Министерство образования и науки

Российской Федерации

ФГБОУ ВО «Тверской государственный университет»

Факультет иностранных языков и международной коммуникации

Направление «Лингвистика»

Кафедра английского языка

РЕФЕРАТ

На тему: «Конкордансер»

По дисциплине «Информационные технологии в лингвистике»

|  |  |
| --- | --- |
|  | Выполнила:  студентка группы 14, дневная форма обучения  Авдеева Вера Павловна |
|  |  |
|  | Проверила:  к.ф.н., доцент кафедры английского языка  Масленникова Е.М. |

ТВЕРЬ

2021

**Конкорданс** – (от нем. konkordanz, англ. concordance – ‘согласие’, ‘соответствие’, лат. conncordare – ‘согласовываться, приводить к согласию’) – особый тип словаря, в котором каждое слово или понятие расположены в алфавитном порядке с минимальным контекстом и всеми случаями употребления в некотором тексте, а также список словоупотреблений с отсылками ко всем контекстам. Кроме того, конкорданс иногда определяют, как словарь сочетаемости языковых единиц или словарь контекстов.

Первый конкорданс был создан к Вульгате в XIII веке Хьюго де Сент-Шером (Hugues de Saint-Cher), которому помогали 500 монахов. Из-за огромного объёма работы в докомпьютерные времена конкордансы были созданы только для важнейших трудов – Библии, Вед, Корана, произведения У.Шекспира.

Конкордансы используются для решения следующих лингвистических задач:

• сравнение различных возможностей использования одного и того же

слова,

• анализ ключевых слов,

• анализ частотности слов и словосочетаний,

• поиск и исследование фраз и идиом,

• поиск перевода, например, терминологии,

• создание списков слов (используется при публикации).

При изучении текста можно выделить поисковую, эвристическую, аналитическую функции конкорданса, а также функции индексации и сравнения.

* Поисковая функция позволяет быстро находить нужный фрагмент текста, используя заданное слово или словосочетание. Данная функция конкорданса используется при подборе цитат и их сверке, при изучении особенностей текста оригинала. Конкорданс может быть использован составителями хрестоматий. Он позволяет искать фразы и идиомы в тексте, осуществлять поиск терминов и идиом в памяти переводов.
* Эвристическая функция конкорданса отличает его от именных указателей наличием контекста. Зачастую, контексты позволяют лингвисту или простому читателю увидеть новое трактование текста. Эвристический потенциал конкордансов состоит именно в способности визуализации сочетаемости слов и их ближайшего окружения.
* Аналитическая функция позволяет проводить анализ различных языковых показателей, таких как лексемы, ключевые слова, частота их употребления в тексте и т.д.
* Функция индексации дает возможность при подготовке текста к публикации создавать индексы и списки слов.
* Функция сравнения применяется при сравнении в тексте всевозможных коннотаций и употреблений слова.

Существуют тематические и многоязычные конкордансы.

**Тематические конкордансы** – это списки тем, которые охватывает книга, с содержанием сути этих тем. Наиболее известный из таких конкордансов – Nave's Topical Bible.

**Многоязычные конкордансы** – конкордансы, основанные на параллельных текстах. В многоязычном конкордансе можно найти переводы определенных слов и словосочетаний как на русский язык, так и с русского языка (например, многоязычные конкордансы Linguee, Reverso).

