views and mixing of grammar and lexical semantics. The article also presents the main features of metalanguage and principles of its creation.

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

В последние десятилетия одной из ключевых проблем лексикологии и семасиологии является отсутствие адекватного метаязыка для семантического описания лексики, несмотря на непрерывающуюся работу в этом направлении.

Данная статья призвана решить следующие задачи:

1. Проанализировать существующие методики семантизации лексики.
2. Сформулировать понятие «метаязык семантического описания лексики», обосновать цели и задачи его построения.
3. Предложить принципы его создания.

Основы семантической классификации лексики заложены в работах П. С. Палласа («Сравнительные словари всех языков и наречий, собранные десницею всевысочайшей особы», 1787–1789), А. С. Шишкова («Собрания языков и наречий с примечаниями на оныя», 1832), Lorge Irving, Thorndike Edvard E. («A Semantic Count of English Words», 1938), В. В. Прокопова («Основные лексико-семантические группы русского глагола», 1945), Ю. Н. Караулова («Русский семантический словарь», 1982), В. В. Морковкина («Лексическая основа русского языка. Комплексный учебный словарь», 1984), The Penguin Roget's thesaurus of English words and phrases (1986), Э. В. Кузнецовой («Лексико-семантические группы русских глаголов», 1988), О. С. Баранова («Идеогра-

фический словарь русского языка», 1995), Н. Ю. Шведовой («Русский семантический словарь», 1998–2007), Л. Г. Бабенко («Толковый словарь русских глаголов: «идеографическое описание», 1999).

Суть этих исследований изложена в работе А. А. Кретова и А. В. Рафаевой «К созданию компьютерной системы семантической классификации лексики» [1, с. 53–56]. Укажем причины, не позволяющие рассматривать предложенные методики как базу для создания семантического метаязыка.

Во-первых, это субъективизм в составлении словаря, вследствие чего словарные статьи оказываются неполными или, наоборот, избыточными. Например, в словаре под редакцией Л. Г. Бабенко [2, с. 75] лексема *взгляд* не входит в группу «процесс зрительного восприятия», хотя в этой группе есть такие лексемы, как *направлять*, *сосредотачивать*, *устремлять*, *вонзять*. Возникает вопрос, что авторы словаря предлагают *направлять*, *сосредотачивать*, *вонзять*, *устремлять* на объект? Слово *глаза* входит только в группу «голова живого существа и ее части», а слова *взор* совсем нет в словаре.

Тем не менее очень детально представлено значение слова *взгляд*, связанное с мыслительной деятельностью. Так, слово *взгляд* входит в группу 7.2.1.3. «результат речемыслительной деятельности» вместе со словами *убеждение*, *воззрение*, а также в группу 7.2.4.2. «результат процесса определения» со словами *мнение*, *воззрение*, *голос*, *позиция*, *соображение*, *суждение*, *понятие*.

Такое разграничение оттенков значения на фоне игнорирования основного значения представляется избыточным и нецелесообразным. Это также указывает на то, что не соблюдена структура тезауруса, так как слово *взгляд* в двух практически тождественных значениях входит в разные синонимические ряды. Более того, члены синонимического ряда также дублируются.

Во-вторых, смешение лексической и грамматической семантики. Большинство словарей в соответствии с взглядами составителей ориентировано на частеречную семантику, поэтому представляют собой чисто грамматические группировки разнородных по своей лексической семантике слов.

Углубление в семантику глагола показывает, что собственно глагольной лексической семантики, по большому счету, не существует – собственно глагольная семантика грамматична: все лексическое в семантике глагола – от имен, например, *идти* – ‘Сaus ноги Func’ (делать так, чтобы *ноги* функционировали).

Анализ семантики прилагательных показывает, что собственно адъективная семантика тоже грамматична – это ‘признак’ какого-то имени: *белый* – ‘цвета снега или молока’, *черный* – ‘цвета сажи’, *голубой* – ‘цвета неба’, *зеленый* – ‘цвета листвы и травы’, *красный* – ‘цвета крови’, *розовый* – ‘цвета розы’, *фиолетовый* – ‘цвета фиалки’, *сиреневый* – ‘цвета сирени’ и т.д. Как видим, лексическая семантика прилагательных также производна – от семантики существительных. О семантике наречий и местоимений говорить не приходится: все лексическое в них – несобственное.

