

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



ЯЗЫКИ СЛАВЯНСКОЙ КУЛЬТУРЫ
МОСКВА 2015

УДК 811
ББК 81
М 54

Работа над монографией осуществлена при поддержке
Российского фонда фундаментальных исследований (РФФИ)
Инициативный проект № 11-06-00301



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. *Заячковская 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 — М.: Языки славянской культуры, 2015. — 344 с. — (Studia philologica).

ISBN 978-5-9445-7209-7

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.

ББК 81

ISBN 978-5-9445-7209-7

© Издательство «Языки славянской культуры», 2015
© 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. Подлеская, 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. Подлеская, 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-

следования дополняется лексикографическим методом, анализом толковых словарей русского языка. Предметом исследования объявляется группа существительных со значением числа, ядро которой составляют такие существительные, как *двушка*, *трёшка*, *тройак*, *пятак*. Кроме того, в исследование вовлекается и лексика, основанная на других словообразовательных моделях: *двойка*, *тройка*, *трешница*, *двадцатник* и др. Лексическое ядро получает полное корпусное и словарное описание, после чего все употребления в корпусе интерпретируются через представленные в словарях толкования. В том случае, если подходящее толкование в словарях русского языка отсутствует, в работе предлагается рабочее толкование. Таким образом, автор получает семантическую сеть, включающую и регулярные, и единичные толкования, которые могут быть охарактеризованы с хронологической точки зрения.

Глава пятая «Поиск и описание коммуникативов на основе Национального корпуса русского языка» подготовлена д. ф. н. проф. И. А. Шароновым и посвящена проблеме выявления особых языковых единиц — коммуникативов, описанию их функционирования и толкования. Коммуникативы — это подвид слов-предложений, идиоматизированные ответные реплики диалога, используемые как клише для выражения оценок, эмоций, отношения говорящего к тому, что он услышал или увидел непосредственно в момент речи. Это единицы типа *Ну вот еще*; *Подумаешь*; *Да-да-да*; *Да ладно*; *Отчего же*; *Нечего*; *нечего*; *Ну дела!*; *Нут уж*; *Вот так-то* и т. п. В статье предлагается классификация коммуникативов, описание их основных семантических и дискурсивных свойств. На примере анализа целого ряда единиц демонстрируется отличие предлагаемого описания коммуникативов с помощью корпусных исследований от описаний в традиционной лексикографии. В качестве базового источника сбора материала выбран Национальный корпус русского языка, поскольку данный корпус наиболее функционально развит, в частности в нем проведена синтаксическая разметка, без которой сбор материала был бы крайне затруднителен.

Шестая глава «Исследование структуры ментального лексикона: вклад компьютерно-корпусной лингвистики» (д. ф. н., проф. Е. М. Позднякова, к. ф. н. Е. В. Суворина) состоит из двух частей. В первой части, написанной Е. М. Поздняковой, дан обзор ряда существующих компьютерных программ, построенных на материалах корпусов, показаны возможности их использования для описания структуры ментального лексикона. Высказаны предположения о семантических и ассоциативных связях слов в ментальном лексиконе. Во второй части главы, написанной Е. В. Сувориной, представлены

данные анализа Британского национального корпуса для слов *emotion/emotions* и *feeling/feelings*. Применительно к настоящему исследованию методология корпусной лингвистики использована для статистического анализа и описания лингвокогнитивных характеристик коллокаций слов *emotion/emotions* и *feeling/feelings* в современном английском языке. Автор вводит в оборот корпусных исследований в когнитивной лингвистике понятия «естественный лексический профиль» и «адаптивный лексический профиль» с целью анализа лингвокогнитивных характеристик исследуемых слов.

