

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ  
**«БЕЛГОРОДСКИЙ ГОСУДАРСТВЕННЫЙ НАЦИОНАЛЬНЫЙ  
ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ»**  
( Н И У « Б е л Г У » )

ИНСТИТУТ МЕЖКУЛЬТУРНОЙ КОММУНИКАЦИИ И  
МЕЖДУНАРОДНЫХ ОТНОШЕНИЙ  
КАФЕДРА АНГЛИЙСКОЙ ФИЛОЛОГИИ И МЕЖКУЛЬТУРНОЙ  
КОММУНИКАЦИИ

**СПЕЦИФИКА ИСПОЛЬЗОВАНИЯ ЛИНГВИСТИЧЕСКОГО  
КОРПУСА ДЛЯ ИЗУЧЕНИЯ ПОЛИСЕМИИ СЛОВА**

Выпускная квалификационная работа

обучающегося по направлению подготовки  
45.03.03. Фундаментальная и прикладная лингвистика  
очной формы обучения,  
группы 04001417  
Тищенко Анны Андреевны

Научный руководитель  
к.ф.н, доцент  
Белкина И.В.

БЕЛГОРОД 2018

## ОГЛАВЛЕНИЕ

|   |    |
|---|----|
| ВВЕДЕНИЕ.....   | 3  |
| ГЛАВА 1. ОСНОВНЫЕ ТЕОРЕТИЧЕСКИЕ ПРОБЛЕМЫ<br>ИССЛЕДОВАНИЯ ЛИНГВИСТИЧЕСКИХ КОРПУСОВ.....  | 6  |
| 1.1 Основные положения корпусной лингвистики.....   | 6  |
| 1.2 Традиционные и корпусные исследования.....  | 8  |
| 1.3 История создания лингвистических корпусов.....  | 10 |
| 1.4 Концептуальная классификация корпусов.....  | 13 |
| ВЫВОДЫ ПО ГЛАВЕ 1.....  | 19 |
| ГЛАВА 2. ЯЗЫКОВОЙ КОРПУС КАК БАЗА ДЛЯ ИССЛЕДОВАНИЯ<br>МНОГОЗНАЧНЫХ СЛОВ.....  | 21 |
| 2.1 Методы автоматического разрешения многозначности.....   | 21 |
| 2.2 Применение корпусных данных при анализе многозначных слов. Роль<br>контекста при выборе семантического варианта в многозначном слове..... | 27 |
| 2.3 Концептуальное моделирование процесса разрешения полисемии с<br>использованием корпуса.....   | 39 |
| ВЫВОДЫ ПО ГЛАВЕ 2.....  | 47 |
| ЗАКЛЮЧЕНИЕ.....   | 49 |
| СПИСОК ИСПОЛЬЗУЕМОЙ ЛИТЕРАТУРЫ.....   | 51 |
| СПИСОК ИСТОЧНИКОВ ФАКТИЧЕСКОГО МАТЕРИАЛА.....   | 53 |

## ВВЕДЕНИЕ

Данная выпускная квалификационная работа посвящена изучению специфики использования лингвистического корпуса для изучения полисемии слова. Повсеместная компьютеризация позволяет без труда создавать большие языковые базы данных для изучения того или иного языка.

Мы живем в эпоху повсеместного становления информационного общества. Процессы информатизации значительно меняют и облегчают нашу жизнь. Наиболее важным ресурсом в современном обществе становится информация. Сохранение, преумножение и рациональное использование этого ресурса стало иметь огромное значение для общества. Всеми этими возможностями обладает лингвистический корпус. Кроме того, лингвистический корпус – это отличный источник поиска контекстуального употребления слов.

**Актуальность** данной работы обусловлена необходимостью изучения лексической многозначности слов в рамках одного из ведущих направлений языкознания – корпусной лингвистики. С развитием компьютерной лингвистики лексическая многозначность слова становится одной из главных проблем для систем автоматической обработки текстов. В данной работе затрагиваются актуальные вопросы выявления и описания лексической многозначности. Лексическая многозначность – неотъемлемая часть естественного языка. Слова и словосочетания могут нести различную смысловую нагрузку в зависимости от контекста, в котором они использовались.

**Цель** работы состоит в выявлении специфики использования лингвистического корпуса для изучения полисемии слова.

Поставленная цель диктует следующие **задачи** исследования:

1. Показать преимущества использования корпусов текстов для решения теоретических и прикладных задач в лингвистике.
2. Проанализировать особенности реализации автоматического разрешения многозначности.
3. Продемонстрировать результаты применения корпуса при изучении многозначных частотных слов в современном английском языке.

**Объектом** исследования является Британский национальный корпус (British National Corpus) и Корпус современного американского английского (Corpus of Contemporary American English).

**Предметом** исследования является семантические характеристики многозначного слова в современном английском языке.

**Материалом** исследования послужили примеры многозначных слов, почерпнутые из электронных текстов корпусов British National Corpus и Corpus of Contemporary American English.

**Методологической основой** исследования послужили научные положения, составляющие суть таких парадигм знания, как:

- компьютерная лингвистика, представленная трудами А.Н. Баранова, В.П. Захаровой, В.В. Рыкова и других ученых;
- лексическая семантика, разрабатываемая трудами Т.И. Вендиной, У. Чейфом и другими.

**Методы** исследования. В процессе работы использовались следующие методы:

- метод анализа словарных дефиниций;
- метод контекстуального анализа;
- метод концептуального моделирования;
- метод корпусного анализа.

**Апробация работы.** Основные положения и результаты данного исследования представлены в докладах на следующих научных конференциях:

- «Проблемы, перспективы и направления инновационного развития науки», 24 ноября 2017 г., г. Омск, Российская Федерация.
- «Интеграционные процессы в науке в современных условиях», 20 марта 2018 г., г. Новосибирск, Российская Федерация.
- Студенческая научная конференция в секции «Концептуализация мира в языке. Компьютерная лингвистика», 18 апреля 2018 г., г. Белгород в рамках Недели Науки ИМКиМО.

**Структура работы.** Данная работа состоит из Введения, двух глав, Заключения, Списка используемой литературы и Списка источников фактического материала.

Во **Введении** обосновывается актуальность данного исследования, формулируются его цель, задачи, определяются методы исследования.

В **первой** главе представлена информация об основных положениях корпусной лингвистики, традиционных и корпусных исследованиях, излагается история создания лингвистических корпусов, а также приводится концептуальная классификация корпусов.

Во **второй** главе представлен анализ некоторых полисемантических слов, значения которых были изучены с помощью Корпуса современного американского английского (Corpus of Contemporary American English, COCA) и Британского национального корпуса (British National Corpus, BNC).

В **Заключении** в обобщенной форме излагаются основные результаты исследования.

# Глава 1. Основные теоретические проблемы исследования лингвистических корпусов

## 1.1 Основные положения корпусной лингвистики

В настоящее время с помощью корпусов текстов можно с достаточно большой точностью анализировать многочисленные факты реализации и функционирования языковых единиц. Теоретически, английское слово “corpus” корпус можно расшифровать, как “capable of representing potentially unlimited selections of texts”, что означает – «способный представлять потенциально неограниченный выбор текстов» (Dash, 2005: 35).

Корпусная лингвистика – многоаспектная область. Это область с широким спектром, охватывающая все варианты языка в различных областях языкового взаимодействия, общения и понимания.

Корпусная лингвистика – это подход, который направлен на изучение языка и всех его свойств путем анализа больших наборов текстовых образцов. Этот подход используется в ряде областей исследований на протяжении веков: от описательного изучения языка до языкового образования и лексикографии. В широком смысле это исчерпывающий анализ любого значительного количества аутентичных, устных и/или письменных образцов текста. В целом, он охватывает большой объем машиночитаемых данных о фактическом использовании языка, который включает в себя сборники литературных и нелитературных образцов текста, чтобы отразить как синхронные, так и диахронные аспекты языка. Корпусная лингвистика подходит к изучению языка в корпусах. Корпус представляет собой большой, принципиальный набор естественных примеров языка, хранящихся в электронном виде (Марчук, 2000).

В контексте систематического исследования языка наиболее актуальными являются возможности именно лингвистического корпуса. В данной связи актуализируется целый ряд значимых научных проблем:

определение пределов допустимого доминирования корпусных данных в сфере теоретического обустройства науки о языке; необходимость формализации, репрезентации комплексной семантики языка и речи; учет и использование новой коммуникативной реальности и другие.

Уникальность корпусной лингвистики заключается в использовании современных компьютерных технологий в сборе языковых данных и методов, используемых в информационном поиске.

Основная философия корпусной лингвистики имеет два направления:

а) у нас есть когнитивная установка, чтобы знать, как люди используют язык в своей повседневной коммуникационной деятельности;

б) создать интеллектуальные системы, которые могут эффективно взаимодействовать с людьми.

С этой мотивацией был разработан языковой корпус, который может быть использован для проектирования интеллектуальных систем (например, система машинного перевода, система обработки языка, система понимания речи, текстовый анализ и т.д.). Все отрасли лингвистики и языковые технологии могут извлечь выгоду из анализа корпусов. Таким образом, описание и анализ лингвистических свойств, собранных из корпуса приобретает первостепенное значение во всех областях человеческого знания и применения ([www.cyberleninka.ru](http://www.cyberleninka.ru)).

Можно сделать вывод о том, что корпусная лингвистика включает в себя два аспекта: во-первых, создание и разметка (аннотирование) корпусов текстов и разработка средств поиска по ним и, во-вторых, собственно лингвистический аспект – экспериментальные исследования на базе корпусов.

Подытоживая, отметим, что корпусная лингвистика сделала возможным:

1. Уточнить результаты и выводы проведённых ранее исследований речи.

2. Произвести новые, более широкие и системные (по охвату эмпирического речевого материала) лингвистические исследования.

Важность подобной проблематики заключается в специфике развития современной лингвистики, которую сегодня уже вполне можно дифференцировать на докомпьютерную и компьютерную (интуитивную и эмпирическую, информационную и семантическую и так далее). В настоящее время уже нет сомнений в самостоятельности корпусной лингвистики в структуре языкознания. Имеется достаточно оснований считать компьютерную парадигматику вспомогательной для корпусной методологии.

