



МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное автономное образовательное учреждение высшего образования

«Дальневосточный федеральный университет»
(ДФУ)

**ВОСТОЧНЫЙ ИНСТИТУТ – ШКОЛА РЕГИОНАЛЬНЫХ И
МЕЖДУНАРОДНЫХ ИССЛЕДОВАНИЙ**

СОГЛАСОВАНО

Руководитель ОП

—  — Спицына Н.А.
(г. Владивосток) (ФИО)

УТВЕРЖДАЮ

Заведующая кафедрой
лингвистики и межкультурной коммуникации

—  — Ловцевич Г.Н.
(подпись) (ФИО)

«21» января 2021 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Технологии корпусной лингвистики

Направление подготовки 45.03.03. Фундаментальная и прикладная лингвистика

(Лингвистика в информационно-коммуникационной цифровой среде)

Форма подготовки очная

курс 3 семестр 5

лекции 16 час.

практические занятия 0 час.

лабораторные работы 16 час.

в том числе с использованием МАО 6 час.

всего часов аудиторной нагрузки 32 час.

самостоятельная работа 40 час.

в том числе на подготовку к экзамену 27 час.

контрольные работы (количество) не предусмотрены

курсовая работа / курсовой проект не предусмотрены

зачет _ семестр

экзамен 5 семестр

Рабочая программа составлена в соответствии с требованиями Федерального государственного образовательного стандарта по направлению подготовки 45.03.03. Фундаментальная и прикладная лингвистика утвержденного приказом Министерства образования и науки РФ от 24.04.2018 г. № 323

Рабочая программа обсуждена на заседании кафедры лингвистики и межкультурной коммуникации протокол № 5 от «21» января 2021 г.

Заведующий кафедрой лингвистики и межкультурной коммуникации Г.Н. Ловцевич

Составители: доцент Спицына Н.А.

Владивосток
2021

I. Рабочая программа пересмотрена на заседании кафедры:

Протокол от « _____ » _____ 20__ г. № _____

Заведующий кафедрой _____
(подпись) (И.О. Фамилия)

II. Рабочая программа пересмотрена на заседании кафедры:

Протокол от « _____ » _____ 20__ г. № _____

Заведующий кафедрой _____
(подпись) (И.О. Фамилия)

III. Рабочая программа пересмотрена на заседании кафедры:

Протокол от « _____ » _____ 20__ г. № _____

Заведующий кафедрой _____
(подпись) (И.О. Фамилия)

IV. Рабочая программа пересмотрена на заседании кафедры:

Протокол от « _____ » _____ 20__ г. № _____

Заведующий кафедрой _____
(подпись) (И.О. Фамилия)

Цели и задачи освоения дисциплины:

Цель: познакомить обучающихся с основами корпусной лингвистики в рамках эмпирического подхода к изучению языка, с концепциями корпусной лингвистики, дать им возможность освоить основы корпусных технологий, приобрести навыки работы с корпусами, определить место дисциплины в ряду дисциплин лингвистического цикла, показать теоретическое и практическое значение корпусов для проведения подлинно научных лингвистических;

Задачи:

- ознакомить с новой парадигмой в лингвистических исследованиях;
- ознакомить с предметами, задачами, методами и понятийным (терминологическим) аппаратом корпусной лингвистики;
- ознакомить с историей корпусных исследований;
- сформировать знания о компьютерном инструментарии корпусной лингвистики;
- ознакомить с практикой составления корпусов, сферами применения корпусной лингвистики;
- способствовать освоению обучающимися методики работы с корпусными базами данных в целях лингвистического анализа;
- способствовать изучению языковых и программных средств корпусной лингвистики;
- сформировать навыки работы с программными средствами и информационными ресурсами корпусной лингвистики.

Профессиональные компетенции выпускников и индикаторы их

достижения:

Тип задач	Код и наименование профессиональной компетенции (результат освоения)	Код и наименование индикатора достижения компетенции
Технологическая сфера компьютерной лингвистики	ПК-9Владеет принципами создания электронных языковых ресурсов (текстовых, речевых и мультимодальных корпусов; словарей, тезаурусов, онтологий; фонетических, лексических, грамматических и иных	ПК -9.1Знает основные принципы обработки информации; базовые принципы корпусной лингвистики, лексикографии, математической статистики; базовые представления о языковом разнообразии; наиболее полные и значимые лингвистические корпуса, электронные словари и базы данных.
		ПК -9.2Умеет пользоваться основными методами, способами

