

ПРОЕКТ МЕТАЯЗЫКА ДЛЯ КОМПЬЮТЕРНОЙ КЛАССИФИКАЦИИ ЛЕКСИЧЕСКИХ ЗНАЧЕНИЙ

Е. Н. Подтележникова

Воронежский государственный университет

Поступила в редакцию 20 апреля 2015 г.

Аннотация: в данной статье рассматривается методика создания метаязыка семантического описания, обсуждается формирование списка переменных и описывается начальный этап формулирования мета-дефиниций.

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

Abstract: this article is devoted to the problem of semantic metalanguage. There is a description of variables and metadefinitions.

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

Создание адекватного метаязыка для семантического описания лексики – одна из актуальных проблем современной лексикологии.

Теоретической базой данного исследования является утверждение представителей Московской семантической школы о том, 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» [1, с. 13].

При этом основной единицей метаязыка мы считаем существительное как наименее маркированную часть речи [2] и лексические константы представляем в виде существительных, а не выводим из глагольных, как это делается в модели «Смысл ↔ Текст» [3].

Описываемый метаязык представляет собой видоизмененный вариант метаязыка, заимствованный из базы данных «Лексикограф», созданной коллективом авторов под руководством Е. В. Падучевой [4] и используемой в Национальном корпусе русского языка. Критика семантической разметки слов в НКРЯ дана в статье [5]. Эта критика оправдана, так как сами авторы статьи отмечают следующее: «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» [6].

На начальном этапе работы был 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 получено 7 классов констант: «натурфакт», «растения», «животные», «человек», «артефакт», «ментефакт», «сверхъестественное». Также выделено 18 классов переменных [7].

Следующим этапом стала обработка дефиниций 7 романско-русских словарей. Из них были 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.

В ходе работы к семи выделенным классам констант были добавлены класс «социум» и резервный класс «прочее». Список переменных был также дополнен и уточнен (табл. 1).

При описании слов с помощью метаязыка мы придерживались правил, изложенных в статье «К проблеме создания метаязыка семантического описания лексики» [8].

Порядок следования единиц метаязыка – от общего к частному. Например, 'Голова: часть' (табл. 2). К частям головы относятся *мозг*, *лицо*, *ухо* и т.д. Метаслова 'внутри' и 'снаружи' различают *лицо* и *ухо* как внешние органы и *мозг* как внутренний, метаслова 'верх' и 'перед' указывают их местоположе-

Т а б л и ц а 1

Переменные лексико-семантической системы

Род	Вид-1	Вид-2	Вид-3	Вид-4	Вид-5	Вид-6	Вид-7	Вид-8
1. Оператор	нет	«-» = и	«/» = или					
2. Движение								
3. Действие								
4. Количество	число	много	мало	пара				
5. Мереология	целое	часть	элемент	род	вид			
6. Норма								
7. Отношение	содействие	антагонизм	собственность					
8. Оценка	плохо	хорошо	своё	чужое				
9. Процесс								
10. Свойство	форма	цвет	чувство	пол	вес	температура	размер	
11. Стабильность	всегда	изменчивость						
12. Субстанция	твёрдое	жидкое	газообразное					
13. Существование	продукт	условие						
15. Фаза	начало	конец	середина	продолжение				
16. Функция	назначение							
17. Пространство	место	верх	низ	горизонталь	перед	зад	внутри	снаружи
18. Время								
19. Цель	результат	обстоятельство						

Т а б л и ц а 2

Метаязыковое описание значений

МОЗГ	голова	часть	внутри	часть	верх				
ЛИЦО	голова	часть	снаружи	часть	перед				
УХО	голова	часть	снаружи	часть	перед	нет	зад	нет	пара
ГЛАЗ	лицо	часть	верх	часть	пара				
ГУБА	лицо	часть	низ	часть	пара				
НОС	лицо	часть	середина						
ГОЛОВА	тело	часть	верх						

ние. Таким же образом рассматриваются части лица.

«Прочитать» метаязыковые дефиниции можно так: *мозг* – это часть головы, которая находится внутри вверху; *лицо* – передняя наружная часть головы; *ухо* – это парная наружная часть головы, находящаяся ни спереди, ни сзади (т.е. сбоку головы); *глаза* – это парный орган, находящийся в верхней части лица; *губы* – это парный орган, находящийся в нижней части лица; *нос* – часть лица, находящаяся посередине; *голова* – верхняя часть тела.

Следующим этапом работы было формирование списка метаслов в количестве 224 и составление их толкований. Это позволило выполнить основное условие создания метаязыка: каждое слово метаязыка само должно быть истолковано. Такое условие оправ-

дывается принципом кумулятивности метаязыка, который обеспечивает его свертываемость и, как следствие, компактность дефиниции, а также их строгость и эксплицитность.

Теперь, используя истолкованные метаслова, можем создать развернутый вариант метаязыка. Вместо *тело* подставляем ‘человек:форма’, вместо *лицо* – ‘голова:часть:снаружи:часть:перед’, вместо *голова* – ‘тело:часть:верх’ (табл. 3). Описание значений получилось, с одной стороны, более громоздкое и трудно воспринимаемое, а с другой стороны, подходящее для автоматической обработки языка.

Таким образом, полученный метаязык представляет собой особый тип дефиниций, ориентированных на компьютерный анализ лексики.

Т а б л и ц а 3

Развернутое описание значений

МОЗГ	человек	форма	часть	верх	часть	внутри	часть	верх				
ЛИЦО	человек	форма	часть	верх	часть	снаружи	часть	перед				
УХО	человек	форма	часть	верх	часть	снаружи	часть	перед	нет	зад	нет	пара
ГЛАЗ	человек	форма	часть	верх	часть	снаружи	часть	перед	часть	верх	часть	пара
ГУБА	человек	форма	часть	верх	часть	снаружи	часть	перед	часть	низ	часть	пара
НОС	человек	форма	часть	верх	часть	снаружи	часть	перед	часть	середина		
ГОЛОВА	человек	форма	часть	верх								

ЛИТЕРАТУРА

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

2. Руделев В. Г. Слово в словаре / В. Г. Руделев // Слово II : сб. науч. работ. – Тамбов, 1997. – С. 104–115.

3. Мельчук И. А. Русский язык в модели «Смысл↔Текст» / И. А. Мельчук. – М. ; Вена : Языки русской культуры, 1995. – 714 с.

4. Красильщик И. С. Предметные имена в системе «Лексикограф» / И. С. Красильщик, Е. В. Рахилина // Научно-техническая информация. – 1992. – № 9. – С. 24–31.

5. 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.

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

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

8. Кретов А. А. К проблеме создания метаязыка семантического описания лексики / А. А. Кретов, Е. Н. Подтележникова // Вестник Воронеж. гос. ун-та. Сер. : Лингвистика и межкультурная коммуникация. – Воронеж, 2014. – № 4. – С. 5–10.

Воронежский государственный университет

Подтележникова Е. Н., доцент кафедры теоретической и прикладной лингвистики

E-mail: podtelezhnikova@yandex.ru

Тел.: 8-915-589-27-34

Voronezh State University

Podtelezhnikova E. N., Associate Professor of the Theoretical and Applied Linguistics Department

E-mail: podtelezhnikova@yandex.ru

Tel.: 8-915-589-27-34