## 1.2 Традиционные и корпусные исследования

Многие русскоязычные источники свидетельствуют, что первая идея о том, что достоверные лингвистические данные могут быть получены только из значительного массива текстов, была высказана Р. Г. Пиотровским в 60-х годах XX века. Но осмысленные исследования в области корпусов проводились еще в 40-е годы американским лингвистом Л. Блумфильдом и другими учеными-лингвистами.

В 50-60-е годы возобладала концепция Н. Хомского, основателя генеративной лингвистики, и она называлась хомскианская лингвистика (англ. *chomskyan linguistics*). Данная концепция заключалась в том, что необходимо изучать лишь языковое знание или «язык» по Соссюру (*competence*), а не языковое употребление или речь по Соссюру (*performance*). Так как число высказываний естественного языка бесконечно, то и исследовать их бессмысленно. А с другой стороны, количество языковых правил, которые и составляют языковое знание, конечно. Следовательно, их можно и нужно исследовать. Таким образом, произошел уход от эмпирики в сторону рационализма (Баранов, 1998). Тем не менее, некоторые ученые продолжали использовать корпусные методики и в период господства генеративной лингвистики.

В последнее время, одна из основных причин возрастания интереса к корпусным исследованиям – повсеместная компьютеризация. Появление компьютеров сделало возможной обработку огромных массивов текстов. Известные корпусные лингвисты Т. Мак-Эннери и Э. Уилсон считали, что необходимо использовать и эмпирику, и интроспекцию, то есть, использование интуиции носителей языка, и искусственные данные, и естественные. Но из корпуса текстов невозможно извлечь все возможные лингвистические выводы, то есть, можно сказать, что корпус текстов не является самодостаточным (Баранов, 2001).

У. Чейф считает, что корпусный лингвист должен не только описывать явления языка, но и стараться объяснить их (Чейф, 2009). В настоящее время, в центре внимания корпусной лингвистики оказалась языковая личность, то есть, ее речевая деятельность, массовая коммуникация и проблема ее описания.

В. Рыков выделил основные отличия корпусной лингвистики от традиционной:

- Основное внимание уделяется изучению речи, а не изучению языка.
- Корпусная лингвистика в своих исследованиях опирается только на данные корпуса текста, а не идет от теории к ее объяснению и подтверждению фактами речи.
- Корпусная лингвистика предпочитает количественные (количественные) методы.
- Видит себя частью традиций, которые базируются на эмпирических методах, а не на рационалистических.
- В корпусной лингвистике текст рассматривается как некоторая физическая сущность.
- Корпусная лингвистика занимается составлением грамматики конкретных языков, а не языковых универсалий.
- Основное внимание уделяется форме и тексты рассматриваются в глобальной перспективе.

- Корпусная лингвистика фокусирует свое внимание на как можно более широком взгляде на текст, неограниченном ни какими догмами, а в своих выводах опирается на наблюдение речевой деятельности, проявленной в виде текстов.

- Часто пользуется вероятностными методами и статистикой для первичной обработки речевого материала.

- Предпочитает индуктивные методы обработки эмпирического словесного материала и считает их сутью научного метода.

- В корпусной лингвистике проводится работа с лингвистическими данными в том виде, в каком они встречались в контексте ([www.rukov-cl.narod.ru](http://www.rukov-cl.narod.ru)).

### **1.3 История создания лингвистических корпусов**

Люди начали составлять и изучать корпусы ещё до появления корпусной лингвистики, начиная с XVIII века. Примерами могут служить исследования Библии (Cruden и многие другие), составление словарей (Johnson, Oxford English Dictionary, Webster Dictionary), преподавание языков (частотный корпус Thorndike'a, 1921), дескриптивная грамматика (Fries, 1940, Quirk, 1968). Так, корпус Квирка (Survey of English Usage) включал в себя один миллион словоупотреблений и изначально представлял собой один миллион карточек размером 6 на 4 дюйма, 17 строк текста на каждой. Этот корпус стал последним не электронным. Его составление заняло 25 лет, и он был закончен к 1989 году, но технология ушла далеко вперёд. Возникла необходимость срочно переводить корпус в цифровую форму ([lab314.brsu.by](http://lab314.brsu.by)).

В 1960-е годы лингвисты впервые собрали корпуса компьютеризированных текстов. Первый компьютеризированный корпус – Брауновский корпус (The Brown Corpus). Полное название корпуса – The Brown Standard Corpus of American English. Он был разработан в Брауновском университете (Brown University) в США в 1963 году. Он включает в себя 500 текстов из американских газет, журналов, книг, которые были опубликованы в США в 1961 году. Каждый текст в Брауновском корпусе имеет длину 2000 слов, а все собрание текстов составляет 1 млн. слов. Авторами данного лингвистического корпуса являются У. Френсис (W. Francis) и Г. Кучера (H. Kucera), которые включили в него огромное количество материала первичной статистической обработки, такого как частотный и алфавитно-частотный словарь и различные статистические распределения.

Целью создания Брауновского корпуса является обеспечение системного изучения отдельных жанров письменного английского языка и сравнение жанров (Захаров, 2011). Появление этого корпуса вызвало большой интерес у лингвистов и множество оживленных дискуссий. Они коснулись, в первую очередь, принципов отбора текстов и состава решаемых на таком корпусе задач. С одной стороны, Брауновский корпус строился на основе статистических данных, а с другой, статистика применялась в сочетании с ролевыми решениями авторов корпуса, которые были основаны лишь на профессиональной интуиции. Для достижения максимальной объективности этого непростого процесса было необходимо построение наиболее формализованных, прозрачных для проверки и контроля процедур (Рыков, 2002).

Спустя некоторое время европейские исследователи составили корпус текстов следуя тем же принципам. В него входили 15 жанров (регистров), 500 текстов по 2000 слов. Таким образом, он включал в себя 1 млн. слов британского варианта английского языка, и его называли корпусом Ланкастер-

Осло-Берген (The Lancaster-Oslo-Bergen Corpus), по названию британского и двух норвежских университетов.

Сбалансированные корпуса типа Брауновского очень важны для исследователей, занимающиеся лингвистикой и которые хотят использовать корпус в целях лингвистического описания и анализа.

Таким образом, самые ранние большие корпуса – это корпуса письменной речи американского и британского вариантов английского языка. Оба корпуса остаются полезными и сейчас и на них основываются исследования английского языка многих лингвистов.

По прошествии нескольких десятилетий, компьютеры стали намного мощнее и облегчили процесс создания корпусов. Таким образом, современные корпуса могут содержать уже миллиарды слов.

К 1990 году было зафиксировано более 600 компьютерных корпусов и в последующие годы количество и многообразие создаваемых корпусов шло по нарастающей.

Среди современных корпусов английского языка наиболее известны Британский национальный корпус (British National Corpus – BNC), Международный корпус английского языка (International Corpus of English – ICE), лингвистический Банк английского языка (Bank of English), Корпус современного американского английского (Corpus of Contemporary American English – COCA) и другие.

К первой половине 1990-х годов корпусная лингвистика уже окончательно сформировалась как отдельное направление науки о языке. «Корпусная лингвистика достигла зрелости» – так Я. Свартвик озаглавил в 1992 году предисловие к материалам первого Нобелевского симпозиума по корпусной лингвистике (Svartvik, 1992).

## 1.4 Концептуальная классификация корпусов

Поскольку электронный корпус – это новое понятие, ученым еще предстоит достичь общего консенсуса в отношении того, что считается корпусом, и как его следует классифицировать. Некоторые ученые предлагают разумный способ классификации корпусов с четко разделенными категориями. Лингвистические критерии могут быть внешними и внутренними. Внешние критерии в значительной степени отображаются на корпусах из типологии текстов, связанной с участниками, социальной обстановкой, коммуникативной функцией языка и др. Внутренние критерии касаются повторения языковых паттернов внутри фрагмента языка. Учитывая все эти пункты, корпусы можно классифицировать следующим образом: жанр текста, характер данных, тип текста, назначение конструкции и характер применения.

По жанру текста среди корпусов различают письменные, речевые и разговорные корпусы. Письменный корпус (Written Corpus) (например, TDIL Corpus) в силу своего жанра содержит только языковые данные, собранные из различных письменных, печатных, опубликованных и электронных источников. Речевой корпус (Speech Corpus) (например, Wellington Corpus of Spoken New Zealand English) содержит все формальные и неформальные обсуждения, дебаты, беседы, экспромт-анализ, диалоги, монологи, различные типы разговора и т. д. В таких текстах нет сферы участия средств массовой информации. Разговорный корпус (Spoken Corpus) (например, London-Lund Corpus of Spoken English) – это техническое расширение речевого корпуса, содержит тексты разговорного языка. В таком корпусе речь представлена в письменной форме без изменений кроме транскрипции.

По характеру данных корпусы бывают общие, специальные, корпусы подъязыка, примерные корпусы, литературные и мониторные.

Общий корпус (General Corpus) (например, British National Corpus) включает общие тексты, принадлежащие к различным дисциплинам, жанрам, предметным областям и регистрам. Количество типов текста и количество слов и предложений в нем ограничено. Данный корпус имеет возможность пополняться с течением времени, посредством добавления новых данных с наличием новых текстов. Общий корпус богат своим разнообразием и широко используется.

Специальный корпус (Special Corpus) (например, база данных CHILDES) разработан из текстов, отобранных в общий корпус для конкретного языка или диалекта, но с акцентом на исследуемую тему. Специальный корпус не сбалансирован (кроме как в пределах его цели) и, если используется для других целей, то может давать искаженный вид языковых сегментов. Его главное преимущество заключается в том, что тексты подобраны таким образом, что искомые явления встречаются в нем чаще, чем в сбалансированном корпусе.

Корпус подъязыка (Sublanguage Corpus) состоит только из одного текста на конкретном языке. Однородность структуры данного корпуса и специализированная лексика позволяет продемонстрировать хорошие свойства при малом количестве данных.

Примерный корпус (Sample Corpus) (например, Zurich Corpus of English Newspapers) – это одна из категорий специального корпуса, который создан из образцов, содержащих конечный набор текстов, подобранных с особой тщательностью и детально изученных. После того, как образец корпуса разрабатывается, он не изменяется каким-либо образом, потому что любое изменение будет нарушать его структуру. Образцы малы по отношению к текстам, поэтому они не квалифицируются как тексты.

