

МИНОБРНАУКИ РОССИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«САРАТОВСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ИМЕНИ Н.Г.ЧЕРНЫШЕВСКОГО»

Кафедра английского языка
и методики его преподавания

**ПРИМЕНЕНИЕ КОРПУСНЫХ ТЕХНОЛОГИЙ В СОВРЕМЕННОМ ЛЕКСИКО-
СЕМАНТИЧЕСКОМ АНАЛИЗЕ**

АВТОРЕФЕРАТ ВЫПУСКНОЙ КВАЛИФИКАЦИОННОЙ РАБОТЫ БАКАЛАВРА

Студентки 4 курса 414 группы
направления 44.03.01 Педагогическое образование
профиль – «Иностранный язык»
факультета иностранных языков и лингводидактики

Люкшиной Кристины Сергеевны

Научный руководитель
профессор кафедры английского языка и
методики его преподавания
д.ф.н., доцент

Е.Ю. Балашова

дата, подпись

Зав. кафедрой
английского языка
и методики его преподавания
канд. пед. наук, доцент

Г.А. Никитина

дата, подпись

Саратов 2022 год

Введение. *Актуальность предпринятого исследования* заключается в необходимости исследования перспектив практического применения корпусных технологий для анализа языка и лексико-семантических отношений слов.

Объект исследования: лексические единицы, представленные в корпусах английского и русского языков.

Предмет исследования: минимальные смысловые компоненты лексических единиц, выделяемые с помощью анализа контекста, в котором они употребляются; возможности использования корпусов текста для воспроизведения лексико-семантического и переводческого анализа.

Цель исследования: определение способов применения корпусов текста для лексико-семантического и переводческого анализа текста в рамках современных технологий.

Гипотеза исследования: корпуса текстов являются ценным инструментом анализа текста, поскольку содержат в себе примеры употребления лексических единиц в том или ином контексте, что позволяет не только оценить значение определенной лексической единицы, но наиболее естественные для языка формы употребления.

Задачи исследования:

1. Определить понятие корпуса и изучить его роль в современной лингвистике.
2. Выявить основные принципы лексико-семантического анализа текста.
3. Рассмотреть примеры анализа лексико-семантических единиц с использованием корпусов.
4. Исследовать способы применения корпусных технологий при лексико-семантическом и переводческом анализе текста.

Методы исследования. В соответствии с характером поставленных задач и с учетом специфики проблемы использовались следующие методы исследования: анализ литературы, классификация, анализ материалов исследования, сравнение, обобщение.

Методологическая и теоретическая база исследования. В основу данного исследования легли научные труды таких отечественных и зарубежных исследователей, как А.Н. Баранов, Е.И. Большакова, И.Р. Гальперин, В.П. Захаров, М.В. Копотев, Е.С. Кубрякова, И.А. Мельчук, Н.Е. Сулименко, R. Montague и другие

Материалы исследования. Материал исследования настоящей работы представляет собой корпуса текстов на английском языке (British National Corpus) и русском языке (Национальный корпус русского языка)

Научная новизна исследования данного исследования заключается в определении способов использования корпусов текстов на английском и русском языках для лексико-семантического и переводческого анализа текста.

Теоретическая значимость исследования настоящей работы заключается в описании принципа использования корпусов английского и русского языков для анализа текста, описании возможностей для работы с текстом при помощи использования корпусов и определении роли использования корпусов при анализе текста на английском и русском языках.

Практическая значимость. Практическая значимость настоящей работы заключается в том, что результаты данного исследования можно применять при лексико-семантическом и переводческом анализе текстов с использованием корпусов.

Структура исследования: Настоящая работа состоит из введения, двух глав основного текста, заключения, списка использованных источников и приложения. Во введении обоснована актуальность исследуемой проблемы, сформулированы объект и предмет, определены цель, задачи и методы исследования.

В первой главе данной ВКР рассматривается история возникновения и развития корпусной лингвистики, раскрывается понятие корпуса, приводятся примеры существующих корпусов английского и русского языков и описываются методы лексико-семантического анализа.

Во второй главе данной ВКР на примере некоторых юридических терминов проведен анализ корпусов текстов, выявляются ядерные и периферийные семы слов и рассматривается роль применения корпусов при лексико-семантическом и переводческом анализе текста.

В заключении представлены выводы по проведенному исследованию.