**Примеры конкордансов:**

Словарь-конкорданс публицистики Ф. М. Достоевского <http://dostoevskii.karelia.ru>

Nave’s Topical Bible <http://www.biblestudytools.com/concordances/naves-topical-bible/>

AntConc <http://www.antlab.sci.waseda.ac.jp/software.html>

GlossaNet <http://glossa.fltr.ucl.ac.be/>

Unitex <http://www-igm.univ-mlv.fr/~unitex/>

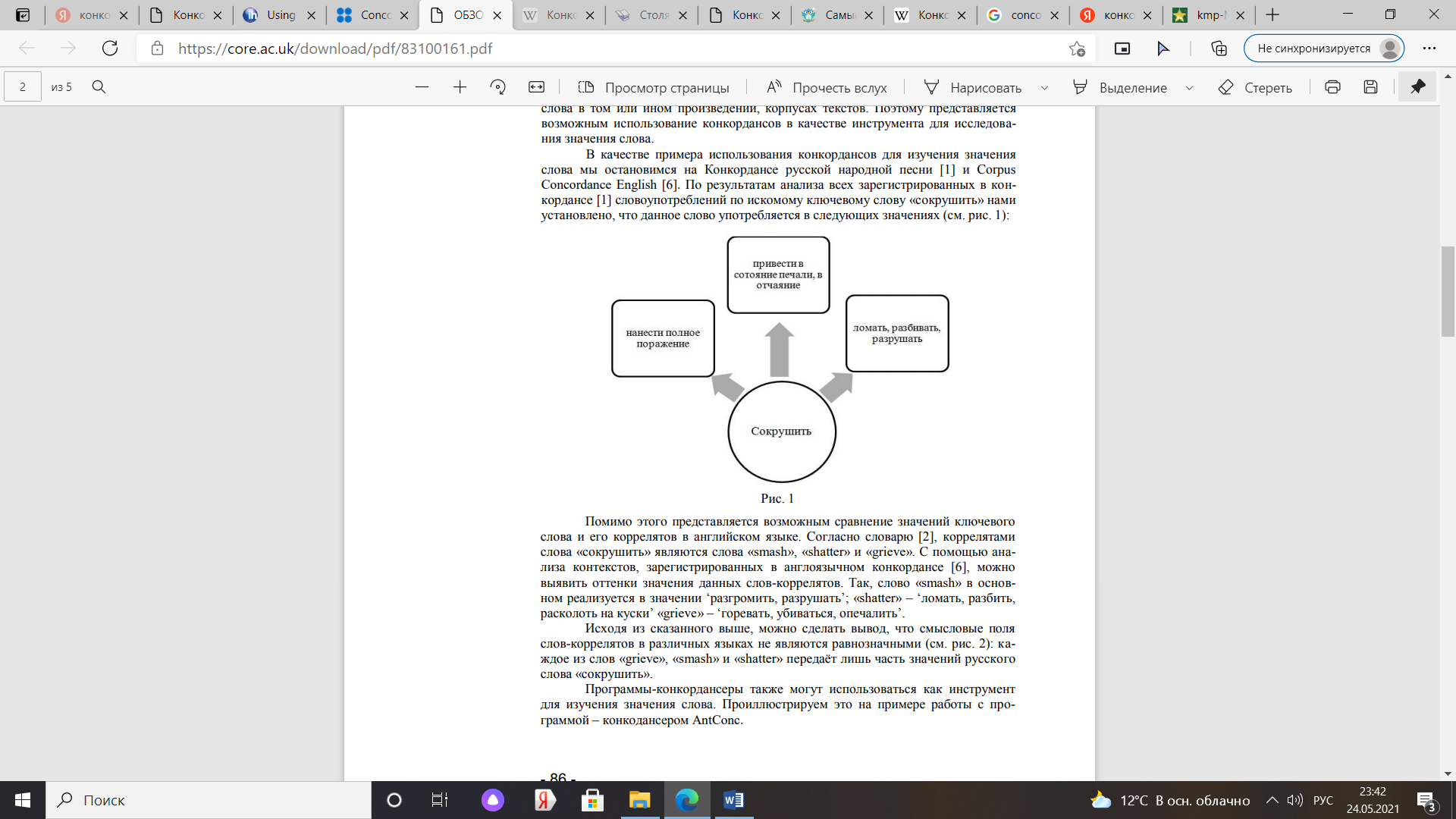
В российской лексикографической традиции практически отсутствуют словари конкордансов, являющиеся важным инструментом изучения языка писателя.