Специальная категория примерного корпуса – литературный корпус (Literary Corpus), который имеет множество видов. Критерии классификации, рассматриваемые для генерации такого корпуса, включают в себя автора, жанр (например, оды, рассказы, выдумки и т.д.), период (например, XV века,

XVIII века и др.), группа (например, поэты-романтики, романисты викторианской эпохи и др.), тема (например, революционные сочинения, семейные повествования, индустриализации и т. д.) и другие вопросы как параметры.

Мониторный (динамический) корпус (Monitor Corpus) (например, Bank of English) представляет собой растущий, не ограниченный набор текстов с возможностью постоянного увеличения объема данных, отражающих изменения в языке. Основой такого корпуса является ссылка на тексты, произнесенные или написанные в течение одного года. Из мониторингового корпуса можно находить новые слова, отслеживать вариации в использовании, наблюдать изменения в значении и получать широкий спектр лексической информации. Со временем баланс компонентов мониторингового корпуса меняется, поскольку появляются новые источники данных.

По типу текста можно выделить одноязычный, двуязычный и многоязычный корпус.

Одноязычный корпус (Monolingual Corpus) (например, Isi Bengali Corpus) содержит репрезентативные тексты одного языка, представляющие его использование в определенный период времени. Он содержит как письменные, так и устные текстовые образцы.

Двуязычный корпус (Bilingual Corpus) (например, TDIL Bengali-Oriya Corpus) формируется при взаимодействии двух родственных или не родственных языков, помещенных в одни рамки. Если эти языки генетически или типологически связаны, они становятся параллельным корпусом, где тексты выровнены по некоторым предопределенным параметрам. Размер, содержание и поле могут варьироваться от корпуса к корпусу, что не допускается в случае параллельного корпуса.

Многоязычный корпус (Multilingual Corpus) (например, Crater Corpus) содержит репрезентативное собрание текстов из более чем двух языков. Как правило, здесь, как и в двуязычном корпусе, используются аналогичные

текстовые категории и идентичные процедуры отбора, хотя тексты принадлежат к разным языкам.

По назначению конструкции можно выделить неаннотированный (неразмеченный) и аннотированный (размеченный) корпус.

Неаннотированный или неразмеченный (Un-annotated Corpus) (например, TDIL Corpus) представляет собой простой текст без дополнительной языковой или неязыковой информации. Он широко используется в изучении языка, но полезность корпуса значительно увеличивается за счет аннотации.

Аннотированный или размеченный корпус (Annotated Corpus) (например, British National Corpus) содержит теги и коды, вставленные извне дизайнерами для записи дополнительной информации (аналитические знаки, части речи, грамматическая информация о категории и т. д.) в текст. В отличие от неразмеченного корпуса, аннотированный корпус более подходит для предоставления актуальной информации, полезной для языковой технологии, включая морфологическую обработку, разбор предложений, поиск информации, смысловую многозначность, машинный перевод и др.

По характеру применения выделяют переводческий, выровненный, параллельный, справочный, сравнительный и оппортунистический (приспособленный) корпус.

Переводческие корпуса (Translation Corpus) обычно состоят из оригинальных текстов исходного языка и их переводов. Эти корпуса, как правило, сохраняют значение и функцию слов и словосочетаний в разных языках и, как следствие, обеспечивают идеальную основу для сравнения реализации конкретных значений в двух разных языках при одинаковых условиях. Более того, они позволяют обнаружить все кросс-лингвистические варианты, т. е. альтернативную визуализацию отдельных значений и понятий. Таким образом, переводческие корпуса более плодотворны для анализа межязыковых данных, так и для формулирования правил, необходимых для перевода (Aijmer, 1991).

Выровненный корпус (Aligned Corpus) (например, The Canadian Hansard Corpus) вид двуязычного корпуса, где тексты на одном языке и их переводы на другой язык соответствуют и следуют предложение за предложением, фраза за фразой или даже слово за словом.

Параллельный корпус (Parallel Corpus) (например, Chemnitz German-English Corpus) содержит тексты и их переводы. Текст на одном языке и его перевод выравниваются. Иногда создаются взаимные параллельные корпуса, в которых используются корпуса, содержащие аутентичные тексты и переводы на каждый из языков.

Справочный корпус (Reference Corpus) (например, Bank of English) создан для того, чтобы предоставлять исчерпывающую информацию о языке. Он достаточно велик, так как представляет все соответствующие разновидности языка. Он может быть использован для написания грамматик, словарей, тезаурусов и других материалов. Он составлен на основе соответствующих параметров, согласованных лингвистическим сообществом. Такой корпус включает устный и письменный, формальный и неформальный язык, представляющий различные социальные и ситуационные регистры.

Сравнительный корпус (Comparable Corpus) (например, Corpus of European Union) представляет собой сборник «аналогичных» текстов на более чем одном языке. Он содержит тексты на разных языках, где тексты не одинаковы по содержанию, жанру или регистру и используется для сравнения различных языков. Сравнительный корпус незаменим для сравнения на разных языках и формирования двуязычных и многоязычных лексиконов и словарей.

Оппортунистический (приспособленный) корпус (Opportunistic Corpus) представляет собой виртуальный корпус, где выбор фактического корпуса (из оппортунистического корпуса) соответствует потребностям конкретного проекта. Мониторный корпус обычно рассматривается как оппортунистический корпус.

Ученые выделяют и другие виды корпусов, такие как закрытый корпус (Closed Corpus), синхронный корпус (Synchronic Corpus), исторический корпус (Historical Corpus), диалектный корпус (Dialect Corpus), идиолектный корпус (Idiolect Corpus) и другие.

## Выводы по Главе 1

Корпусная лингвистика включает в себя два аспекта: во-первых, создание и разметка (аннотирование) корпусов текстов и разработка средств поиска по ним и, во-вторых, собственно лингвистический аспект – экспериментальные исследования на базе корпусов.

Подытоживая данные, отметим, что корпусная лингвистика сделала возможным:

1. Уточнить результаты и выводы проведённых ранее исследований речи .
2. Произвести новые, более широкие и системные (по охвату эмпирического речевого материала) лингвистические исследования.

Важность подобной проблематики заключается в специфике развития современной лингвистики, которую сегодня уже вполне можно дифференцировать на докомпьютерную и компьютерную (интуитивную и эмпирическую, информационную и семантическую и так далее). В настоящее время уже нет сомнений в самостоятельности корпусной лингвистики в структуре языкознания. Имеется достаточно оснований считать компьютерную парадигматику вспомогательной для корпусной методологии.

В последнее время, одна из основных причин возрастания интереса к корпусным исследованиям – повсеместная компьютеризация. Появление компьютеров сделало возможной обработку огромных массивов текстов.

Первым компьютеризированным корпусом является Брауновский корпус, разработанный в Брауновском университете в США в 1963 году. Цель его создания – системное изучение отдельных жанров письменного английского языка, а также сравнение различных жанров.

Самые ранние большие корпуса – это корпуса письменной речи американского и британского вариантов английского языка. Оба корпуса

остаются полезными и сейчас и на них основываются исследования английского языка многих лингвистов.

Среди современных корпусов английского языка наиболее известны Британский национальный корпус (British National Corpus – BNC), Международный корпус английского языка (International Corpus of English – ICE), лингвистический Банк английского языка (Bank of English), Корпус современного американского английского (Corpus of Contemporary American English – COCA) и другие.

Учитывая особенности каждого корпуса, их можно классифицировать следующим образом: по жанру текста, характеру данных, типу текста, назначению конструкции и характеру применения.

По жанру текста среди корпусов различают письменные (Written Corpus), речевые (Speech Corpus) и разговорные корпуса (Spoken Corpus).

По характеру данных корпуса бывают общие (General Corpus), специальные (Special Corpus), корпуса подязыка (Sublanguage Corpus), примерные корпуса (Sample Corpus), литературные и мониторные (Monitor Corpus).

По типу текста можно выделить одноязычный (Monolingual Corpus), двуязычный (Bilingual Corpus) и многоязычный корпус (Multilingual Corpus).

По назначению конструкции выделяют неаннотированный (неразмеченный) (Un-annotated Corpus) и аннотированный (размеченный) корпус (Annotated Corpus).

По характеру применения корпус бывает переводческий (Translation Corpus), выровненный (Aligned Corpus), параллельный (Parallel Corpus), справочный (Reference Corpus), сравнительный (Comparable Corpus) и оппортунистический (приспособленный) корпус (Opportunistic Corpus).

## Глава 2. Языковой корпус как база для исследования многозначных слов

### 2.1 Методы автоматического разрешения многозначности

Научные исследования по разрешению лексической многозначности находятся в поле зрения прикладной и компьютерной лингвистики достаточно давно и имеют многолетнюю историю. С течением лет количество предложенных решений и их эффективность неуклонно росли до тех пор, пока не достигли определённого уровня сравнительно-эффективных показателей точности для определённого спектра слов и типов многозначностей. Полного решения проблема пока не получила, поскольку на пути успешного решения стоит много вопросов, напрямую связанных с языковыми особенностями человеческой речи.

Исходя из определения корпуса текстов, можно предположить, что содержащаяся в нем языковая информация представляет собой богатый материал для исследований в разных областях лингвистики.

Еще в 1986 году М. Леск предложил использовать электронную версию словаря Oxford Advanced Learner's Dictionary of Current English для автоматического определения значения слова в заданном словосочетании. Например, с помощью этой программы можно определить, что в сочетании *pine cone* слово *pine* с наибольшей вероятностью используется в значении «сосна» (а не «ананас», «тосковать», «лимонка» и так далее), а слово *cone* – в значении «шишка» (а не «конус», «пучок», «рожок» и так далее). Это связано с тем, что в словарных толкованиях этих лексем больше всего пересечений (например, *pine* “kind of evergreen trees...” и *cone* “fruit of certain evergreen trees”) (Lesk, 1986).

Другое важное направление, сформировавшееся немного позднее, – использование корпусов текстов. В корпусах разрешение лексической многозначности осуществляется с опорой на сочетаемость компонентов.