Глава седьмая «Концептуальный анализ семантики эмоционального лексикона с опорой на корпусные данные», написанная О. О. Заячковой, рассматривает методы концептуального анализа семантики слов, называющих эмоции, с опорой на корпусные данные. В области когнитивной семантики широко применяются такие положительно зарекомендовавшие себя методы исследования, как выявление конструкционносочетаемого профиля и коллострукционный анализ. Анализ этих методов и посвящена данная глава.

Метод определения конструкционносочетаемого профиля (St. Gries) применен О. О. Заячковой в исследовании концептуальной семантики эмоционального лексикона. При построении конструкционносочетаемого профиля осуществляется оценка данных, для чего используют иерархические агломеративные методы кластерного анализа. Кластерный анализ (*cluster analysis*; см. работы S. Gries, A. Stefanowitch) получил развитие в последние годы в связи с возможностью компьютерной обработки больших баз данных. Кластерный анализ предполагает выделение по какому-либо признаку компактных групп данных, которые представлены в виде матриц близости или расстояний между объектами либо в виде точек в многомерном пространстве.

О. О. Заячковская отмечает, что исследователи находятся в постоянном поиске новых стратегий анализа семантики слова, ищут пути взаимодействия и точки соприкосновения когнитивной лингвистики и корпусной лингвистики. Так, метод выявления конструкционносочетаемого профиля имеет сходство с новым, разрабатываемым в настоящее время методом *коллострукционного анализа* (*collostruction, collostruction analysis*).

Глава восьмая «Когнитивные сдвиги в семантике темпоральных лексем: корпусный аспект» (к. ф. н. М. Н. Коннова) посвящена анализу когнитивных механизмов изменения значения прилагательного *вечный*, в смысловой структуре которого когнитивная (темпоральная) составляющая теснейшим образом сопряжена с ценностным, аксиологическим компонентом. Происходящая в диахронии трансформация ценностного компонента значения оказывает влияние на языковой узус, модифици-

руя возможности совместного употребления прилагательного *вечный* с субстантивами различных лексико-семантических групп.

В современном русском языке субстантивные конструкции с прилагательным *вечный* соотносятся с полярными бытийными планами — вневременным и неизменным и повседневно-обыденным. Когнитивные сдвиги в значении прилагательного *вечный* являются непосредственным следствием сдвигов прагматического, а точнее, аксиологического плана и отражают глубинные перемены мировоззренческого характера. Эта трансформация свидетельствует о перво-степенном значении для формирования аксиологического потенциала слова культурного контекста, предопределяющего его ценностные смыслы. Утрата живой связи с культурным контекстом приводит в конце концов к десемантизации, к потере аксиологического компонента значения. Данное исследование выполнено на материале НКРЯ.

Глава девятая «Методология и методы разработки мультимодальных корпусов», подготовленная к. ф. н. А. А. Котовым, завершает монографию исследованием с когнитивно-коммуникативных позиций не только вербальной, но и невербальной коммуникации. Среди достижений последних лет в области компьютерной и корпусной лингвистики следует отметить создание мультимодальных корпусов. Зарубежный опыт предлагает такие разработки, как проект CogAff (университет г. Бирмингем, Великобритания); компьютерный агент Грета (университет Париж 8, Франция), агент Макс (университет Билефельда, Германия). В американской лингвистике представителем данного направления исследований является Stefan Th. Gries (University of California, Santa Barbara). В его работах, в частности, представлена связь поведенческих моделей в корпусной лингвистике. В работах Dylan Glynn и Kerstin Fischer рассматривается когнитивная семантика в ее связи с корпусной лингвистикой. В данной книге А. А. Котовым предложен новый подход к построению мультимодального корпуса, в котором представлены материалы вербальной и невербальной коммуникации, описан метод создания русскоязычного эмоционального корпуса, показана возможность создания компьютерного агента на базе такого корпуса. Проведена выборочная разметка подобного корпуса по следующим дорожкам: речь, сегментация речи, иллокутивные цели, стратегии вежливости участников диалога, разметка мимики участников диалога. Выделены и проанализированы основные эмоциональные стратегии взаимодействия. На основе корпуса выделены и проанализированы основные мимические паттерны, а также типовые мимические и жестовые окружения высказываний. Выделенные паттерны использованы для пополнения библиотеки мультимодального взаимодействия эмоциональных компьютерных агентов.