Словари конкордансов включают примеры употребления всех слов в отдельном произведении или во всем творчестве писателя. Известны словари-конкордансы к поэмам О. Мандельштама, к поэзии А. Пушкина. Относительно недавно опубликован конкорданс к русскому варианту библии. Устройство словаря-конкорданса довольно просто. Словарным входом служит словоформа. Она помещается в центр строки и отделяется от текста примера дополнительными пробелами слева (возможно использование и других способов выделения). Пример имеет фиксированную длину, поэтому концы приводимого контекста обычно обрезаются. Пример сопровождается индексом, привязанным к какому-то достаточно полному изданию, что позволяет найти пример в тексте произведения.

Как отмечалось ранее, конкорданс отображает все случаи употребления слова в том или ином произведении, корпусах текстов. Поэтому представляется возможным использование конкордансов в качестве инструмента для исследования значения слова.

В качестве примера использования конкордансов для изучения значения

слова остановимся на Конкордансе русской народной песни и Corpus Concordance English. По результатам анализа всех зарегистрированных в конкордансе словоупотреблений по искомому ключевому слову «сокрушить» установлено, что данное слово употребляется в следующих значениях (рис. 1):

Рис.1

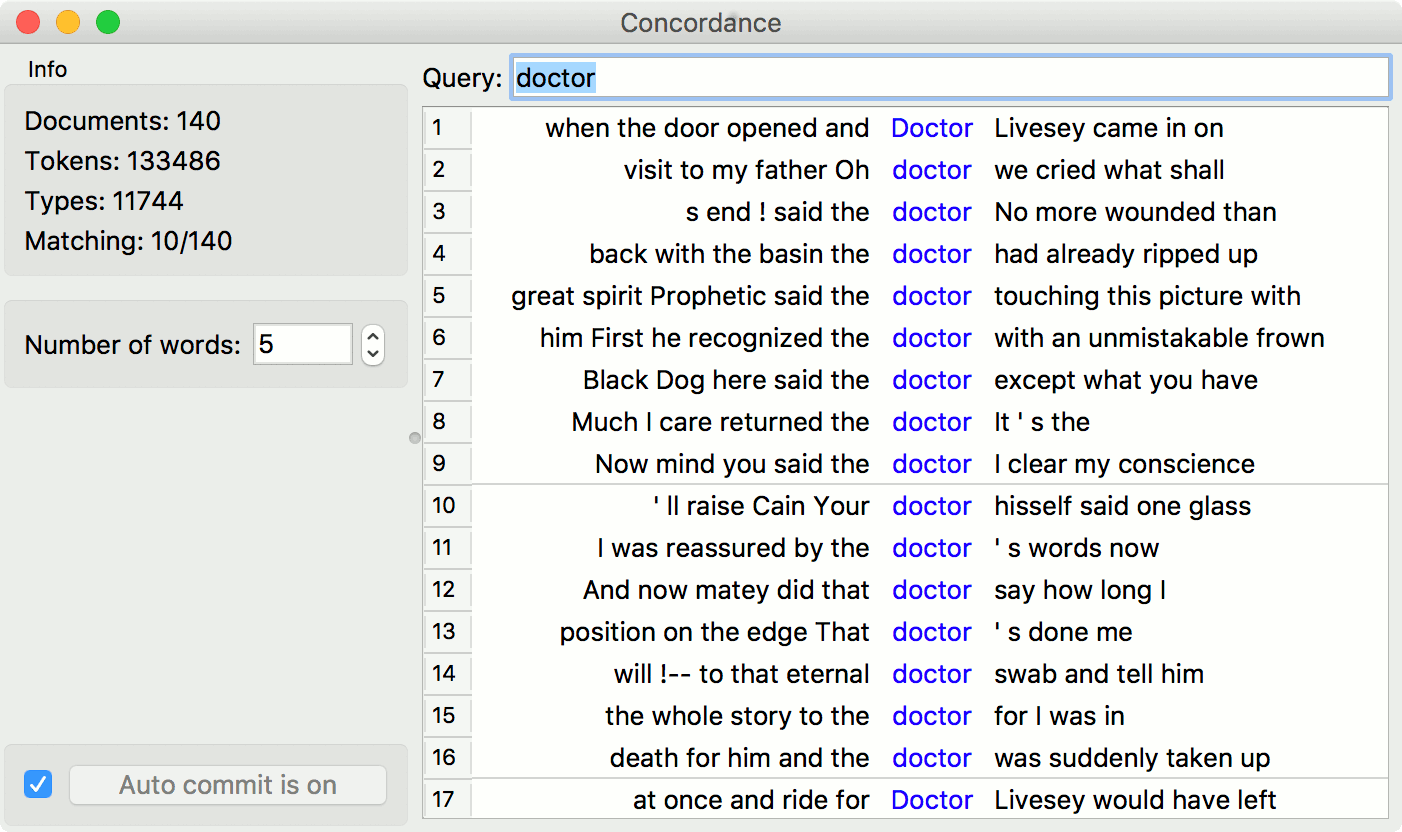
Помимо этого, представляется возможным сравнение значений ключевого слова и его коррелятов в английском языке. Согласно словарю, коррелятами слова «сокрушить» являются слова «smash», «shatter» и «grieve». С помощью анализа контекстов, зарегистрированных в англоязычном конкордансе, можно выявить оттенки значения данных слов-коррелятов. Так, слово «smash» в основном реализуется в значении ‘разгромить, разрушать’; «shatter» – ‘ломать, разбить, расколоть на куски’ «grieve» – ‘горевать, убиваться, опечалить’.