Большие объемы современных корпусов позволяют выполнять статистически значимые наблюдения о совместной встречаемости слов в разных значениях. Наилучшие результаты дают аннотированные корпуса, в которых у слов размечены значения. Примером такого корпуса является Британский национальный корпус (British National Corpus).

Чтобы понять масштаб задач, стоящих перед системой разрешения лексико-семантической неоднозначности, надо иметь в виду, что на сегодняшний день в мире насчитывается очень небольшое число корпусов с семантической разметкой. Различия в разметке и в системах автоматического разрешения неоднозначности (WSD – word-sense disambiguation) определяются, прежде всего, тем, каковы потребности пользователей конечного продукта и каким способом (и с какими затратами) разработчики собираются добиться нужного результата. От этого зависят:

- 1) «глубина» различения многозначности;
- 2) выбор словаря или лексической классификации, к которому привязана семантическая аннотация;
- 3) ручной или автоматический способ разрешения неоднозначности.

Например, если конечным результатом является правильная морфологическая разметка текста (POS-tagging), которую затем можно будет использовать в системе машинного перевода, то задачи системы автоматического разрешения неоднозначности ограничиваются снятием частеречной омонимии и вовсе не требуют обращения к семантической многозначности внутри одной части речи.

В последнее время перспективными оказываются разработки, совмещающее в себе использование корпусов текстов и словарей (Roth, 2008). Тем не менее проблема автоматического разрешения неоднозначности не потеряла актуальность, и это может заметить любой пользователь поисковых систем. К основным трудностям, с которыми сталкиваются разработчики относятся:

- отсутствие единообразного описания значений и недостаточная системность существующих лексикографических источников;

- принципиальная недискретность полисемии и неоднородность решений, принимаемых разными лексикографами;

- частое несогласие носителей языка при выделении значений слов и интерпретации конкретных употреблений, затрудняющее работу над созданием аннотированных корпусов текстов.

Как разнообразны бы ни были методы автоматического разрешения многозначности, их объединяет одно: опора на контекст. «Контекст – это единственное средство идентификации значения многозначного слова. Таким образом, вся деятельность по разрешению многозначности опирается на контекст анализируемого слова, в котором и содержится информация, используемая для определения его значения» (Ide, 1998: 18).

Исследований такого рода меньше, чем работ в области автоматического определения значения в контексте. Можно привести характерную цитату из монографии О. Куонг: “Believe it or not, there is literally almost no study in the vast volume of psychology literature which tells us how humans disambiguate word senses. Rather, indirect evidence is found within the broader topic of human lexical processing, which mainly deals with the organization of words and the representation of meanings in the mental lexicon, and the mechanisms of lexical access, especially for words with multiple meanings” (Kwong, 2012: 53) («Трудно поверить, но в литературе по психологии не существует практически ни одного исследования, которое бы показало, как человек определяет значение многозначного слова. Есть лишь косвенные свидетельства в более широкой сфере лексической обработки текстов человеком, в основном связанные с организацией слов и представлением значений в ментальном лексиконе и механизмами лексического доступа, в том числе для многозначных слов»).

Выявление, например, неологизмов в условиях существования текстов в электронной форме технически достаточно легко осуществимо, регулярно

выпускаются словари новых слов, разрабатываются словари неологизмов на основе корпусов, создаются программы по ежедневному мониторингу социальных сетей на предмет выявления неологизмов по ключевым запросам (например, “coined the term jargon for”, “new word”, “slang (expression or phrase) for” и так далее) (Cook, 2012). Процесс выделения у слов новых значений, оказывается значительно сложнее. Тем не менее, в европейской лексикографии и в этой области ожидаются существенные продвижения: в некоторых работах описывается успешное автоматическое выделение новых значений путем сравнения корпусов текстов разных лет и статистического анализа контекстов. Такие употребления начинают появляться и на русском языке, например, в выражениях со словом *vest*. На русский язык слово *vest* в этом значении, следует переводить словом «пояс», то есть выражения *suicide vest* и *explosive vest* целесообразно переводить «пояс шахида» (Lau, 2012).

В настоящее время наиболее известными среди корпусов английского языка являются Британский национальный корпус (British National Corpus). Первоначально он был создан издательством Оксфордского университета в 1980-х-начале 1990-х годов и содержит 100 миллионов слов из широкого спектра жанров (например, устная речь, художественная литература, журналы, газеты и другое). Британский национальный корпус (British National Corpus) связан со многими другими корпусами английского языка. Корпус охватывает британский английский конца XX в. и задуман как образец типичного разговорного и письменного британского английского языка того времени.

Создание корпусов текстов большого объема резко расширило возможности наблюдения и изучения речи как устной, так и письменной. Многие традиционные проблемы языкознания в настоящее время решаются по-новому, причем их решение достигается гораздо легче, а полученные результаты обладают высокой степенью объективности, так как основаны на изучении языкового материала в его реальном контексте.

Одним из очевидных преимуществ использования корпуса текстов в лингвистических исследованиях является удобство и простота его использования, которые достигаются за счет структурности корпуса, т. е. разделения на главы, абзацы, предложения и словоформы, а также его размеченности (tagging, annotation), которая заключается в том, что тексты и отдельные компоненты получают лингвистические (описывающие морфологические, синтаксические, семантические, просодические и прочие характеристики элементов текста) и экстралингвистические (автор, название, год и место издания, жанр, тематика и т. д.) характеристики. Такого рода метаданные во многом определяют возможности корпуса, так как пользователей корпуса, как правило, интересует не содержание конкретных текстов, а их метатекстовая информация и примеры употребления тех или иных языковых элементов и конструкций (Захаров, 2005: 6).

Другой важной характеристикой корпуса текстов является большой объем языкового материала (в настоящее время не меньше 100 млн словоупотреблений), который обеспечивает реальные статистические данные (данные о частоте употребления словоформ, лексем, грамматических категорий), на основе которых можно вывести новые закономерности в языке или проверить ту или иную гипотезу. Более того, только языковая база корпуса дает возможность изучения языковых единиц средней и низкой частотности.

Отметим, что опираясь на данные лингвистического корпуса, основное внимание уделяется работе с языковыми данными (словоупотреблениями) в их реальном контексте и язык описывается в том виде, как он проявляет себя в речи, представленной в специально подобранном корпусе текстов.

Первоначальные лингвистические исследования, проводившиеся с помощью корпусов, сводились к подсчету частотности различных языковых элементов. В настоящее время статистические данные корпуса используются в решении сложных лингвистических задач, например на материале корпуса

статистическими методами можно определить какие слова встречаются вместе регулярно и могут быть отнесены к устойчивым словосочетаниям.

Есть ряд областей, где языковой корпус используется непосредственно. Так, например, языковой корпус используется в языковом описании, исследовании синтаксиса, фонетики и фонологии, интонации, морфологии, лексикологии, семантики, лексикографии, а также в дискурсе, прагматике, языковом планировании, социолингвистике, психолингвистике, семиотике и когнитивной и вычислительной лингвистике.

На самом деле, вряд ли найдется область лингвистики, где корпус не нашел бы своей полезности. Это стало возможным благодаря большим возможностям, которые предоставляет компьютер для сбора, хранения и обработки природных языковых баз данных. Наличие компьютеров и машиносчитываемых корпусов позволило не только быстро и легко получить данные, но и представить их в формате, подходящем для анализа (Захаров, 2005: 5).

Таким образом, языковой материал, представленный в корпусе, служит экспериментальной базой для доказательства лингвистических теорий, особенно в тех случаях, когда корректные лингвистические сравнения возможно успешно провести только на большом объеме речевого материала. Также следует отметить, что языковые данные большинства лингвистических корпусов, так называемых мониторинговых корпусов, постоянно обновляются и расширяются. Это позволяет, с одной стороны, оценить современное состояние того или иного языка, а с другой – проследить изменение частот и контекстов в различные периоды времени и изучить динамику процессов изменения лексического состава языка.

## **2.2 Применение корпусных данных при анализе многозначных слов. Роль контекста при выборе семантического варианта в многозначном слове**

Самым важным преимуществом корпуса текстов является то, что языковые данные разного типа находятся в корпусе в своей естественной контекстной форме. Корпус текстов позволяет по любому слову построить конкорданс – список всех употреблений данного слова в контексте со ссылками на источник. Составление конкорданса дает возможность изучить реальные примеры употребления тех или иных языковых элементов и позволяет выявить особенности сочетаемости (коллокации) языковых единиц. В то время как традиционные грамматики и словари дают информацию о том, что слово должно значить, только корпус текстов, содержащий в себе примеры реального употребления языковых единиц, может дать ответ на вопрос о том, что слово значит на самом деле.

Классическое определение полисемии дает Т.И. Вендина. Исследователь определяет полисемию как наличие у одного слова нескольких связанных между собой значений. Данное определение полисемии в лингвистике и является классическим. Многозначность слова возникает в процессе исторического развития языка, когда слово вследствие семантических переносов, наряду с обозначением одного предмета или явления объективной действительности, начинает использоваться для обозначения другого, сходного с ним по некоторым признакам или свойствам. Потенциально любое слово языка может приобрести новое значение, когда в этом возникает потребность, поэтому многозначных слов в языке, как правило, больше, чем однозначных (Вендина, 2001: 136).

Так, языковой корпус служит уникальной базой данных для любого лингвистического исследования, в том числе для изучения полисемантических

слов, что можно продемонстрировать на примере исследования многозначного прилагательного *sustainable*.

Данное прилагательное обладает абстрактной семантикой и неоднозначностью, а следовательно его трудно адекватно перевести на русский язык. Но с другой стороны, слово *sustainable* обладает высокой частотой употребления в современных англоязычных текстах. По данным Оксфордского этимологического словаря, прилагательное *sustainable*, образованное от глагола *sustain*, имеет французское происхождение и первоначально имело значение «устойчивый, постоянный (able to continue or last for a long time)» ([www.etymonline.com](http://www.etymonline.com)).

В качестве базы языковых данных обратимся к Корпусу современного американского английского (*The Corpus of Contemporary American English*). Проанализировав статистические данные из этого корпуса, можно сделать вывод, что употребление прилагательного *sustainable* наиболее характерно для научного и публицистического стилей. Что касается частоты употребления данного слова, то она, начиная с 1990 года, значительно возросла.

