

S T U D I A P H I L O L O G I C A



РОССИЙСКИЙ ГОСУДАРСТВЕННЫЙ
ГУМАНИТАРНЫЙ УНИВЕРСИТЕТ

МЕТОДЫ КОГНИТИВНОГО АНАЛИЗА СЕМАНТИКИ СЛОВА

Компьютерно-корпусный подход

Ответственный редактор
В. И. Заботкина

2-е издание



Издательский Дом ЯСК

УДК 811
ББК 81
М 54

Под общей редакцией:
д. ф. н., проф. В. И. Заботкиной

Коллектив авторов:
д. ф. н., проф. Заботкина В. И., д. ф. н., проф. Голубкова Е. Е.,
д. ф. н., проф. Кронгауз М. А., д. ф. н., проф. Позднякова Е. М.,
д. ф. н., проф. Шаронов И. А., к. ф. н., доц. Боярская Е. Л.,
к. ф. н., доц. Заячкова О. О., к. ф. н., доц. Коннова М. Н.,
к. ф. н. Котов А. А., к. ф. н. Суворина Е. В.

М 54 Методы когнитивного анализа семантики слова: компьютерно-корпусный подход / Под общ. ред. В. И. Заботкиной. — 2-е изд. — М.: Издательский Дом ЯСК, 2019. — 344 с. — (Studia philologica).

ISBN 978-5-907117-73-0

Данная монография представляет собой пример интегрированной методологии в двух современных научных парадигмах: когнитивной и корпусной. В ней решаются общие вопросы методологии когнитивного анализа семантики слова на основе компьютерно-корпусных исследований, изложены результаты, объединенные единой целью — разработать или развить методы в области когнитивной семантики на базе корпусов. Авторы используют Национальный корпус русского языка, Британский национальный корпус и ряд других корпусов. Впервые представлена методология создания мультимодального русскоязычного эмоционального корпуса, включающего материалы вербальной и невербальной коммуникации.

Книга рассчитана на лингвистов, специалистов в области компьютерной и корпусной лингвистики, семиотики, когнитивной психологии и психолингвистики.

Монография адресована не только ученым, но и преподавателям русского и английского языков, бакалаврам, магистрам и аспирантам гуманитарных дисциплин.

УДК 811
ББК 81

В оформлении переплета использована работа Александра Григорьева из серии «Свечение» (1995)

ISBN 978-5-907117-73-0

© Издательский Дом ЯСК, 2019
© Авторы, 2015

СОДЕРЖАНИЕ

Предисловие (Ред. д.ф.н., проф. <i>В. И. Заботкина</i>)	7
Глава 1. От интеграционного вызова в когнитивной науке к интегрированной методологии (<i>В. И. Заботкина</i>)	15
Глава 2. Использование лингвистических корпусов при решении семантических проблем (<i>Е. Е. Голубкова</i>)	39
Глава 3. Разработка методологии когнитивного анализа многозначного слова (<i>Е. Л. Боярская</i>)	81
Глава 4. Корпусные методы исследования сложных случаев полисемии (<i>М. А. Кронгауз</i>)	119
Глава 5. Поиск и описание коммуникативов на основе Национального корпуса русского языка (<i>И. А. Шаронов</i>)	141
Глава 6. Исследование структуры ментального лексикона: вклад компьютерно-корпусной лингвистики (<i>Е. М. Позднякова, Е. В. Суворина</i>)	185
Глава 7. Концептуальный анализ семантики эмоционального лексикона с опорой на корпусные данные (<i>О. О. Заячковская</i>)	243
Глава 8. Когнитивные сдвиги в семантике темпоральных лексем: корпусный аспект (<i>М. Н. Коннова</i>)	269
Глава 9. Методология и методы разработки мультимодальных корпусов (<i>А. А. Котов</i>)	311
ЗАКЛЮЧЕНИЕ	339

ПРЕДИСЛОВИЕ

Когнитивная парадигма лингвистического знания начала свое интенсивное развитие с 80-х годов XX в. в зарубежной лингвистике и с 90-х годов XX в. в России. Первыми аналитическими публикациями стали публикации 1992 и 1994 гг. Института языкознания РАН по структурам представления знаний в языке. В течение последних двух десятилетий наибольшее внимание российских ученых-лингвистов было привлечено к проблемам категоризации и концептуализации, когнитивному анализу дискурса (Е. С. Кубрякова, В. З. Демьянков, В. И. Заботкина, Н. Н. Болдырев, А. А. Кибрик и др.). Однако современное состояние развития когнитивной науки характеризуется динамизмом и тенденцией к освоению нового инструментария и новых методов анализа, в том числе статистического анализа больших массивов данных.