Исходя из сказанного выше, можно сделать вывод, что смысловые поля слов-коррелятов в различных языках не являются равнозначными: каждое из слов «grieve», «smash» и «shatter» передаёт лишь часть значений русского слова «сокрушить».

Существуют специальные программы составления конкордансов по некоторому корпусу текстов, так называемые конкордансеры.

**Конкордансер** (англ. concordancer) (рис.2) – это программное обеспечение, предназначенное для извлечения (и отображения на экране компьютера) всех случаев вхождения слова или фразы в контексте языковых элементов, располагаемых справа и слева от ключевой единицы, на основе загружаемого для целей исследования определенного корпуса текстов. Программа-конкорданс позволяет быстро установить лингвистические условия функционирования той или иной лингвистической единицы и поэтому представляет огромный интерес для лингвиста, лексикографа и переводчика. Программы-конкордансы не заменяют такие средства, как словари и глоссарии, а обеспечивают альтернативный метод обработки текста при переводе. Цель такого поиска – установление структур, которые вызывают при переводе сомнения или требуют уточнения. Для поиска можно вводить часть слова, например, префикс или суффикс, целиком слово или группу слов. Программы-конкордансы предоставляют также статистические сведения: количество слов, классифицируют слова по частотности или в алфавитном порядке и самое важное для переводчика – показывают искомую единицу именно в том контексте, в котором она встречается. Полученная информация может накапливаться и сохраняться по мере перевода текстов. Программы-конкордансы позволяют переводчику получить информацию о той или иной лингвистической единице в тексте любой длины и сложности.

Рис.2



Программа-конкондансер

Следует отметить, что с помощью электронного корпуса текстов можно также получить определённую фоновую информацию по определённой тематике. Часто это играет чуть ли не решающую роль при переводе на иностранный язык. Например, прежде чем приступать к переводу документации нового истребителя, было бы полезно узнать общую информацию о боевых самолетах такого класса. Скорее всего, новый образец самолета основан на одном из предыдущих вариантов и вобрал в себя большинство свойств и качеств последнего. Для этого необходимо составить конкорданс для слова “fighter”. После предварительного анализа предложенных вариантов можно приступить к тщательному просмотру заинтересовавших нас вариантов, нажав на центральное слово. В данном случае стоит отыскать цель, как появляются названия оборудования и детали, характерные для истребителей.