Составление конкорданса по данной лексической единице позволило определить особенности сочетаемости прилагательного *sustainable* и выявить основные сферы его употребления, которые включают в себя следующие области: экология (*sustainable environment*), деятельность человека (*sustainable farming*), человеческие сообщества (*sustainable community*), потребление природных ресурсов (*sustainable energy*) и образ жизни (*sustainable lifestyle*) ([corpus.byu.edu/coca/](http://corpus.byu.edu/coca/)).

Выявление и тщательное изучение синонимов прилагательного *sustainable* позволило разделить их на две группы. К первой группе относятся слова, объединенные значением «экологически безопасный» (*ecological, green, natural, organic*), а вторая группа включила в себя слова с общим значением «жизнеспособный, приемлемый, оправданный» (*viable, workable, defensible, bearable, supportable, maintainable*). Опираясь на семантику

синонимов, можно дать следующее определение прилагательному *sustainable*: рациональный, оправданный, способный существовать не нанося ущерба окружающей среде, обеспечивающий учет будущих потребностей. Причем значение «рациональный», заключенное в прилагательном *sustainable*, понимается как целесообразный с точки зрения безопасности для окружающей среды. Таким образом, самым близким по значению русским переводом данного прилагательного будет «экологически рациональный, обоснованный».

Что касается диверсификации синонимов, так, в отличие от *sustainable*, прилагательные *green*, *organic* и *ecological* в сочетании с существительными, обозначающими продукты, сырье, материалы (*green / organic food*, *green / organic / ecological material*, *green / organic product*, *green / ecological goods*) и существительными, описывающими деятельность человека, (*green business / marketing*, *green housing*, *organic agriculture / gardening*, *ecological engineering / technology / farming*) выражают значение «безопасный для здоровья человека», в то время как прилагательное *sustainable* в комбинациях с теми же существительными (*sustainable farming*, *agriculture*) подчеркивает значение «безопасный для экологии»:

1. The original stated purpose of organic agriculture was not just to protect the quality of foods, but also to safeguard farm environments and communities... ([corpus.byu.edu/coca/](http://corpus.byu.edu/coca/)).

2. ...*sustainable agriculture*... emphasizes a self-sustaining ecosystem in which conditions are right for plants to practically take care of themselves ([corpus.byu.edu/coca/](http://corpus.byu.edu/coca/)).

Однако в таких словосочетаниях, как *sustainable / organic / green / ecological lifestyle*, *green / ecological living* значения синонимичных прилагательных чрезвычайно близки и могут быть определены как «сбалансированный», «гармоничный», «обеспечивающий учет будущих потребностей», «не наносящий вреда окружающей среде и здоровью человека»;

3. ...they get serious about their green home, energy independence and the personal satisfaction of their sustainable lifestyle (corpus.byu.edu/coca/).

4. I would get away from any kind of pollutant environment, carcinogenic and go towards a very healthy organic lifestyle (corpus.byu.edu/coca/).

Как было отмечено ранее, прилагательное *sustainable* несет в себе значение «жизнеспособный, устойчивый к неблагоприятным условиям», «приемлемый». Однако, в отличие от синонимов *viable* и *workable*, семантика прилагательного *sustainable* не ограничена данным значением и всегда несет в себе дополнительное значение «благоприятный», «безопасный для окружающей среды». Например, понятия *sustainable business*, *sustainable industry* трактуются как «деятельность», «производство безопасное для экологии», «экологически чистые, рационально использующиеся природные ресурсы, учитывающие потребности будущих поколений»:

5. ...in the early 90s, when, really, newspapers still were thriving and there was a viable business there and a lot of jobs (corpus.byu.edu/coca/).

6. America will help Afghanistan to develop a stable, free government, an educational system, and a viable economy (corpus.byu.edu/coca/).

7. A sustainable industry uses natural resources wisely and distributes costs and benefits equitably (corpus.byu.edu/coca/).

8. Sustainable industry is a business method that seeks to reduce negative impacts on the environment (corpus.byu.edu/coca/).

Значения, выраженные словосочетаниями *sustainable business*, *sustainable industry*, являются составной частью понятия *sustainable economy*, которое понимается как новый вид экономики, ориентированной на сохранение окружающей среды, использование возобновимых источников энергии и уменьшение негативного влияния человека на экологию:

9. A transition must be made to a sustainable economy, with renewable energy sources and a stable population; but to do so, the perception of our relationship to the Earth and to each other must change (www.natcorp.ox.ac.uk).

Новый тип экономической деятельности может быть сформирован лишь при

условии перехода к новому типу общества, который выражен в понятии sustainable society.

10. As a result, the development of an ecologically literate and sustainable society has emerged in the environmental education ([corpus.byu.edu/coca/](http://corpus.byu.edu/coca/)).

Идея гармонично развивающегося общества, заключенная в данном понятии, выражена на разных уровнях:

- на уровне деятельности человека (sustainable agriculture, sustainable forestry);
- на уровне потребления, использования природных ресурсов (sustainable energy, water, land, wood use);
- на уровне индивидуального образа жизни (sustainable lifestyle, sustainable living).

Новая концепция общества получает широкое освещение не только в научных журналах и отдельных публикациях на тему экологии, но и в различных популярных изданиях, ориентированных на массового читателя, что, несомненно, свидетельствует о том, что идея гармонично развивающегося общества (sustainable society) активно продвигается и в сознании читателя формируется идея нового образа жизни (sustainable lifestyle), ключевой составляющей для которого является гармоничное сосуществование с окружающей средой.

Таким образом, по данным Корпуса современного американского английского, можно обнаружить, что реальное контекстное употребление прилагательного sustainable свидетельствует о поляризации его значения и выделении двух концептуальных компонентов: экологически безопасный и жизнеспособный. При этом в настоящее время более частотным становится употребление прилагательного sustainable в значении «экологически безопасный, учитывающий потребности будущих поколений». Этот феномен прежде всего связан с возрастанием значения экологических проблем для общества, вызванных факторами, связанными с деятельностью человека, не только на уровне стран, но и в масштабе целой планеты.

Как было упомянуто выше, контекст играет огромную роль при выборе нужного значения слова. Используя корпусные данные можно подсчитать, что 500 самых употребительных слов английского языка передают более 10000 значений, причем, чем употребительнее слово, тем больше значений оно имеет. Выбор нужного значения подсказывается окружением слова, то есть контекстом.

Примером можно привести перевод различных словосочетаний с существительным stub: a stub of pencil – огрызок карандаша (“He unfolded it carefully, took a stub of pencil from his ammunition pouch, and marked a cross ...”) (corpus.byu.edu), a stub of candle – огарок свечи (“...he said, handing me a saucer with a stub of candle on it”) (corpus.byu.edu), a stub of a tooth – обломок зуба (“Jack ran to mum, holding in a hand the stub of his tooth”) (corpus.byu.edu), a stub of a check (ticket) – корешок чека (билета) (“Well, I found a check stub that would suggest differently”) (corpus.byu.edu).

Из примеров видно, что все значения многозначного слова связаны одним центральным смыслом, группируются вокруг одного стержня. Центральный смысл (стержневой смысл) – это общий смысл, объединяющий главное значение слова и большинство подчиненных. Центральный смысл в словарях не фиксируется. Определить его можно, проанализировав все значения многозначного слова и значения родственных ему слов. Так, центральным смыслом существительного stub является понятие остаток, часть чего-либо, оставшаяся после удаления большей или главной части. Вокруг этого стержня группируются значения многозначного слова и значения родственных слов.

Анализ другой лексемы – roundabout показал, что данное слово имеет следующие значения:

- 1)участок дороги с круговым движением;
- 2)юбка с фижмами;
- 3)карусель;
- 4)кресло с круглой спинкой;

5) овсяная лепёшка ([www.lingvolive.com](http://www.lingvolive.com)).

По данным Большого оксфордского словаря слово *roundabout* определяется как:

1. A road junction at which traffic moves in one direction round a central island to reach one of the roads converging on it ([en.oxforddictionaries.com](http://en.oxforddictionaries.com)).

2. A large revolving device in a playground, for children to ride on ([en.oxforddictionaries.com](http://en.oxforddictionaries.com)).

3. A revolving machine with model horses or cars on which people ride for amusement ([en.oxforddictionaries.com](http://en.oxforddictionaries.com)).

4. A merry-go-round ([en.oxforddictionaries.com](http://en.oxforddictionaries.com)).

Хотя центральный смысл и не указан в словарной статье, но из примеров перевода видно, что стержнем данного слова является понятие «предмет, имеющий очертание круга».

В Корпусе американского английского (*Corpus of Contemporary American English*) можно встретить следующие примеры со словом *roundabout*: “...who got his idea for a business in a kind of a roundabout way”; “He'd been associating with westerners so much he actually chafed at the roundabout discussion”; “On the screen is a virtual representation of a roundabout”; “But it is also a genre that, in its roundabout way”; “All cardholders pay in a roundabout way for all of the fraud costs a card provider incurs”; “Yes, I suppose it is a bit of a roundabout way”. Анализируя изученные примеры, с данным словом наиболее частотным словосочетанием является “*roundabout way*”, которое используется в переносном значении «обходной путь».

У слова *brown* несколько значений, и все они подчиняются главному значению «коричневый» (“of a colour produced by mixing red, yellow, and blue, as of dark wood or rich soil”), но если мы говорим о цвете кожи, то выбираем значение «смуглый»; «загорелый» (“...with a pleasant, open face, brown skin, and lovely, expertly made-up gray-green eyes...”); о глазах «карий» (“She had his sharper jaw and dark brown eyes flecked with gold”); о виде человека «хмурый, суровый» (“He`s rather brown today”) ([corpus.byu.edu/coca/](http://corpus.byu.edu/coca/)).