Идеи Л. В. Щербы [3], Ю. Д. Апресяна [4], В. В. Морковкина [5] и других предполагают «вынесение за скобки» всего регулярного в словаре, переосмысление сущности и соотношения лексической и грамматической семантики. *Лексическая семантика* – это семантика, направленная на внеязыковую действительность, это выделенные языком константы внеязыковой действительности, составляющие словарь, а семантические, синтаксические, морфологические функции, выполняемые этими константами, описываются как множество *переменных*. В плане выражения к семантическим константам ближе всего *корни*, в которых, собственно говоря, и концентрируется лексическая семантика. Соответственно, константы лексической семантики следует искать в формально простейших – корневых словах, а в общем случае – в корневых морфемах.

Таким образом, мы приходим к идее, противоположной общепринятой: лексическая и грамматическая семантика не «едины и неделимы», а, наоборот, принципиально противопоставлены друг другу как «константы и переменные языка», если воспользоваться счастливо найденным А. Е. Кибриком назва-

нием его книги [6]. *Лексическая семантика* – это константы, *грамматическая семантика* – переменные. Лексическая семантика первична, грамматическая – производна.

Целью семантической классификации должно быть выделение констант лексической семантики, выражаемых универбами. В таком случае лексическая семантика будет представлять собой лес семантических деревьев, в корне каждого из которых содержится исходное, далее неделимое средствами естественного языка значение, передаваемое, как правило, неизменяемым (корневым) словом.

Набор семантических констант ближе к набору корней (корневых слов) данного языка и не имеет ничего общего с семантическими примитивами А. Вежбицкой [7] (которые при ближайшем рассмотрении оказываются некоторым отдаленным и малоинформативным подобием грамматических переменных). Набор этих корней с неизбежностью будет отражать архаичное мышление и архаичную картину мира.

В силу того, что лексическая семантика внеположна частям речи, а русский язык не может передавать значение, не приписав его к какой-либо части речи, целесообразно ориентироваться на наименее маркированную часть речи – существительное [8, с. 35] и лексические константы представлять в виде существительных, а не выводить из глагольных, как это делается в модели «Смысл ↔ Текст» [9].

Таким образом, наиболее близким аналогом данного исследования можно считать опыт Ю. Н. Караулова [10, с. 57–63], предложившего особый тип дефиниций и особый метаязык (на базе русского), ориентированных на компьютерный анализ и приспособленных для него: предполагается снятие неоднозначности (асимметрии) единиц – как в виде омонимии-полисемии, так и в виде синонимии. Этот метаязык должен быть ориентирован скорее на компьютер, чем на человека.

К сожалению, слова-аттракторы в «Русском семантическом словаре» задавались а priori, а не получались в ходе исследования. Кроме того, этот словарь, как и словарь екатеринбургского коллектива [11], показал, что словарные дефиниции – независимо от их качества – лишь полуфабрикат для семантизации лексики. Понятие «квазиосновы», верное по существу, так как ориентировано на лексическую семантику, к сожалению, оказалось недостаточно продуктивно: от грамматической информации в дефиниции мы отказаться не можем.

Мы полностью согласны с представителями Московской семантической школы, считающими, что семантика конкретного естественного языка должна строиться на основе ограниченного подязыка самого естественного языка. Этот язык, называемый

«естественным семантическим метаязыком», должен быть максимально прост: «в подязык отбираются относительно простые слова, грамматические формы и синтаксические конструкции в их основных значениях», а каждое слово подязыка «должно в идеале удовлетворять требованию взаимно-однозначного соответствия имени и смысла [4, с. 13].