Данная монография посвящена разработке проблем методологической интердисциплинарности. В ней представлены и обобщены методы когнитивного анализа, семантики слова с позиции корпусно-компьютерного подхода. В данном случае речь идет не о простом сложении существующих методов, а об их интеграции — когда методы двух наук дают принципиально новую методологию. Необходимость привлечения корпусных методов в когнитивную лингвистику диктуется прежде всего тем, что концептуальный анализ в рамках когнитивной лингвистики в интерпретации, предложенной Дж. Лакоффом [Lakoff 1987], страдает от недостатка строгих методов тестирования. Менталисты, так же как и структуралисты, работали с моделями языка, которые можно было тестировать интуитивно. Однако эти методы больше не работают для анализа идеализированных когнитивных моделей и культурно-специфических концептов. Обзоры новых методов анализа когнитивной семантики, основывающихся на корпусных данных, были представлены в двух монографиях Гриса и Стеффановича [Gries, Stefanovitsch 2006], [Stefanovitsch, Gries 2006], а также в моно-

графии Глина и Фишер [Glynn, Fischer forthcoming]. Известно применение мультивариативного статистического метода для исследования полисемии и синонимии. [Glynn 2008]. Следующим шагом в развитии данного метода является анализ того, как этот метод позволяет эмпирически определить концептуальные структуры. В работах британских лингвистов-когнитологов доказывается то, что частотность употребления слова тесно взаимосвязана со строением концептуальной структуры, лежащей в основе его значения [Diviak 2014].

В отечественной лингвистике корпусные подходы к анализу когнитивной семантики слова разрабатываются в трудах таких ученых, как Е. В. Рахилина, В. И. Подлесская, А. А. Кибрик, Н. Б. Гвишиани и Е. Е. Голубкова.

Корпусные данные английского языка стали привлекаться в России как основание для лингвистических, в том числе лингвокогнитивных, исследований в основной своей массе уже в начале XXI в. Большая часть исследований велась на основе ставшего наиболее популярным в то время Британского национального корпуса (годы создания 1991—1994, второе издание с небольшими изменениями — 2001 г., третье — XML edition 2007 г.). BNC — не самый первый электронный корпус, созданный в мире, но один из лучших, крупнейших и известных. Практически он стал эталоном лингвистического корпуса.

Национальный корпус русского языка (НКРЯ) — доступный для поиска электронный онлайн-корпус русских текстов. Открыт 29 апреля 2004 г. по адресу <http://ruscorpora.ru/>. Известен вклад в разработку этого корпуса В. А. Плунгяна, Е. В. Рахилиной и многих других лингвистов.

В России был создан ряд специальных корпусов текстов (А. А. Кибрик, В. И. Подлесская, А. Н. Баранов и др.), предназначенных для анализа устного дискурса.

Авторами настоящей монографии поставлена и решена проблема разработки общей методологии когнитивного анализа семантики слова на основе компьютерно-корпусных методов исследования. Основные результаты работы коллектива состоят, прежде всего, в преодолении фрагментарности применения корпусных методов для анализа семантики слова с когнитивных позиций, а также в создании новых или усовершенствовании существующих методов в области когнитивной семантики. Важно отметить, что *впервые* корпуса английского и русского языков были здесь системно использованы с целью моделирования когнитивных структур и когнитивных процессов. С этой целью авторы данной монографии обращаются к НКРЯ, а также к BNC и другим англоязычным корпусам.