Аналогичным образом можно рассмотреть многообразие значений прилагательного *loose*. Оно описывает предмет, который не закреплен с одного конца, и закреплен с другого (при этом иногда закреплен непрочно). В Британском национальном корпусе можно встретить такие примеры: “She wears her hair loose” – Она ходит с распущенными волосами; “a loose button” – пуговица, которая “болтается”, “a loose nail” – гвоздь, который плохо держится, “a loose tooth” – шатающийся зуб, “loose candies” – развесные конфеты, “loose flowers” – одиночные цветы (не в букете), “loose soil” – рыхлая почва, “loose tongue” – болтливый язык; язык без костей, “loose translation” – вольный (приблизительный) перевод, “loose statement” – расплывчатое заявление и другие (corpus.byu.edu).

Из приведенных примеров видно, что многозначность обеспечивает, с одной стороны, экономность и обозримость языка, а с другой стороны, его гибкость и способность обслуживать все потребности в обозначении многообразия познанного человеком мира.

Так, обратившись к семантической структуре прилагательного *cold*, можно обнаружить целый ряд производных значений, каждое из которых семантически связано с ядерным значением “having a low temperature” – “The office always feels so cold first thing on Monday morning” (corpus.byu.edu/coca/). Реализацию производных значений можно проиллюстрировать следующими примерами:

1. Lacking normal human feelings such as sympathy, pity, humor etc.: He is a very cold man, very aloof and arrogant (corpus.byu.edu/coca/).

2. Unfriendly: “Martin was really cold towards me at the party” (corpus.byu.edu/coca/).

3. Cold colour: “The moon shone with a cold, clear light” (corpus.byu.edu/coca/).

Все лексико-семантические варианты данного слова отражают разные грани его семантики, но при этом объединяются в единое целое на основе

семантических ассоциаций, что характерно для полисемантов с радиальным типом связи.

Гораздо сложнее дело обстоит при наличии в слове нескольких номинативных значений с большой степенью семантической самостоятельности. Так, в структуре существительного *state* налицо два равноправных, параллельно представленных номинативных значения – *state* 1 («штат, страна») и *state* 2 («состояние»), между которыми не так просто установить семантические связи и, следовательно, однозначно ответить на вопрос, являются ли данные значения лексико-семантическими вариантами полисемантического слова или же различными словами-омонимами. Данные значения представлены в следующих контекстах:

1. The drought is worse in the central African states ([corpus.byu.edu/coca/](http://corpus.byu.edu/coca/)).

2. They complained about the untidy state the house had been left in ([corpus.byu.edu/coca/](http://corpus.byu.edu/coca/)).

Данные значения семантически весьма удалены друг от друга, и возможность их объединения в рамках одного слова представляется далеко не однозначной.

Отсутствие общепринятых критериев, в соответствии с которыми можно было бы установить, где заканчивается одно слово и начинается другое, отражается в трактовке проблемы полисемии-омонимии в различных словарях. Например, *Cambridge International Dictionary of English* толкует практически все случаи лексической неоднозначности как омонимию, в то время как *Longman Dictionary of Contemporary English* представляет их как полисемию. В этой связи возникает необходимость выработки новых подходов к проблеме разграничения различных видов лексической неоднозначности.

Информация, требующаяся для определения семантического варианта многозначного слова, т.е. необходимая доля контекста, называется «указательным минимумом». Положение его может варьироваться, т.е. «указательным минимумом» может быть, как информация, заключенная в

одном словосочетании, так и сведения о целой эпохе, в контексте которой определяемое слово было употреблено.

Необходимо упомянуть, что понятие контекст делится на три основных типа – узкий, широкий и экстралингвистический. Узкий контекст (микро контекст) – это контекст нескольких лингвистических единиц, таких как словосочетание или предложение.

Например: “Note that when power is first applied to the circuit, there will be a slight delay before the amplifier begins to work” (corpus.byu.edu) – «Обратите внимание, что при первом включении питания на схему, будет небольшая задержка, прежде чем усилитель начинает работать». Наблюдая слово power в языковом окружении терминов, имеющих отношение к электронике (circuit, amplifier), можно сделать вывод о том, что здесь его аналогом в русском языке станет слово питание.

Существуют и конструкции, в которых ядро категорически меняет свое значение под влиянием факторов узкого контекста. Некоторые из таких словосочетаний даже можно отнести к так называемым «ложным друзьям переводчика», поскольку если при переводе допустить, что они употреблены в основном значении, все высказывание сразу же лишится какого-либо смысла. “Power of attorney” буквально переводится как «сила адвоката»; однако при переводе на основе узкого контекста мы получим значение «доверенность». При условии, что оба слова имеют основное предметно-логическое значение, словосочетание “power supply” переводится как «энергетическая поставка/поставка энергии», но в случае сочетания данных слов между собой, их эквивалентом станет словосочетание «источник питания».

Широкий контекст (макро контекст), в отличие от узкого, выходит за рамки предложения. Во внимание принимается целый абзац, глава или все произведение в целом. Следующая цитата принадлежит повести Дж.Р.Р.Толкина «Хоббит или туда и обратно»: “He wanted it because it was a ring of power, and if you slipped that ring on your finger, you were invisible”

(madbook.org) («Это было кольцо Всевластия, и тот, кто надевал его на палец, становился невидимым»). Известно, что слово *power* многозначно. Однако обратившись к широкому контексту произведения, можно сразу понять, что в данном отрывке речь идет именно о власти. Поскольку далее по контексту произведения становится понятно, что такие кольца были выкованы специально для великих владык Средиземья, а то кольцо, о котором идет речь в цитате даже правит и повелевает другими волшебными кольцами. Поэтому общепризнанным переводом Толкиновской фразы “ring of power”, дублирующимся как во всех его книгах, так и в их экранизациях, будет «Кольцо Всевластия».

Нередко встречаются и случаи, когда выбор семантического варианта оказывается невозможен даже в контексте всего произведения, в таких случаях необходимо выйти за его рамки и привлечь так называемый экстралингвистический контекст. Экстралингвистический контекст затрагивает различные внеязыковые факторы — особенности эпохи, местоположения, обстоятельств, к которым относится высказывание. Как правило, экстралингвистический контекст бывает напрямую связан с историей и культурой определенного народа. Так для перевода следующей цитаты требуется информация о фактах, связанных с политической обстановкой в Гане во второй половине XX века. “Ghana agreed to provide cheap power, but the Ghanaians never expected the smelter to take as much as 65 percent of the dam's power” – «Гана согласилась предоставлять дешевую электроэнергию, но жители Ганы не ожидали, что завод возьмет аж 65 процентов мощности дамбы». Здесь речь идет о плотине, построенной в Гане в 1965 году и о заводе американской алюминиевой компании Volta. Эта информация помогает нам определить правильный перевод существительного *power* как «электроэнергия».

Стоит отметить, что часто контекстуальное значение может быть использовано в структуре каких-либо тропов или фигур речи. К примеру, нередко можно встретить контекстуальное значение в метафоричных

высказываниях, например, “the lyrical power of his prose” – «Лирическая сила его прозы.» В данном примере слово power имеет переносное значение, подчеркивающее способность прозы влиять на читателей так же сильно, как может сделать это сила физическая.

Интересно и то, что, как правило, именно на полисемии основывается такой литературный прием, как каламбур (или игра слов). Практически любая многозначная конструкция пригодна для каламбура. Эффект игры слов заключается в контрасте между смыслом многозначных единиц. Примером этому может служить цитата Джими Хендрикса: “When the power of love overcomes the love of power the world will know peace” – «Когда власть любви превзойдет любовь к власти, настанет мир на земле» ([www.aphorisme.ru](http://www.aphorisme.ru)). Стоит отметить, что перевод каламбуров на другие языки практически не представляется возможным, поскольку полные совпадения всей системы лексических значений определенного слова встречаются в разных языках крайне редко.

После перевода и анализа 40 предложений с существительным power, отобранных в случайном порядке системой сайта Британского Национального Корпуса, можно вывести следующую статистику частоты употребления лексико-семантических вариантов этого слова: власть – 51,4%; сила, мощь – 14,2%; держава – 8,5%; электрическое питание, энергия – 11,4%; полномочия – 2,8%; право, возможность – 11,4%. Следовательно, на сегодняшний день значение слова power на основе различных видов контекста чаще всего определяется как «власть».

### **2.3 Концептуальное моделирование процесса разрешения полисемии с использованием корпуса**

Использование данных Британского национального корпуса, анализ случаев употребления как одного и того же, так и разных значений многозначного слова в естественном и максимально широком контексте позволило сделать некоторые предположения касательно особенностей протекания когнитивных процессов.

Когнитивные контексты, ассоциированные с отдельными значениями многозначного слова, не статичные образования; они отражают динамический характер познания в процессе восприятия реальности, приобретение нового опыта и его оценки.

Когнитивный контекст отдельных значений многозначного слова может содержать культурно-специфический компонент. Соответственно, можно вести речь и о наличии индивидуально-личностного элемента когнитивного контекста, так как невозможно не принимать во внимание субъективные факторы прагматического (социального) характера при восприятии когнитивного контекста, ассоциированного с определенным смыслом многозначного слова. Разные носители языка при использовании многозначного слова в одном и том же значении оперируют различным объемом информации, которая фиксирована данным смыслом. В момент формирования когнитивного контекста употребления данного смысла каждый индивид обладает своим представлением о мире, порядке его устройства, его ценностных аспектах, которые были заложены всем его предыдущим опытом.

Так, многозначное слово может присутствовать в ментальном лексиконе, комплексной ментальной репрезентации – наборе когнитивных контекстов, ассоциированных со смыслами многозначного слова, фиксируя при этом значительный объем концептуальной информации, относящейся к

самым разным областям. Модель распознавания значения многозначного слова может быть представлена следующим образом:

- предъявление слова, употребленного в одном из своих смыслов;
- активизация набора когнитивных контекстов, ассоциированных со словом;
- идентификация соответствующего когнитивного контекста (или его элемента);
- идентификация соответствующего значения многозначного слова.

Необходимо заметить, что указанная последовательность когнитивных операций, связанных с формированием, переработкой, хранением и извлечением концептуальной информации, имеет гипотетический характер и является результатом применения метода концептуального моделирования.

Можно привести примеры, которые подтверждают данную гипотезу. Такие примеры были отобраны из Британского национального корпуса. Именно возможности корпуса, фиксирующего фрагменты живой речи, а также корпусной лингвистики, рассматривающей единицы текста в глобальной перспективе позволяют приблизиться к решению некоторых проблем лингвистики, таких, как многозначность. Более того, корпусная лингвистика фокусирует внимание на как можно более широком взгляде на текст, не ограниченном никакими догмами и прибегает к вероятностным методам и обработке эмпирических данных ([www.rykov-cl.narod.ru](http://www.rykov-cl.narod.ru)).