Основное содержание. В первой главе «Корпусная лингвистика как инструмент анализа современного текста» приводится понятие и значение корпуса, классификация корпусов текста, описаны элементы и принципы лексико-семантического анализа.

В настоящее время в связи с тенденциями развития информационного общества все большее значение уделяется направлениям, связанным с современными технологиями. Такие направления распространены практически повсеместно, что оказывает неизбежное влияние и на такую область научных знаний, как лингвистика. Кроме того, еще одной современной тенденцией является объединение разных наук, что способствует развитию новых направлений, как, к примеру, в случае с когнитивной лингвистикой, возникшей из «симбиоза» лингвистической и психологической перспектив.

Как следствие, в науке о языке зародилось новое направление – корпусная лингвистика.

Корпусная лингвистика – это раздел лингвистики, в центре внимания которого находится разработка общих принципов построения и использования лингвистических корпусов (корпусов текстов) с применением компьютерных технологий.

Корпус текста описывается некоторыми исследователями как «большой, представленный в машиночитаемом формате, унифицированный, структурированный, размеченный, филологически компетентный массив языковых данных, предназначенный для решения конкретных лингвистических задач». На данный момент существует множество лингвистических корпусов, среди которых можно назвать Британский национальный корпус, Международный корпус английского языка Национальный корпус русского языка и другие.

Научные исследования в рамках корпусной лингвистики предполагают изучение текста или набора текстов и включают в себя сначала выбор материала, а затем анализ и обработку больших объемов текстов для выявления определенных языковых моделей.

Традиционные лингвистические методы анализа текста позволяют решать все вышеперечисленные задачи, но их низкая эффективность приводит к все более широкому использованию методов компьютерного анализа текста, что позволяет сократить трудозатраты лингвиста, значительно увеличивает объем обрабатываемых данных, а также предотвращает неточности и ошибки в расчетах. Компьютерный анализ текста позволяет создавать лингвистические формулы, основанные не на теоретических, а на эмпирических данных, что является новым шагом в развитии лингвистики как науки и имеет особое значение на современном этапе развития данного направления.

В связи с развитием технологий, что одновременно облегчает и усложняет задачу исследования вследствие изобилия материала, и в целях эффективности изучения языка, корпус текстов должен отвечать определенным требованиям. Так, корпус должен обладать репрезентативностью, что заключающееся в статистически достоверном представлении языка или его части и достигаемое за счет необходимого объема и жанрового разнообразия текстов, сбалансированностью и достаточным объемом.

Существует несколько классификаций типов корпусов. Согласно В. П. Захарову, выделяется три типа корпусов: универсальные корпуса, специфичные корпуса, отражающие проявление культурного или языкового явления, и специфичные корпуса, создающиеся с определенной целью. Согласно А. Н.

Баранову, выделяется пять типов корпусов: исследовательские, иллюстративные, динамические, статические и корпуса параллельных текстов.

Нельзя не отметить, что текст является сложным лингвокультурным образованием, и к его анализу следует подходить ответственно. Так, обычно анализ текста проводится путем разделения на этапы графематического анализа, морфологического анализа, синтаксического анализа, семантического и прагматического анализа, на котором и фокусируется настоящая работа.

В лингвистике существует несколько распространенных моделей, используемых для семантического анализа текста. Среди них можно назвать модель «смысл – текст», модель компонентного анализа, формально-логическую модель, модель, основанную на ролевой структуре предложения, и модель семантических классов. Однако следует понимать, что у каждой из вышеперечисленных моделей есть особенности, которые необходимо учитывать при анализе текста.

Таким образом, возможности и перспективы разработки речевых корпусов объясняются:

- представлением лингвистических данных в реальном контексте;
- достаточно большим объемом текстового материала;
- возможностью многократного использования корпуса для решения различных лингвистических задач, таких как, например, выполнение лексико-грамматического анализа текста и др.;
- способностью производить быстрые и надежные вычисления частоты употребления слов в языке и разделять их по различным параметрам;
- возможностью выбора сочетаний слов, а не самого слова и наличием общих информационных инструментов, которые устанавливают взаимосвязь между вхождениями слов, а также обеспечивают более последовательное распределение словосочетаний.

Лексико-семантический анализ лингвистических корпусов может способствовать более глубокому пониманию языка, включая его функционирование в повседневной жизни при использовании в речи. Переводческий анализ лингвистических корпусов, в свою очередь, позволяет осуществлять перевод наиболее точным и актуальным для языка перевода способом.