Подводя итог, следует подчеркнуть, что все описанные выше направления анализа семантики слова в когнитивном аспекте, представленные в монографии, основаны на методах лингвистических исследований, применяемых или созданных в России и за рубежом в последнее десятилетие. Однако результаты исследований, предложенные авторами монографии, таковы, что можно утверждать, что в настоящей работе сделан определенный шаг вперед как в разработке методов, так и методологии когнитивного анализа семантики на основе корпусных данных в целом. При этом некоторые из предложенных методов еще не вошли в активный научный оборот не только в российской, но и в мировой когнитивной науке.

Таким образом, работа выполнена в русле новейших исследований в области когнитивных наук и, следовательно, соответствует требованиям современного междисциплинарного подхода, сочетающего данные гуманитарных и естественных наук.

Работа над монографией осуществлялась при поддержке Российского фонда фундаментальных исследований в рамках инициативного научного проекта «Когнитивный анализ семантики слова (компьютерно-корпусный подход)» № 11-06-00301 в 2011—2013 гг., а публикация монографии — в рамках издательского проекта № 14-06-07022 «Издание монографии «Методы когнитивного анализа семантики слова: компьютерно-корпусный подход», поддержанного РФФИ в 2014 г.

В. И. Заботкина

ЛИТЕРАТУРА

- Lakoff 1987 — *Lakoff G. Women, Fire and Dangerous Things*. London: UCP, 1987.
- Gries, Stefanowitsch 2006 — *Gries St., Stefanowitsch A. Corpora in cognitive linguistics*. Berlin: Mouton, 2006.
- Stefanovitsch, Gries 2006 — *Stefanovitsch A., Gries St. Corpus-Based Approaches to Metaphor and Metonymy*. Berlin: Mouton, 2006.
- Glynn 2008 — *Glynn D. Polysemy, Syntax, and Variation. A usage-based method for Cognitive Semantics*. New Directions in Cognitive Linguistics / Eds. by V. Evans, S. Pourcel. Amsterdam: Benjamins, 2008.

ГЛАВА 1

В. И. Заботкина

ОТ ИНТЕГРАЦИОННОГО ВЫЗОВА В КОГНИТИВНОЙ НАУКЕ К ИНТЕГРИРОВАННОЙ МЕТОДОЛОГИИ

Одним из основных вызовов XXI в., стоящих перед когнитивной наукой, является проблема интеграции различных сфер знания и дисциплин, входящих в данную науку. Как справедливо указывает директор Курчатовского института М. В. Ковальчук, «сейчас цивилизация находится в точке бифуркации, смены методов познания. Процесс познания до середины XX в. представлял собой путь анализа: мы углубляли и продолжаем углублять свои знания о мире, подстраивая под них науку, технологии, а затем и промышленность. В середине прошлого века начала формироваться принципиально новая линия — линия синтеза. Сегодня анализ поглощается синтезом, но вместе с тем не исчезает, а продолжает развиваться параллельно, оставаясь его основой и предпосылкой... Мы достигли той стадии, когда дальнейшее развитие науки, образования, промышленности возможно только на междисциплинарной основе, конвергенции, взаимопроникновении наук и технологий, ведь сама природа конвергентна по своей сути [Ковальчук и др. 2013].

Как указывалось в наших предыдущих публикациях [Заботкина 2012], современный этап развития общества и науки характеризуется переходом к новой модели порождения знания. Если традиционная модель была узкоспециальной, гомогенной, иерархической и определялась, как правило, академическим сообществом [Fox 2009], то новая модель междисциплинарна, генерируется в прикладном контек-