Тип задач	Код и наименование профессиональной компетенции (результат освоения)	Код и наименование индикатора достижения компетенции
	баз данных и знаний) и умеет пользоваться такими ресурсами	исредствамиполучения, хранения, переработки информации; пользоваться лингвистически ориентированными программными продуктами
		ПК -9.3Имеет практический опыт разработки электронных языковых ресурсов; опыт применения основных методов, способов и средств получения, хранения, переработки информации

Код и наименование индикатора достижения компетенции	Наименование показателя оценивания (результата обучения по дисциплине)
ПК -9.1Знает основные принципы обработки информации; базовые принципы корпусной лингвистики, лексикографии, математической статистики; базовые представления о языковом разнообразии; наиболее полные и значимые лингвистические корпуса, электронные словари и базы данных.	Знает основные корпуса и положения корпусной лингвистики.
	Умеет пользоваться корпусами и корпусными менеджерами.
	Владеет навыками составления корпусов.
ПК -9.2Умеет пользоваться основными методами, способами исредствами получения, хранения, переработки информации; пользоваться лингвистически ориентированными программными продуктами	Знает основные принципы устройства корпуса.
	Умеет искать информацию в корпусе.
	Владеет навыками алгоритмизации составления корпуса.
ПК -9.3Имеет практический опыт разработки электронных языковых ресурсов; опыт применения основных методов, способов и средств получения, хранения, переработки информации	Знает основные принципы разработки электронных языковых ресурсов.
	Умеет применять основные методы, способы и средства получения, хранения, переработки информации в лингвистическом корпусе.
	Владеет навыками переработки языковых данных для составления лингвистического корпуса.

1. Трудоёмкость дисциплины и видов учебных занятий по дисциплине

Общая трудоёмкость дисциплины составляет __2__ зачётных единиц (__72__ академических часа).

(1 зачетная единица соответствует 36 академическим часам)

Видами учебных занятий и работы обучающегося по дисциплине могут являться:

Обозначение	Виды учебных занятий и работы обучающегося
Лек	Лекции
Лаб	Лабораторные работы
СР	Самостоятельная работа обучающегося в период теоретического обучения
Контроль	Самостоятельная работа обучающегося и контактная работа обучающегося с преподавателем в период промежуточной аттестации

Структура дисциплины:

Форма обучения – ____ очная _____.

№	Наименование раздела дисциплины	Семестр	Количество часов по видам учебных занятий и работы обучающегося						Формы промежуточной аттестации
			Лек	Лаб	Пр	ОК	СР	Контроль	
1.	Раздел 1. Введение в корпусную лингвистику	5	6	6	-	-	6	7	экзамен
2.	Раздел 2. Корпусные лингвистические исследования	5	10	10	-	-	7	20	экзамен
	Итого:		16	16			13	27	

3. СТРУКТУРА И СОДЕРЖАНИЕ ТЕОРЕТИЧЕСКОЙ ЧАСТИ КУРСА

Раздел 1. ВВЕДЕНИЕ В КОРПУСНУЮ ЛИНГВИСТИКУ (6 ЧАС.)

Тема 1: предмет, цели и задачи корпусной лингвистики (2 час., в том числе 2 часа с применением МАО)

Введение в корпусную лингвистику. Понятие лингвистического корпуса. История корпусной лингвистики. Корпусный (эмпирический) подход в сравнении с хомскианской лингвистикой. Задачи и основные направления корпусной лингвистики. Корпусная лингвистика и компьютерная лингвистика. Предмет исследования.

Тема 2: методика лингвистического корпуса (2 час.)

Задачи: создание корпусов и лингвистические исследования на их базе, объективное лингвистическое описание языковой системы. Направления: лексикографические исследования, исследования лексического состава языков, исследования относительных частот употребления тех или иных слов, изучение изменений в лексическом составе языков, различных его вариаций, изучение грамматики естественных языков. Корпусная

лингвистика занимается в основном описанием и объяснением сущности, структуры и использования языка. Компьютерная лингвистика выступает для корпусной в качестве «поставщика» инструментов анализа и обработки корпусов. Понятие лингвистического корпуса. Создание и типология корпусов. Развитие лингвистических корпусов в мире. Первое и второе поколение корпусов. Разметка. Виды разметок.

Тема 3: типы корпусов (2 час., в том числе 2 часа с применением МАО)

История лингвистических корпусов: от картотеки к корпусу. Корпусная лингвистика в России. История электронных лингвистических корпусов. Четыре группы корпусных лингвистов. Первое поколение корпусов: TheBrownCorpus, Lancaster-Oslo/Bergen (LOB) Corpus, London-LundCorpus (LLC). Второе поколение корпусов: The Cobuild Project / The Bank of English, The Longman Corpus Network, British National Corpus (BNC), The International Corpus of English (ICE), American National Corpus, Gigaword corpora. Типология корпусов. Типы корпусов: устные и письменные, одноязычные и многоязычные, аннотированные и неаннотированные.