Во второй главе «Применение корпусных технологий в современном лексико-семантическом и переводческом анализе текста» приведены примеры использования корпусов текстов с целью выявления ядерных и периферийных компонентов лексических единиц (лексико-семантический анализ текста) и осуществления перевода словосочетаний и предложений, в состав которых входят определенные лексические единицы, с английского на русский язык (переводческий анализ текста).

Цель исследования – произвести лексико-семантический и переводческий анализ лексических единиц с помощью применения корпусов языка, рассмотрев принципы работы с языковыми корпусами и способы их использования в рамках прикладной лингвистики.

Исследование проводилось в три этапа:

- на первом этапе был произведен анализ функционала сервисов, созданных для работы с корпусами (в настоящей работе использовались Британский национальный корпус и Национальный корпус русского языка);
- на втором этапе был произведен лексико-семантический анализ некоторых единиц языка с помощью использования корпусов;
- на третьем этапе с использованием корпусов был произведен переводческий анализ некоторых языковых единиц в контексте.

На первом этапе исследования рассматривались сервис «Sketch Engine» и поисковая система сайта Национального корпуса русского языка.

Сервис «Sketch Engine» предоставляет большое количество возможностей, обеспечивающих удобную работу с корпусом. На главном экране сервиса представлены основные способы его использования. Сервис позволяет:

- осуществлять поиск слов и их сочетаний,
- сравнивать словосочетания с помощью анализа случаев их употребления,
- искать синонимы и похожие слова,
- смотреть случаи употребления языковых единиц,
- просматривать список частоты употребления лексических единиц,
- искать выражения, в состав которых входит большее количество слов,
- выделять слова, относящиеся к терминологии,
- производить диахронический анализ языка, предоставляя список неологизмов,
- определять тип текста и анализировать статистику корпуса,
- создавать автоматический словарь лексических единиц.

Сервис Sketch Engine включает в себя большое количество корпусов текстов, в том числе:

- корпуса на различных языках (китайский, немецкий, французский, итальянский, американо-испанский и т.д.),

- корпуса, основанные на источниках различного вида (веб-корпуса, корпуса субтитров, библейские корпуса и т.д.),
- корпуса устного языка,
- корпуса письменного языка и многие другие.

При поиске определенной языковой единицы так же можно выбрать следующие критерии:

- часть речи,
- тип текста,
- дата публикации,
- домен (для корпусов письменного текста),
- источники (для корпусов письменного текста),
- классификация Дэвида Ли (по типам источников),
- идентификация воспроизводящего текст лица.

Таким образом, данный сервис предоставляет пользователю достаточное количество настроек, с помощью которых можно осуществить анализ языковых единиц, таким образом, выбрав необходимые для конкретного анализа случаи употребления.

Сервис официального сайта Национального корпуса русского языка в сети Интернет также представляет большое количество возможностей. На главной странице сайта представлены вкладки, с помощью которых можно подобрать соответствующие задачам тексты и осуществить поиск с помощью них. Среди типов корпусов представлены следующие:

- основной,
- синтаксический,
- газетный,
- параллельный,
- обучающий,
- диалектный,
- поэтический,
- устный,
- акцентологический,
- мультимедийный,

- мультипарк,
- исторический.

Кроме того, при поиске той или иной лексической единицы на экран выводятся дополнительные настройки, которые позволяют выбрать дополнительные признаки, среди которых:

- повтор (часть речи, падеж, число, время, род, лицо, одушевленность),
- тип оборота (mw:ADV, mw:ADVPRO, mw:CONJ, mw:PARENTH, mw:PART, mw:PR, mw:PRAEDIC, mw:SPRO, mw:degr:max, mw:dir, mw:dist:max, mw:dur:max, mw:ev, mw:ev:neg, mw:ev:posit, mw:loc:body, mw:place, mw:rep, mw:shift, mw:space, mw:speed:max, mw:time),
- положение слова относительно знаков препинания (перед/после любого знака препинания, точки, запятой, двоеточия, точки с запятой, тире, восклицательного знака, вопросительного знака),
- написание слова (с заглавной буквы, в начале предложения, в конце предложения),
- наличие слова в словаре.

Таким образом, данный сервис предоставляет пользователю достаточное количество настроек, удобных и эффективных в том числе при переводческом анализе текста, демонстрируя различный контекст употребления языковых единиц и существующие формы и выражения с данными лексическими единицами.