Проанализировать различные пути разрешения многозначности можно на примере существительного *bird*. Данная лексическая единица имеет несколько значений, зафиксированных в Большом оксфордском словаре:

1. A warm-blooded egg-laying vertebrate animal distinguished by the possession of feathers, wings, a beak, and typically by being able to fly ([en.oxforddictionaries.com](http://en.oxforddictionaries.com)).

1.1 A bird that is hunted for sport or used for food ([en.oxforddictionaries.com](http://en.oxforddictionaries.com)).

1.2 An aircraft, spacecraft, or satellite (North American informal) (en.oxforddictionaries.com).

2. A person of a specified kind or character (informal) (en.oxforddictionaries.com).

3. A young woman or a girlfriend (British informal) (en.oxforddictionaries.com).

4. A cry or noise made to express displeasure or contempt (en.oxforddictionaries.com).

5. Badminton equipment consisting of a ball of cork or rubber with a crown of feathers (en.oxforddictionaries.com).

С помощью Британского национального корпуса можно рассмотреть несколько примеров употребления данной лексической единицы в некоторых из указанных значений. Первая группа примеров включает в себя слово *bird* в его исходном значении:

1.a. Glancing up, I saw a beautiful yellow bird perched on a telegraph wire, looking like a prize long-tailed canary (corpus.byu.edu).

b. A bird shuffled along the perch (corpus.byu.edu).

c. A bird tweetered in the gray light through the blinds (corpus.byu.edu).

d. It`s 10 minutes today: less than a cat puts in on a bird (corpus.byu.edu).

e. A bird had gained entry through one of the broken windows and flown helplessly around until it collided with her (corpus.byu.edu).

f. It must be great to be a bird – you know, just flying over people and buildings and that (corpus.byu.edu).

g. I expect he was after a squirrel or a bird up there; he`s a regular hunter (corpus.byu.edu).

Данная группа примеров демонстрирует употребление рассматриваемого существительного в его исходном прототипическом значении. При этом следует упомянуть, что понятие «прототипичность» трактуется по-разному в когнитивной и корпусной лингвистике: в когнитивной лингвистике прототип – наилучший представитель некой

категории, в то время как в корпусной лингвистике это самая частотная единица (Gilquin, 2006: 159).

В примере 1a ментальная репрезентация содержит концептуальную информацию, которая ассоциативно и категориально связана с анализируемой лексической единицей – указание на месторасположение по направлению вверх от говорящего (*glancing up, perched on a telegraph wire*), внешние характеристики (*beautiful and yellow*), отсылку у определенной концептуальной категории (*long-tailed canary*). В сознании человека пробуждаются концептуальные контексты, которые свойственны только одному значению данного слова, то есть исходному значению полисемантического слова. Даже при наличии минимального объема концептуальной информации – “*glancing up, I saw a beautiful and yellow bird ...*” – определение значения данного слова не представляет никакого труда.

В примере 1b существительное *bird* также употреблено в своем исходном значении, как и в случае с 1a. Однако указание на характер передвижения (*shuffled*) формирует совсем иной когнитивный контекст, тоже ассоциативно связанный с птицей, но более крупной, тяжелой, возможно, даже неспособной летать, то есть совершать действие, являющееся одной из основных черт прототипического представления о птице как таковой. Итак, значение одно, оно имеет сходные описания в лексикографических источниках, но объем концептуальной информации, фиксированный данным значением, когнитивные контексты, ассоциированные с ним, различны.

Так же в примерах 1c, 1e и 1g происходит похожий по характеру когнитивный процесс инференции с элементами специализации, то есть сужения первоначального объема концептуальной информации. В данных примерах когнитивные контексты активизируются в сознании и способствуют более точной идентификации, по сути, сужая концептуальную информацию о птицах вообще до представителей определенного класса (типа) птиц.

В примерах 1d и 1f активизируются когнитивные контексты, ассоциируемые с собственно прототипическим значением без сужения объема концептуальной информации, то есть «a warm-blooded egg-laying vertebrate animal distinguished by the possession of feathers, wings, a beak, and typically by being able to fly».

Далее перейдем к рассмотрению второй группы примеров:

2. a. Victoria Wood's quite an attractive bird for a fat lady (corpus.byu.edu).

b. Richie left with a bird yesterday (corpus.byu.edu).

c. Oh look ain't that, that same bird (corpus.byu.edu).

d. He had risen in the City to a financial position which permitted him to park his BMW beside the flash docklands flat, close to his rough but pure roots, in which he kept his nice French bird (corpus.byu.edu).

Во второй группе примеров активизируется комплекс когнитивных контекстов, связанных с категорией «человек» — одушевленный объект, в данном случае женского пола (Victoria Wood, lady), а также фиксирующих некоторые физические (fat) и оценочные (attractive) характеристики объекта. Активизация данных областей в сознании помогает идентифицировать метафорическое, прагматически окрашенное значение — «привлекательная женщина».

В примере 2b происходит активизация когнитивного контекста, отражающего вполне стереотипную ситуацию — мужчина покидает ресторан/кафе/вечеринку в компании привлекательной девушки, которая идентифицируется метафорическим значением существительного bird.

В примере 2с лингвистический контекст употребления слова bird настолько узок, что не дает возможности однозначной идентификации смысла. Однако узнав из корпуса информацию о говорящем — молодом человеке 19 лет, беседующем со своими сверстниками, по фонетическим характеристикам его речи можно предположить, что в данном случае анализируемая лексическая единица использована в значении «молодая привлекательная женщина». Данный вывод можно сделать по стереотипным

представлениям о молодых людях определенного возраста, особенностях их поведения, возможных темах разговоров, то есть исходя из имеющегося у нас конкретного когнитивного контекста, который является частью комплексной ментальной репрезентации многозначного слова, иными словами, на основе ряда прагматических стратификационных характеристик, которые в данном случае и позволили идентифицировать значение слова.

В примере 2d активизируется когнитивный контекст, фиксирующий стереотипную информацию об успешном в финансовом и карьерном плане чиновнике и наборе тех характеристик, которыми такой мужчина должен обладать: престижная и высокооплачиваемая работа (“had risen in the City to a financial position”), элитное жилье (“flash docklands flat”), дорогая машина у дома (“to park his BMW”) и так далее. Когнитивный контекст, который уже перестал быть культурно-специфическим, может вызвать сходные ассоциации у многих людей, которые неоднократно наблюдали такого рода ситуацию. Поэтому из всех значений полисемантического существительного сознание активизирует именно это.

Третья группа примеров демонстрирует еще одно метафорическое значение данного полисеманта, которое не зафиксировано большинством традиционных лексикографических источников, однако присутствует в корпусе, — «летательный аппарат, самолет»:

3. a. How soon could Kirov get his bird off the ground, do you think (corpus.byu.edu)?

b. That's a \*\*\* there is still no other bird that can take off (corpus.byu.edu)!

Ментальная репрезентация в примере 3a содержит разные когнитивные контексты — эпоха (Киров), место (Россия) и действие над объектом — «поднять что-либо в воздух» (“get smth. off the ground”). В результате в сознании активизируется когнитивный контекст, ассоциируемый с метафорическим значением слова — «летательный аппарат, самолет, ракета». Наши фоновые знания о времени события, состоянии научно-технического

прогресса сужают объем значения слова *bird* до значения «самолет», так как конструкции ракет в то время еще не были разработаны.

Во втором примере данной группы активизируется совсем иной когнитивный контекст — невозможность вылететь (*take off*), покинуть аэропорт и вызванное этим раздражение. Стереотипная ситуация, знакомая всем, кто часто совершает перелеты. Данный когнитивный контекст способствует идентификации единственно подходящего значения.

Интересным представляется следующий пример:

4. *Ninety seven there it is now showing, Lot ninety seven, singing bird in a cage I have three hundred offered for this, three hundred pounds and twenty, three fifty, three eighty, four hundred and twenty, fifty four eighty five hundred fifty six hundred and fifty, seven hundred fifty eight hundred eight hundred pounds seated now, any more at eight hundred and fifty nine hundred and fifty nine fifty nine fifty, one thousand new bidder thousand one hundred one thousand one hundred pounds, any more (corpus.byu.edu)?*

В этом случае ментальная репрезентация многозначного слова фиксирует информацию о типичной конвенциональной ситуации — проведение торгов на аукционе. В данном контексте присутствует указание на все необходимые элементы стереотипной ситуации — продавец (ведущий аукционных торгов), покупатели (“*bidder*”), деньги (перечисление предлагаемых сумм) и, наконец, товар-лот (“*bird in a cage*”), обозначенный многозначным словом. Понятно, что речь вряд ли идет о продаже живой птицы, так как стереотипное представление о данной ситуации — проведении аукциона — не ассоциировано в первую очередь с продажей живых птиц. Речь, вероятно, идет о некоем предмете — картине, статуэтке и так далее. Отсутствие категориальной отсылки делает невозможной более точную идентификацию смысла, создавая тем самым эффект неоднозначности.

В результате осуществленного концептуального моделирования процесса разрешения полисемии можно сделать вывод о важной роли

когнитивных контекстов разного типа, являющихся частями комплексной ментальной репрезентации, ассоциированной с определенным значением слова, и способствующих идентификации соответствующего значения.

Таким образом, данные фактического материала продемонстрировали преимущества сочетания методов эмпирической лингвистики, концептуального моделирования и данных корпусов для осуществления анализа такого сложного феномена, как многозначность.

## Выводы по Главе 2

В данной Главе были рассмотрены различные типы контекста и затронуты нестандартные ситуации, возникающие при переводе, а также проведен анализ лексико-семантической структуры различных слов. Данные фактического материала показывают, что роль контекста очень важна при переводе многозначных слов, так как в одних случаях он способствует выбору надлежащего семантического варианта, а в других, только благодаря контексту и возникает определенное значение слова. Таким образом, необходимо всегда обращаться к контексту для раскрытия значения слова и адекватной его передачи в переводе.