В лексику естественного семантического языка предполагают включать единицы следующих типов:

1) носители так называемых «системообразующих смыслов»:

а) семантические примитивы, т.е. нетолкуемые слова;

б) носители некоторых более сложных смыслов, недалеко отстоящих от примитивов;

в) некоторые более простые, невербализуемые смыслы – «кварки» (например, пересечение смыслов русских слов *хотеть* и *желать*);

2) «промежуточные слова-смыслы», для которых «характерно вхождение в состав многих лексических единиц данного языка» [4, с. 12–13].

Синтаксис естественного семантического языка необходимо также унифицировать: в него не должны входить лингвоспецифичные конструкции естественного языка (например, конструкции так называемого малого синтаксиса, свойственные разговорному языку).

Практическим воплощением этой концепции является «Перспектив активнаго словаря русскаго языка» [12]. В словаре представлено примерно 1100 слов метаязыка (из них – 40 служебных). К толкованиям предъявляются следующие требования:

1) нетавтологичность (отсутствие порочных кругов – определений слов друг через друга);

2) необходимость и достаточность (толкование должно быть семантически эквивалентно толкуемой единице).

Толкования, построенные на таком метаязыке и отвечающие указанным требованиям, – главный инструмент для установления и демонстрации системных семантических связей между лексемами в словаре (синонимии, антонимии, конверсии, регулярной полисемии, семантической производности и т.п.), а также основа для формулировки правил взаимодействия значений.

Однако Н. В. Перцов в статье «К проблеме построения семантического метаязыка» [13] считает, что особенности языковой единицы «должны отражаться посредством самой конфигурации элементов семантического метаязыка, но не посредством особых единиц данного метаязыка. При описании семантики конкретного естественного языка неправомерно вводить в семантический метаязык специальные элементы, исключительно ориентированные на сильно лингвоспецифичные единицы данного ЕЯ» [13, с. 419].

По его мнению, естественный семантический язык работает в рамках конкретного языка, но ограничен при сопоставлении семантических систем разных языков. Универсальный же семантический язык, наоборот, ограничен при решении первой задачи, но должен быть более эффективен при решении второй. Для компьютерной обработки текстов больше подходит универсальный семантический язык.

Практическим применением идей семантической классификации лексики является также создание «универсального сетевого языка» «Universal Networking Language» (далее – UNL). Эта работа проводится в рамках международного проекта под эгидой Организации Объединенных Наций с 1996 г. [14, с. 53–58].

UNL – это компьютерный язык, разработанный для представления информации в таком виде, который позволял бы порождать тексты, содержащие эту информацию, на самых разнообразных языках. UNL состоит из лексических значений английского языка, дополненных особыми атрибутами и семантическими метаязыковыми ограничителями; синтаксис UNL строится как набор конструктов-отношений. Таким образом, проблема создания метаязыка на базе русского в проекте UNL не рассматривается.

В статье А. А. Кретьова и А. В. Рафаевой «К созданию компьютерной системы семантической классификации лексики» [15] описана компьютерная методика объединения всех дефиниций словаря в ограниченное множество деревьев, содержательно соответствующих семантическим полям. Для решения этой задачи авторами создана компьютерная программа ПроСеКа (ПРОграмма СЕмантической Классификации) [15, с. 230], служащая инструментом создания, проверки и сохранения цепочек лексико-семантических единиц, заданных пользователем.

Основная задача программы – сохранять цепочки и текстовые примеры к ним, введенные пользователем, и создавать словарь всех встретившихся в этих цепочках лексико-семантических единиц (далее – ЛСЕ). Кроме того, в функции программы входит контроль за согласованностью данных: каждая ЛСЕ может иметь не более одной дефиниции (финальные узлы цепочек могут не иметь дефиниции), и последовательность ЛСЕ во всех цепочках должна сохраняться неизменной. Узлами цепочек являются фактически семемы (отдельные значения) слов.

На сегодняшний день стремление построить концептосферу русского языка как лес не связанных между собой деревьев представляется чрезмерно жестким. Более реалистичным решением представляется связный граф, предполагающий снятие запрета на связи между деревьями [15, с. 231].