В рамках монографии осуществлено развитие методов фреймового анализа, пропозиционального анализа, анализа qualia-структур, метода концептуального моделирования, метода анализа лексического ядра группы слов с опорой на корпусные данные, метод концептуального прайминга. Предложены новые методы, основанные на данных, предоставляемых корпусами: метод анализа структуры ментального лексикона с помощью компьютерных программ, построенных на корпусных данных; метод построения лексических профилей слова. Представлены методы, позволяющие создавать мультимодальные корпуса. Созданы компьютерные агенты, использованные для моделирования вербального и невербального эмоционального поведения человека.

Актуальность, значимость и новизна проведенного исследования состоит в комплексном и системном применении корпусного подхода в области когнитивной семантики.

Полученные результаты можно рассматривать как вклад в развитие междисциплинарных связей, затрагивающих такие сферы когнитивной науки, как психология, философия, семиотика, компьютерная лингвистика. Проблема междисциплинарности, являющейся методологической основой когнитивной науки, получила здесь новое освещение. Результаты, представленные в монографии, могут быть использованы не только в сфере лингвистики, но и в сфере исследования информационных процессов, компьютерного моделирования и разработки методов построения мультимодальных корпусов.

Монография открывается главой «От интеграционного вызова в когнитивной науке к интегрированной методологии», написанной д. ф. н., профессором В. И. Заботкиной. В этой главе обсуждаются общие вопросы методологии когнитивных исследований. Тот факт, что глава о роли междисциплинарного подхода открывает монографию, является совершенно оправданным. Ведь использование компьютерно-корпусного подхода в лингвистике — явление междисциплинарное.

В России когнитивные исследования активно развиваются в нескольких направлениях. Прежде всего это лингвистика, психология, философия, искусственный интеллект и нейронауки. Сложность задач, решаемых когнитивистикой, на практике делает подобное деление весьма условным. Интеграционный вызов, стоящий перед когнитивной наукой, заключается в необходимости разработки общего каркаса (framework), который мог бы объединить все составляющие данной науки на основе общих принципов, общего объекта исследования.

Вторая глава «Использование лингвистических корпусов при решении семантических проблем» написана д. ф. н., профессором Е. Е. Голубковой. Она посвящена сопоставлению двух подходов к использованию корпуса в лингвистических исследованиях, так называемым *Corpus-driven* и *Corpus-based* подходам. Первый из подходов позиционируется как дедуктивный, эмпирический, полностью основывающийся на корпусных данных. Второй рассматривает корпус как методологию исследования и способ верификации выдвигаемой гипотезы. Практическая часть представляет собой корпусно-когнитивное исследование структурных и семантических различий в номенклатуре и функционировании фразовых глаголов современного английского языка в его британском и американском вариантах. Исследование опирается на словарные данные и информацию из корпусов. В целом глава показывает преимущества использования корпусных данных в когнитивной лингвистике.

В третьей главе «Разработка методологии когнитивного анализа многозначного слова», написанной к. ф. н., доцентом Е. Л. Боярской, представлены структура методологии когнитивного анализа слова и требования, которым она должна удовлетворять.

Автор указывает, что, прежде всего, разрабатываемая методология должна носить комплексный и комплементарный характер, т. к. применение традиционных методов исследования к такому сложному когнитивному и языковому феномену, как полисемия, может оказаться недостаточным. Следовательно, методология должна включать несколько методик, применение которых в комплексе позволит исследователям приблизиться к разгадке механизмов развития и функционирования многозначности в целом и полисемии как одной из ее форм.

Особая роль в рамках разрабатываемой методологии принадлежит эмпирическим методам исследования — вероятностному прогнозированию и концептуальному моделированию как формам познания. Так, для исследования процесса развития нового смысла в структуре многозначного слова и его представления используется методика, разработанная на основе понятия *Qualia*-структуры. В основу второго направления положены исследования ментального лексикона, в то время как третья группа методов основывается на данных корпусных исследований. Таким образом, соблюдается принцип комплементарности при рассмотрении процессов формирования смыслов, их накопления, усвоения и распознавания.

Четвертая глава «Корпусные методы исследования сложных случаев полисемии», подготовленная д. ф. н., проф. М. А. Кронгаузом, рассматривает один из сложных случаев полисемии, описание которого основывается на корпусном исследовании. Метод корпусного ис-