Во второй главе на примере английского прилагательного “sustainable” выявляются преимущества использования лингвистических корпусов при изучении полисемантических лексических единиц; рассматриваются сферы его функционирования, особенности сочетаемости и диверсификация его синонимов. Помимо этого, делается попытка определения понятий “sustainable economy”, “sustainable society”, “sustainable lifestyle”, которые отражают актуальные тенденции развития современного общества. Проанализировав с помощью Корпуса современного американского английского данные по контекстному употреблению прилагательного “sustainable” можно выделить два концептуальных компонента: «экологически безопасный» и «жизнеспособный». Наиболее частотным является значение «экологически безопасный», так как в современном мире возрастает значение экологических проблем, вызванных антропогенными факторами в масштабе целой планеты.

На примере перевода различных словосочетаний со словами “stub”, “roundabout”, “brown”, “loose”, “cold” и “state” показан выбор нужного значения, которое определяется с помощью окружения данных слов, то есть контекстом.

Кроме того, представлен анализ различных случаев влияния контекста на выбор семантического варианта многозначного английского существительного “power”, рассмотрены способы перевода этого слова на русский язык в различном грамматико-контекстуальном окружении, а также продемонстрирована статистика использования различных значений существительного “power” в английском языке на сегодняшний день.

Использование корпусов открывает новые возможности исследования когнитивных механизмов формирования структуры многозначного слова, а также идентификации его отдельных значений.

Многозначное слово может присутствовать в ментальном лексиконе, комплексной ментальной репрезентации – наборе когнитивных контекстов, ассоциированных со смыслами многозначного слова. Во второй главе представлена вероятностная модель (алгоритм) концептуальных операций, в результате которых становится возможной идентификация соответствующего значения многозначного слова. Данный алгоритм является результатом применения метода концептуального моделирования. Кроме того, приводятся примеры, отобранные из Британского национального корпуса, которые подтверждают данную гипотезу.

В результате осуществленного концептуального моделирования процесса разрешения полисемии можно сделать вывод о важной роли когнитивных контекстов разного типа, являющихся частями комплексной ментальной репрезентации, ассоциированной с определенным значением слова, и способствующих идентификации соответствующего значения.

## ЗАКЛЮЧЕНИЕ

В рамках данной дипломной работы лексическая многозначность была изучена как лингвистическая проблема.

В первой главе данного исследования было изучено то, что принято понимать под корпусной лингвистикой, а также выявлены признаки уникальности корпусной лингвистики и ее основные отличия от традиционной лингвистики. Кроме того, подробно разобрана и изучена концептуальная классификация корпусов по различным признакам.

Таким образом, была создана теоретическая база для дальнейшего практического исследования во второй главе, где с помощью языковых корпусов, а именно с помощью Британского национального корпуса (British National Corpus) и Корпуса современного американского английского (Corpus of Contemporary American English) проанализированы некоторые многозначные лексические единицы. В ходе этого анализа был установлен ряд наиболее часто употребляющихся переводов данных слов в соответствии с контекстом.

В корпусах разрешение лексической многозначности осуществляется с опорой на сочетаемость компонентов. Большие объемы современных корпусов позволяют выполнять статистически значимые наблюдения о совместной встречаемости слов в разных значениях. Наилучшие результаты дают аннотированные корпуса, в которых у слов размечены значения. Примером такого корпуса является Британский национальный корпус.

В ходе данного исследования было выявлено несколько преимуществ использования корпуса текстов в лингвистических исследованиях:

- 1) удобство и простота его использования, которые достигаются за счет структурности корпуса;
- 2) большой объем языкового материала, который обеспечивает реальные статистические данные, на основе которых можно

вывести новые закономерности в языке или проверить ту или иную гипотезу.

- 3) корпус имеет дело с языковыми данными (словоупотреблениями) в их реальном контексте;
- 4) язык описывается в том виде, как он проявляет себя в речи, представленной в специально подобранном корпусе текстов;
- 5) языковые данные большинства лингвистических корпусов, постоянно обновляются и расширяются, что позволяет оценить современное состояние языка, проследить изменение частот и контекстов в различные периоды времени и изучить динамику процессов изменения лексического состава языка.

Более того, представлена модель распознавания значения многозначного слова, которая носит гипотетический характер и выглядит следующим образом:

- предъявление слова, употребленного одним из своих смыслов;
- активизация набора когнитивных контекстов, ассоциированных со словом;
- идентификация соответствующего когнитивного контекста (или его элемента);
- идентификация соответствующего значения многозначного слова.

Представляется, что результаты данной работы могут быть полезны для лингвистов, которые сталкиваются при переводе с проблемами лексической многозначности.

## Список используемой литературы

1. Баранов, А.Н. Автоматизация лингвистических исследований: корпус текстов как лингвистическая проблема [Текст] / А.Н. Баранов. Русистика сегодня. 1998. № 1 – 179-191 с.
2. Баранов А.Н. Введение в прикладную лингвистику: Учебное пособие [Текст] / А.Н. Баранов — М.: Эдиториал УРСС, 2001. — 360 с.
3. Вендина Т.И. Введение в языкознание: Учебное пособие для педагогических вузов [Текст] / Т.И. Вендина – М., Высш. шк., 2001. – 288 с.
4. Захаров В.П., Богданова С.Ю. 3-38 Корпусная лингвистика: учебник для студентов гуманитарных вузов [Текст] / В.П. Захаров, С.Ю. Богданова – Иркутск: ИГЛУ, 2011. – 161 с.
5. Захаров В.П. Корпусная лингвистика: учебное пособие [Текст] / В.П. Захаров – СПб.: СПбГУ, 2005. – 48 с.
6. Кутузов А.Б. Курс «Корпусная лингвистика» [Электронный ресурс] / А.Б. Кутузов. – URL: <http://lab314.brsu.by/kmp-lite/kmp2/CL/CorporeLingva.pdf>
7. Марчук Ю.Н. Основы компьютерной лингвистики. Учебное пособие [Текст] / Ю.Н. Марчук. Издание 2-е дополненное. – М.: Изд-во МПУ «Народный учитель», 2000. – 226 с.
8. Нагель О.В., Корпусная лингвистика и ее использование в компьютеризированном языковом обучении // Язык и культура. 2008. [Электронный ресурс] / О.В. Нагель. – URL: <http://cyberleninka.ru/article/n/korpusnaya-lingvistika-i-ee-ispolzovanie-v-kompyuterizirovannom-yazykovom-obuchenii>.
9. Рыков В.В., Корпус текстов как реализация объектно-ориентированной парадигмы // Труды Международного семинара Диалог-2002. [Текст] / В.В. Рыков. – М.: Наука, 2002. – 59-61 с.
10. Рыков В.В. Курс лекций по корпусной лингвистике [Электронный ресурс] / В.В. Рыков. – URL: <http://rykov-cl.narod.ru/c.html>

11. Чейф У.Л. Значение и структура языка [Текст] / У.Л. Чейф. Пер. с англ. Изд.3 URSS. 2009. – 424 с.
12. G. Gilquin The place of prototypicality in corpus linguistics // Corpora in cognitive linguistics: corpus-based approaches to syntax and lexis [Text] / G. Gilquin. Berlin, 2006. – 157 p.
13. J.H. Lau, P. Cook, D. McCarthy, D. Newman, T. Baldwin. Word sense induction for novel sense detection [Text] / J.H. Lau, P. Cook, D. McCarthy, D. Newman, T. Baldwin. // Proceedings of the 13th Conference of the European chapter of the association for computational linguistics, 2012. – 103-111 pp.
14. J. Svartvik Directions in Corpus Linguistics. Proceedings of Nobel Symposium 92, Stockholm, 4-8 August 1991 [Text] / J. Svartvik. Berlin: Mouton de Gruyter, 1992. – 158 p.
15. K. Aijmer and B. Altenberg, (Eds.) 1991. English Corpus Linguistics: Studies in Honour of Jan Svartvik [Text] / K. Aijmer, B. Altenberg. London: Longman – 62 p.
16. M. Roth, S. Schulte im Walde. Corpus co-occurrence, dictionary and Wikipedia entries as resources for semantic relatedness information [Text] / M. Roth, S. Schulte. Proceedings of the 6th conference on language resources and evaluation. Marrakesh, 2008. – 203 p.
17. N. Ide, J. Véronis. Introduction to the special issue on word sense disambiguation: the state of the art // Computational linguistics [Text] / N. Ide, J. Véronis. 1998. V. 24. № 1. – 175 p.
18. N. S. Dash: Corpus Linguistics: A General Introduction [Text] / N. S. Dash. CIIL, Mysore, 2010. – 69 p.
19. O.Y. Kwong. New perspectives on computational and cognitive strategies for word sense disambiguation [Text] / O.Y. Kwong. New York, 2012. – 156 p.
20. P. Cook. Using social media to find English lexical blends // Proceedings of the 15th EURALEX International congress (EURALEX 2012) [Text] / P. Cook. Oslo, 2012. – 846 p.

## Список источников фактического материала

1. Афоризмы великих [Электронный ресурс]. – URL: <http://www.aphorisme.ru/by-authors/hendriks/?q=1921>
2. Электронный словарь АБВУ LingvoX5/АБВУ, 2011. [Электронный ресурс]. – URL: <https://www.lingvolive.com/en-us>
3. M. Lesk. Automatic sense disambiguation using machine readable dictionaries: How to tell a pine cone from an ice cream cone [Text] / M. Lesk. Proceedings of SIGDOC-86: 5th International conference on systems documentation. Toronto, 1986. – 24-26 p.
4. British National Corpus [Электронный ресурс]. – URL: <https://corpus.byu.edu/bnc/>
5. Corpus of Contemporary American English [Электронный ресурс]. – URL: <https://corpus.byu.edu/coca/>
6. English Oxford Living Dictionaries [Электронный ресурс]. – URL: <https://en.oxforddictionaries.com/>
7. Oxford Dictionary of English Etymology [Электронный ресурс]. – URL: <https://www.etymonline.com/>
8. J.R.R. Tolkien. “Hobbit” [Электронный ресурс] / J.R.R. Tolkien. – URL: <http://madbook.org/view?book=70&page=1>