В рамках настоящего исследования лексико-семантический анализ проводится на примере терминов, относящихся к сфере юриспруденции, а именно contract, trial, justice, crime, procedure. Исследование проводится следующим образом: слово вводится в поисковую строку корпуса, после чего анализируется частота употреблений данного слова, его ядерные и периферийные семы, сочетаемость и др.

В результате анализа были выявлены ядерные и периферийные семы вышеупомянутых слов.

1. Ядерными семами концепта «contract» в английском языке можно назвать term, fulfill, sign, negotiate. Периферийными семами данного концепта в английском языке можно назвать enforce, cancel, money, hire.
2. Ядерными семами концепта «trial» в английском языке можно назвать crime, judge, undergo. Периферийными семами данного концепта в английском языке можно назвать sentence, process, compare, arrest.

3. Ядерными семами концепта «crime» в английском языке можно назвать violation, commit, prevention, deviance. Периферийными семами данного концепта в английском языке можно назвать corporate, organize, syndicate, investigation, statistic.
4. Ядерными семами концепта «justice» в английском языке можно назвать system, truth, ministry, equality. Периферийными семами данного концепта в английском языке можно назвать peace, principle, applicability, review, freedom.
5. Ядерными семами концепта «procedure» в английском языке можно назвать standard, detail, procedure, enforce. Периферийными семами данного концепта в английском языке можно назвать follow, assessment, apply, rule.

Переводческий анализ в данном исследовании был проведен на примере вышеуказанных терминов, относящихся к сфере юриспруденции, а именно contract, trial, justice, crime, procedure.

Для производства переводческого анализа были найдены предложения на английском языке, предоставляющие контекст употребления данных слов. При переводе предложений фокус внимания состоял не только на самой лексической единице, но и на словах, которые сочетаются с ней. В процессе анализа в особые поля вводились нужные значения, определяющие необходимую часть речи и возможное расстояние между словами. Затем был произведен анализ результатов и были определены подходящие варианты перевода лексических единиц в рамках данного в предложении контекста.

Таким образом, корпус текста является важным инструментом прикладной лингвистики. На данный момент представлено множество различных корпусов текста (начиная от дифференциации относительно устного и письменного текста и заканчивая различными жанрами и типами текстов), по причине чего они представляют большую ценность для изучения языка.

Корпусная лингвистика, считаясь относительно новым направлением в лингвистике, экспериментально показала свою актуальность. Корпуса текстов на английском языке предоставляют незаменимую информацию, как для исследования языка, так и для его практического применения, что и было подтверждено в результате проведения исследования на основе лексико-семантического и переводческого анализа лексических единиц.

Заключение. Таким образом, корпусная лингвистика, считаясь относительно новым направлением в лингвистике, экспериментально показала свою актуальность. Корпуса текстов на английском языке предоставляют незаменимую информацию как для исследования языка, так и для его практического применения.

Проведенное исследование позволяет сделать следующие выводы.

1. Современные сервисы по работе с корпусами предоставляют множество возможностей по работе с языковым материалом. Корпуса языков предоставляют актуальный в различных целях контекст тех или иных лексических единиц, являясь собранием разнообразных текстов. Кроме того, функционал корпусов позволяет в весомой степени упростить работу исследователя, поскольку предоставляет возможность выбора стиля и жанра текста, указать необходимую или ожидаемую роль лексической единицы в предложении, показывает статистику употребления синонимичных или контекстно связанных лексических единиц. Так, с помощью исследованных сервисов можно произвести лексико-семантический и переводческий анализ текста или отдельных лексических единиц.
2. Лексико-семантический анализ в рамках исследования языковых корпусов, проведенный на примере лексики из сферы юриспруденции, показал, что корпусы текстов являются ценным инструментом для понимания языка. Контекст употребления лексических единиц и статистика слов, наиболее часто употребляемых в сочетании с данными лексическими единицами, способствует проведению более глубоко лексико-семантического анализа текста.
3. Переводческий анализ в рамках исследования языковых корпусов, проведенный на примере лексики из сферы юриспруденции, показал, что корпусы текстов также являются незаменимым инструментом при переводческой деятельности. Сервисы по работе с корпусами текстов позволяют не только определить сочетающиеся и актуальные в конкретной ситуации лексические единицы, но и определить наиболее точный эквивалент для конкретного переводческого случая, помимо этого предоставляя употребляемое в языке грамматическое оформления сочетаний слов.