В статье «Константы и переменные лексико-семантической системы» [16, с. 46] проведен анализ

китайской письменности, где более 50 000 иероглифов и передаваемых с их помощью значений сведены к 214 ключам-иероглифам, выступающим в качестве классификаторов. Путем интерпретации классификаторов с помощью метаязыка получены 7 классов констант: «натурфакт», «растения», «животные», «человек», «артефакт», «ментефакт», «сверхъестественное». Также выделено 18 классов переменных (таблица).

Для проверки метаязыка обработаны дефиниции 7 романско-русских словарей, из которых удалены метаслова, встретившиеся только в одном словаре, метаслова с частотой менее 4, не субстантивные метаслова и метасловоформы, не являющиеся леммами существительных, что дало 3767 метаслов-существительных, подлежащих описанию с помощью метаязыка.

Взаимодействие лексико-семантических констант и переменных дает новый класс сущностей – *метаязыковые дериваты*. Например, ‘человек:форма > тело’; ‘тело:часть:верх > голова’; ‘голова:часть:перед > лицо’; ‘лицо:часть:вид > глаза’; ‘глаза:функция > зрение’; ‘артефакт:зрение:вид’ > бинокль, подзорная труба, телескоп, микроскоп, лорнет, пенсне, монокль, очки, контактные линзы’.

Этот пример дает представление о синтаксисе предлагаемого метаязыка, правила которого приведены ниже.

1. Первое место в дефиниции всегда занимает семантическая константа высшего уровня.

2. Первое место может занимать уже описанное с помощью метаязыка метаслово (семантический дериват). Например, свернутый вариант ‘тело:часть:верх’ вместо развернутого ‘человек:форма:часть:верх’ при *тело* = ‘человек:форма’. При этом указанное требование все равно выполняется.

3. Семантический дериват может занимать место после константы, выступая в функции переменной. Например, ‘артефакт:зрение:вид’ вместо ‘артефакт:человек:форма:часть:верх:часть:перед:вид:функция:вид’ при *зрение* = ‘человек:форма:часть:верх:часть:перед:вид’ < ‘человек:форма:часть:верх:часть:перед’ = *лицо*.

4. Семантические комбинации, не имеющие лексического выражения в языке, полностью подходят под понятие *кварков*, предложенное Ю. Д. Апресяном [4, с. 15].

5. Правила 2 и 3 объединены *принципом кумулятивности* метаязыка: каждое истолкованное с помощью метаязыка значение (семантический дериват) может быть использовано в качестве метаслова. Принцип кумулятивности обеспечивает такое свойство метаязыка, как *свертываемость*, за счет чего обеспечивается, с одной стороны, компактность дефиниции, а с другой – строгость и эксплицитность.

Т а б л и ц а

Переменные лексико-семантической системы

Род	Вид-1	Вид-2	Вид-3	Вид-4	Вид-5
1. <Оператор>	Нет				
2. Движение	Покой				
3. Действие	Бездействие				
4. Количество	Число	Много	Мало		
5. Мереология	Целое	Часть	Элемент	Род	Вид
6. Норма	Аномалия				
7. Отношение	Содействие	Антагонизм	Собственность		
8. Оценка	Плохо	Хорошо			
9. Процесс					
10. Свойство	Состояние	Форма	Цвет	Чувство	Пол
11. Стабильность	Всегда	Изменчивость			
12. Субстанция					
13. Существование					
14. Условие	Еда				
15. Фаза	Начало	Конец	Середина	Продолжение	
16. Функция	Назначение				
17. Хронотоп	Пространство	Верх–Низ	Перед–Зад	Время	
18. Цель	Результат	Обстоятельство			

6. Порядок следования единиц метаязыка – от общего к частному. Например, ‘тело:часть’. К частям тела относятся *голова, туловище, руки и ноги*. Метаслово ‘верх’ указывает на голову (тело:часть:верх), метаслово ‘низ’ указало бы на ноги (тело:часть:низ), метаслово ‘середина’ указало бы на туловище с руками (тело:часть:середина). Различить туловище и руки помогло бы метаслово ‘конечность’ (рука – ‘тело:часть:середина:конечность’). Соответственно, использование метаслова ‘не’, помогло бы различить руки и туловище: ‘тело:часть:середина:конечность:не’ – туловище.