Кроме того, с помощью программы-конкордансера можно отыскать в специализированном корпусе терминологические дублеты, антонимы, референты и номенклатурные наименования, а также расшифровку встретившихся в оригинале сокращений.

Сразу отметим, что эффективность использования корпусов для этих целей зависит от опыта работы с ними. Существует много нюансов, которые необходимо учитывать при работе с электронным корпусом текстов. Один из них заключается в том, что программа находит в корпусе именно ту комбинацию слов, которая вводится в строку поиска, если таковая есть вообще. Любое дополнительное слово в строке поиска или иное расположение слов в корпусе текстов приводит к тому, что программа не может отыскать ни одного совпадения. Это отрицательно сказывается на результатах поиска. Невозможно ввести в поле запроса, например, аббревиатуру для расшифровки и служебную фразу “(is) short for”, так как никто не знает точного их расположения в корпусе, если такой вариант есть.

В. П. Захаров предлагает более полный и конкретный список функций, выполняемых такими поисковыми системами, как конкордансер и корпусный менеджер:

– поиск конкретных словоформ и вывод результатов в виде конкорданса;

– поиск словоформ по леммам;

– поиск группы словоформ в виде разрывной или неразрывной синтагмы;

– поиск словоформ по набору морфологических признаков;

– отображение информации о происхождении, типе текста и т. п.;

– вывод результатов поиска с указанием контекста заданной длины;

– получение различных лексико-грамматических статистических данных;

– сохранение отобранных строк конкорданса в отдельном файле на компьютере пользователя и др.

Система УНИЛЕКС-Т предназначена как для получения традиционных частотных словарей, словоуказателей и конкордансов, так и для формирования базы данных, позволяющей работать со словником и текстами в режиме "запрос - ответ", то есть в режиме автоматического конкорданса. В системе не предусмотрено задание параметров для красивой (полиграфической) печати, поскольку основные пользователи системы - филологи-исследователи, для которых частотные словари, словоуказатели и конкордансы являются не конечным продуктом, предназначенным для издания, а промежуточными рабочими материалами. В частности, результаты обработки текста могут быть использованы в качестве заготовок словарных статей или как картотека при создании, пополнении или коррекции словаря, причем вход в картотеку возможен по разным параметрам: по словоформе, частоте словоформы. Система все же более ориентирована именно на конкордансы, но нет никаких препятствий, которые мешали бы использовать ее только для получения частотных словарей.

Работу системы можно разделить на несколько этапов. Первый - первичная обработка текста, которая заключается в составлении словника по тексту, причем каждому слову приписываются адреса и ссылки на исходный текст. После этого можно либо осуществить лемматизацию, которая припишет словоформам леммы (словарные формы) и части речи, либо обойтись без нее, то есть остаться только со словоформами. Следующий шаг - сортировка словника по алфавиту и подсчет частот для словоформ и, если они есть, для лемм. Если пользователю не нужен диалоговый режим (автоматический конкорданс), то этим можно и ограничиться, разве что воспользоваться еще возможностью отсортировать словник по разным параметрам: по алфавиту (прямой и обратный словарь), по частоте, а если есть леммы и части речи, то по ним. Для получения словоуказателя или конкорданса по некоторому списку слов пользователь должен составить запрос по определенным правилам (если по всему словнику, то запрос не нужен). Элементами запроса могут быть словоформы, леммы, части речи, адреса. Размер контекста задается или числом символов, или числом фрагментов (если текст фрагментирован). Так же можно получить и выборку из частотного словаря, при этом подсчитывается вторая относительная частота - относительно общего объема выборки.

Как упомянуто выше, конкордансеры относят к лингвистическим поисковым системам, как и корпусные менеджеры. Более того, эти два понятия объединены практически одинаковыми функциональными возможностями. Возникает вопрос о том, чем отличаются эти два вида корпусных технологий. В связи с чем, предлагаем рассмотреть определение термина «корпусный менеджер».