7. Определение следует за определяемым, отрицание – за отрицаемым.

8. Члены дефиниции могут повторяться (ср. выше повторение переменных ‘часть’ и ‘вид’).

Так в метаязыке порождаются производные семантические комплексы (*семантические дериваты*), способные выступать в функции констант или переменных, будучи уже иерархически соотношенными с первичными семантическими константами.

Описанный метаязык представляет собой видоизмененный вариант метаязыка, заимствованный из базы данных «Лексикограф», созданного коллективом авторов под руководством Е. В. Падучевой [17, с. 24; 18, с. 35] и используемого в Национальном корпусе русского языка. Критика семантической разметки слов в НКРЯ дана в статье А. А. Кретова «Анализ семантических помет НКРЯ» [19, с. 240–250]. Эта критика оправдана, так как сами авторы статьи признают: «Корпус в принципе не может ставить перед собой задачу описания лексико-семантической системы русского языка... Его задача – обеспечение максимально удобного поиска примеров для максимально широкого круга пользователей» [20].

Как показал опыт, метаязык позволяет эффективно упорядочивать слова по семантике. Например, дефиниция ‘человек:возраст’ объединяет метаслова *дитя, малыши, малютка, ребенок*. Дефиниция ‘человек:возраст:пол’ объединяет метаслова *девочка, девушка, женщина, старуха; мальчик, мальчуган, мужчина, парень, старик, юноша*. Метаслова *девочка, девушка, женщина, старуха* описываются дефиницией ‘человек:возраст:пол: мужской:не’.

В заключение выделим основные требования к метаязыку семантической классификации:

1. Метаязык должен представлять собой особый тип дефиниций, ориентированных на компьютерный анализ лексики. Он должен быть ориентирован скорее на компьютер, чем на человека. Для этого предполагается снятие неоднозначности (асимметрии) единиц – как в виде омонимии-полисемии, так и в виде синонимии.

2. Лексические константы целесообразно представлять в виде существительных.

3. Поскольку слово или словосочетание текста на любом языке может быть семантизировано по-русски, в перспективе за основу можно взять русский язык.

4. Поскольку метаязык предназначен для семантической классификации, т.е. для обобщения, а не для различения лексических значений, нижним пределом обобщения целесообразно выбрать синонимический ряд метаслов или множество эквонимов (метаслов с общим гиперонимом).

ЛИТЕРАТУРА

1. Кретов А. А. К созданию компьютерной системы семантической классификации лексики / А. А. Кретов // Проблемы лингвистической прогностики : сб. науч. трудов / под ред. А. А. Кретова. – Вып. 4. – Воронеж, 2007. – С. 53–69.

2. Бабенко Л. Г. Большой толковый словарь синонимов русской речи : идеографическое описание 2000 синонимических рядов, 10 500 синонимов / Л. Г. Бабенко [и др.] ; под общ. ред. Л. Г. Бабенко. – М. : АСТ-ПРЕСС, 2008. – 753 с.

3. Щерба Л. В. Опыт общей теории лексикографии / Л. В. Щерба // Языковая система и речевая деятельность. – М. : Наука, 1974. – С. 265–304.

4. Апресян Ю. Д. О московской семантической школе / Ю. Д. Апресян // Вопросы языкознания. – 2005. – № 1. – С. 3–30.

5. Морковкин В. В. Основы теории учебной лексикографии / В. В. Морковкин. – М., 1990. – 72 с.

6. Кибрик А. Е. Константы и переменные языка / А. Е. Кибрик. – СПб., 2005. – 719 с.

7. Wierzbicka A. Semantic primitives / A. Wierzbicka. – 1972.

8. Руделёв В. Г. Слово : сб. науч. работ, посв. 20-летию Тамбовской лингвистической школы / В. Г. Руделёв. – Тамбов : Тамбов. ун-т, 1996. – 247 с.

9. Мельчук И. А. Русский язык в модели «Смысл-Текст» / И. А. Мельчук. – М. ; Вена, 1995.

10. Караулов Ю. Н. Русский семантический словарь. Опыт автоматического построения тезауруса : от понятия к слову / Ю. Н. Караулов [и др.]. – М. : Наука, 1982.

11. Лексико-семантические группы русских глаголов / под ред. Э. В. Кузнецовой. – Свердловск, 1988.

12. Апресян В. Ю. Проспект активного словаря русского языка / В. Ю. Апресян [и др.] ; отв. ред. Ю. Д. Апресян. – М. : Языки славянских культур, 2010. – 784 с.

13. Перцов Н. В. К проблеме построения семантического метаязыка / Н. В. Перцов // Компьютерная лингвистика и интеллектуальные технологии : труды Междунар. конф. «Диалог 2006» (Бекасово, 31 мая – 4 июня 2006 г.) / под ред. Н. И. Лауфер, А. С. Нариньяни, В. П. Селегея. – М. : Изд-во РГГУ, 2006. – С. 419–425.

14. Богуславский И. М. Модуль универсального сетевого языка (UNL) в составе системы ЭТАП-3 / И. М. Богуславский [и др.] // Труды Междунар. семинара «Диалог’2000» по компьютерной лингвистике и ее приложениями : в 2 т. – Протвино, 2000. – Т. 2.

15. *Кретов А. А.* Программа семантической классификации ПроСеКа : теоретические и прикладные аспекты / А. А. Кретов, А. В. Рафаева // Компьютерная лингвистика и интеллектуальные технологии : по материалам ежегодной Международной конференции «Диалог 2009» (Бекасово, 27–31 мая 2009 г.). – Вып. 8 (15). – М. : РГГУ, 2009. – С. 230–235.

16. *Кретов А. А.* Константы и переменные лексико-семантической системы / А. А. Кретов // Язык – когнитивия – коммуникация : тезисы Междунар. науч. конф. (3–6 ноября 2010 г.) / отв. ред. З. А. Харитончик. – Минск : МГЛУ, 2010. – С. 46–47.

17. *Красильщик И. С.* Предметные имена в системе «Лексикограф» / И. С. Красильщик, Е. В. Рахилина //

Научно-техническая информация. Сер. 2. – 1992. – № 9. – С. 24–31.

18. *Кустова Г. И.* Словарь как лексическая база данных / Г. И. Кустова, Е. В. Падучева // Вопросы языкознания. – 1994. – № 4. – С. 96–106.

19. *Кретов А. А.* Анализ семантических помет в НКРЯ / А. А. Кретов // Национальный корпус русского языка : 2006–2008. Новые результаты и перспективы. – СПб. : Нестор-История, 2009. – С. 240–257.

20. *Рахилина Е. В.* Задачи и принципы семантической разметки лексики в НКРЯ / Е. В. Рахилина [и др.] // Национальный корпус русского языка : 2006–2008. Новые результаты перспективы. – СПб : Нестор-История, 2009. – С. 215–239.

Воронежский государственный университет

Кретов А. А., доктор филологических наук, профессор, заведующий кафедрой теоретической и прикладной лингвистики

E-mail: a_a_kretov@rambler.ru

Тел.: 8 (473) 220-41-49

Подтележникова Е. Н., кандидат филологических наук, доцент кафедры теоретической и прикладной лингвистики

E-mail: podtelezchnikova@yandex.ru

Тел.: 8 (473) 220-41-49

Voronezh State University

Kretov A. A., Doctor of Philology, Professor, Head of the Theoretical and Applied Linguistics Department

E-mail: a_a_kretov@rambler.ru

Tel.: 8 (473) 220-41-49

Podtelezchnikova E. N., Candidate of Philology, Associate Professor of the Theoretical and Applied Linguistics Department

E-mail: podtelezchnikova@yandex.ru

Tel.: 8 (473) 220-41-49