Итак, по определению В. П. Захарова, корпусным менеджером (или корпус-менеджером) (англ. corpus manager) называют специализированную поисковую систему управления текстовыми и лингвистическими данными, которая включает программные средства для поиска в корпусе, получения статистической информации и предоставления результатов пользователю в удобной форме.

Также В. П. Захаров наделяет корпусный менеджер как поисковую систему такими характеристиками, как:

– формальная релевантность;

– информационно-поисковый язык фактографического типа;

– НО: умение работать с лексемами и словоформами;

– операции над запросами;

– сравнение с архитектурой поисковых систем в сети Интернет: роботы, программы загрузки индексов, собственно поисковые системы, с которыми работают пользователи. Их автономность (распределенность) и взаимосвязь;

– аналогичная картина в корпусах.

Исходя из вышесказанного, следует заключение о том, что корпусный менеджер стоит на порядок выше конкордансера. Так, например, к задачам простого конкордансера можно отнести построение конкорданса отдельных слов, словосочетаний, частей слов, знаков пунктуации и т. д. в контекстном окружении. Тогда как более сложные программы, которые называются корпусными менеджерами, способны строить полные конкордансы, включающие в себя не только слова, но и другие элементы корпуса. Корпусные менеджеры предназначены для работы с такими явлениями, как лемма и морфологические характеристики слова, позиция слова в предложении и в структуре размеченного текста (HTML, XML), библиографические и типологические признаки документа, из которого выбран контекст (автор, название, источник, год издания, тип текста и т. д.), статистические данные и многим другим.

**Конкондансеры в образовании**

Спектр использования конкордансеров для построения конкордансов безграничен: они могут быть использованы как при обучении профессионально ориентированному иностранному языку на неязыковых факультетах, так и при обучении языку для специальных целей (ЯСЦ) на языковых факультетах в рамках занятий по устной и письменной речи, практикуму по культуре речевого общения на старших курсах бакалавриата, в магистратуре и в системе дополнительного и послевузовского образования (аспирантуре, докторантуре).

Исходя из практического применения конкордансеров в обучении иностранному языку, выделим конкордансеры, основанные на Интернет-технологиях и работающие непосредственно в режиме online («online, web-based»), и стационарные программные средства составления конкордансов («concordancing software»).

Первая группа представлена, во-первых, персональными средствами поиска в сети Интернет («personal Internet search agent»). К ним относится прежде всего KWiCFinder и GlossaNet. Во-вторых, это онлайн конкордансеры, рассматривающие сеть как один большой корпус текстов (концепция «Web as a Corpus»): ВNCWeb, WebCorp, WebCONC, WebKWiC, Linguist’s Search Engine. Особо можно выделить корпусы «Деловая переписка» (Business Letter Corpus) и «Письма личного характера» (Personal Letter Corpus). Оба корпуса оснащены онлайн конкордансерами (<http://ysomeya.hp.infoseek.co.jp>, <http://ysomeya.hp.infoseek.co.jp> ).

Стационарные программные средства составления конкордансов подразделяются на шесть подгрупп: 1) конкордансеры («concordancers» — AntConc, MicroConcord); 2) инструменты для анализа корпусов («corpus analysis tools» — International Corpus of English Corpus Utility Programme (ICECUP), SGML — Aware Retrieval Application (SARA) for British National Corpus, Xaira, WordCruncher, WordSmith Tools); 3) системы анализа («analysis system» — Wmatrix); 4) приложения для составления конкордансов («concordance application» — ConcApp); 5) пакеты программ для составления конкордансов («concordance packages» — Corpus Wizard, Institut für Maschinelle Sprachverarbeitung (IMS) Corpus Work Bench, Multiconcord, Multilingual Concordancer); 6) программы составления конкордансов («concordance programmemes» — Concordance (R. J. C. Watt), Concordancer/Le Concordanceur (D. W. Rand), MonoConc, Oxford Concordance Programme (OCP), ParaConc).

В информационную эру развития компьютерных технологий и широкой доступности интернета открываются новые возможности для повышения эффективности процесса обучения иностранным языкам. Одним из современных и актуальных направлений в методике преподавания иностранных языков является обучение разным аспектам языка на базе корпусов (corpus-based teaching and learning). Корпусные технологии используются при обучении лексике, грамматике, переводу и т. д.

Конкордансы представляют собой специальные программы, предназначенные для обработки текста с той или иной лингвистической задачей, заключающейся в поиске морфем, слов, словосочетаний в контексте. В качестве примера заданного в конкордансе поиска можно привести отслеживание в группе текстов вариантов использования тех или иных грамматических конструкцийили всех слов с определенным окончанием. Конкорданс выдаст все слова с данным окончанием или примеры заданной грамматической конструкции вместе с контекстом. Полученная информация полезна как для преподавателя, так и для студента, поскольку представляет собой множество примеров как грамматической, так и лексической формы слова. Кроме того, благодаря репрезентативности (одного из главных преимущественных свойств лингвистического корпуса), примеры, представленные в конкордансе, позволяют обучаемым проводить собственные исследования, прослеживая изучаемые лексические и грамматические явления в многочисленных естественных языковых контекстах. Таким образом, полученные теоретические знания легче закрепляются на практике, осуществляемой подобным образом. Кроме того, конкордансы могут использоваться при обучении иностранному языку для самоконтроля, а также при выполнении перевода. Например, если студент сомневается в правильности какого-либо выражения или словосочетания, благодаря полученным по его запросу данным в программе-конкордансере, он может выявить по встречаемости введенного выражения или словосочетания, насколько оно правомерно для употребления в данном языке.

Также компьютерные конкордансы внедряются в учебный процесс для подготовки и написания рефератов, курсовых работ по иностранным языкам, а также для научно-исследовательской работы студентов. Возможности конкорданса по сравнению с текстовым редактором значительно шире. Конкорданс анализирует не один, а сразу несколько текстов или корпусы электронных текстов и предоставляет информацию о контексте использования языковых единиц. Кроме того, конкорданс может использовать более сложные, по сравнению с текстовым редактором, формы запроса искомых единиц (например, исключать из поиска определённые слова), выводить на экран информацию о частотности употребления языковых единиц и их сочетаемости в данном корпусе текстов; предоставлять возможность обращения к конкретному тексту, в котором был найден пример; предлагать различные варианты вывода информации (в виде законченных предложений или с пропуском искомой лексической единицы) и др.

Конкордансы на базе корпусов предлагают выбрать такие параметры, как используемый корпус (от самых многочисленных – до ограниченных), конфигурация страницы (выбор количества строк, расположения контекста и ключевого слова) и т. д. Например, при вводе искомого ключевого слова «punishment» с другими заданными параметрами, мы получим иллюстрацию его использования в разных контекстах.

Мульти-конкорданс предлагает также возможность автоматического составления заданий (task-builder option (worksheet or interactive)). При вводе слова или фразы (возможно также введение целого блока лексических единиц), выводятся готовые задания с пропуском для того, чтобы студенты самостоятельно их вставили, тем самым отрабатывая на практике их употребление. Так, например, при обучении студентов института права юридической лексике на английском языке можно воспользоваться конкордансером как для иллюстрации примеров употребления изучаемых лексических единиц, так и для их практического закрепления, выполняя упражнения и тесты, в данном случае – выполняя интерактивный квиз (interactive quiz), который можно выполнить либо в режиме online, либо в форме обычного теста, составленного преподавателем с использованием заданий, автоматически разработанных программой – конкордансером. Обычные задания с пропуском лексических единиц можно также получить с помощью мульти-конкорданса.

Использование корпусных технологий открывает новые перспективы в обучении иностранным языкам, способствуя более эффективному процессу преподавания и обучения, поскольку облегчает работу и преподавателя, и студентов. В частности, как конкордансы, так и корпусные менеджеры могут эффективно использоваться в обучении иностранным языкам, предоставляя студентам возможность самостоятельно осуществлять поиск конкретных словоформ, словоформ по леммам, группы словоформ в виде разрывной или неразрывной синтагмы, словоформ по набору морфологических признаков. Кроме того, данные лингвистические поисковые системы осуществляют отображение информации о происхождении, типе текста и т. п., а также вывод результатов поиска с указанием контекста заданной длины, получение различных лексико-грамматических статистических данных, сохранение отобранных строк конкорданса в отдельном файле на компьютере пользователя и др

**Преимущества конкондансеров в обучении иностранным языкам**

Одним из преимуществ программ-конкондансеров является автономное обучение. Автономное использование конкордансеров помогает ученикам развивать словарный запас, грамматику и жанровые знания в аутентичных, значимых контекстах. Юн (2008) сообщил, что с помощью корпусной технологии студенты осознали свои трудности с написанием текстов на английском языке для конкретных жанров.

Еще одним преимуществом является исследовательское обучение. Хафнер и Кэндлин (2007) обнаружили, что конкондансеры дают мотивацию для непрерывного исследовательского изучения, поскольку удовлетворяет их личное и профессиональное любопытство к новым словам. Студенты могут использовать конкордансер как средство решения проблем, замечая словосочетания, условные обозначения и связи между грамматикой и лексикой в ​​профессиональных текстах. Это также может позволить им перенести знание слов и шаблоны их употребления на чтение.

Конкордансеры могут быть чрезвычайно полезны для создания собственных учебных материалов или для студентов, чтобы они сами проводили некоторые исследования.

Конкондансеры также полезны в понимании значения слов. Например, учитель может собрать примеры предложений из определенного корпуса и попросить учеников сгруппировать их в соответствии со значением.

Список использованной литературы

Вестник ТвГУ. Серия "Филология". 2016. № 4. С. 85–89. А.И. Доминикан «Конкорданс как инструмент для изучения значения слова» (Дата обращения 25.05.2021)

URL: <https://poisk-ru.ru/s13289t1.html> (Дата обращения 25.05.2021)

Using electronic concordancers in the language classroom – URL: <https://ihworld.com/ih-journal/issues/issue-45/using-electronic-concordancers-in-the-language-classroom/> (Дата обращения 25.05.2021)

Concordancers in ELT – URL: [Concordancers in ELT | TeachingEnglish | British Council | BBC](https://www.teachingenglish.org.uk/article/concordancers-elt) (Дата обращения 25.05.2021)

Википедия – URL: [Конкордансер — Википедия (wikipedia.org)](https://ru.wikipedia.org/wiki/%D0%9A%D0%BE%D0%BD%D0%BA%D0%BE%D1%80%D0%B4%D0%B0%D0%BD%D1%81%D0%B5%D1%80) (Дата обращения 25.05.2021)

URL: <http://lab314.brsu.by/kmp-lite/kmp2/Translation/MT/MT-Corpus/MT-Corpus.htm> (Дата обращения 25.05.2021)

URL: <https://human.snauka.ru/2017/02/21074> (Дата обращения 25.05.2021)

«Корпусные технологии в методике преподавания иностранных языков» М. С. Мальцева: <https://cyberleninka.ru/article/n/korpusnye-tehnologii-v-metodike-prepodavaniya-inostrannyh-yazykov> (Дата обращения 31.05.2021)

«Использование конкордансеров при обучении английскому языку для специальных целей» А. И. Глазырина: <https://cyberleninka.ru/article/n/ispolzovanie-konkordanserov-pri-obuchenii-angliyskomu-yazyku-dlya-spetsialnyh-tseley> (Дата обращения 31.05.2021)