РАЗДЕЛ 2. КОРПУСНЫЕ ЛИНГВИСТИЧЕСКИЕ ИССЛЕДОВАНИЯ (10 час.)

Тема 4: лексические корпусные лингвистические исследования(2 час.)

Понятие разметки. Типы разметки. Автоматический морфологический и синтаксический анализ. Металингвистическая разметка. Параллельные корпусы. Проблема выравнивания. Языковые средства представления размеченных текстов Экстралингвистическая разметка. Методы и средства разметки и создания корпусов. Лингвистические исследования на базе корпуса: изучение лексики. Лингвистические исследования на базе корпуса: изучение грамматики.

Тема 5: узус в корпусе(2 час.)

Лингвистические исследования на базе корпуса: изучение лексики, изучение грамматики. Распределение и функции номинализаций. Распределение грамматических категорий. Обзор существующих корпусов различных типов. Сравнительный анализ корпусов и корпусных менеджеров. Брауновский корпус. Британский национальный корпус. Корпус современного американского английского. Русско-английский корпус С. Шарова. Технология создания корпусов. Классификация (типология) корпусов по различным основаниям. Типы корпусов по задачам. Типы

корпусов по формальным признакам. Корпус первого типа. Корпус второго типа. Лингвистические исследования на базе корпусов. Исследование нормы/узуса. Социолингвистические исследования. Возможности преподавания иностранного языка на основе корпусной лингвистики

Тема 6: практические сферы применения корпусов: определения и основные понятия (4 час.)

Лингвистические исследования на базе корпусов. Стилистические исследования. Лингвистические исследования на базе корпусов. Практические сферы применения: установление плагиата.

Тема 7: Программы для работы с корпусом (2 ч.)

Семантическое наполнение Британского национального корпуса. Методы извлечения информации из корпуса. Типы извлекаемой информации. Конкорданс. Программы для работы с корпусом: конкордансеры и другие. Параллельные корпуса. Использование корпусов в других науках. Статистические меры вычисления совместной встречаемости. Методы извлечения информации из корпуса. Типы извлекаемой информации. Конкорданс. Программы для работы с корпусом: конкордансеры и другие. Параллельные корпуса.

4. СТРУКТУРА И СОДЕРЖАНИЕ ПРАКТИЧЕСКОЙ ЧАСТИ КУРСА

ЛАБОРАТОРНЫЕ РАБОТЫ

ЛАБОРАТОРНАЯ РАБОТА 1. 2 час. Введение в корпусную лингвистику.

Понятие лингвистического корпуса. История корпусной лингвистики. Корпусный (эмпирический) подход в сравнении с хомскианской лингвистикой. Задачи и основные направления корпусной лингвистики. Корпусная лингвистика и компьютерная лингвистика. Предмет исследования.

В ходе лабораторной работы студенты отвечают подготовленный теоретический материал и выполняют индивидуальные практические задания.

Задание 1.

Показатели	Традиционная лингвистика	Корпусная лингвистика
Объект изучения		
Цель		
Методы		
Материал		
ресурсы		

Связь с дисциплинами		
История возникновения		

Заполните таблицу, сопоставив корпусную и традиционную лингвистику.

ЛАБОРАТОРНАЯ РАБОТА 2. 2 час. Понятие корпуса.

Определение лингвистического корпуса. Отличие корпуса от электронной библиотеки. История создания лингвистических корпусов. Основные понятия. Лингвистические и нелингвистические корпуса. Проблема репрезентативности и сбалансированности корпуса. Развитие лингвистических корпусов в мире. Первое, второе и третье поколение корпусов.

В ходе лабораторной работы студенты отвечают подготовленный теоретический материал и выполняют индивидуальные практические задания.

Задание 2. Дайте характеристику корпусов в соответствии с классификацией: А) Компьютерный корпус текстов русских газет конца XXого века (http://ns2.philol.msu.ru/~lex/corpus/corpus_descr.html);

Б) Корпус современного американского английского - Corpus of Contemporary American English (COCA) (<http://corpus.byu.edu/coca/>);

В) Национальный корпус русского языка (<http://www.ruscorpora.ru/>).

Классификация корпусов
 Признак Типы корпусов Тип данных
 Письменные Речевые Смешанные Язык текстов Русский Английский и т.д.
 «Параллельность» Одноязычные Двужязычные Многоязычные
 «Литературность», специфичность Литературные Диалектные
 Разговорные Терминологические Смешанные Жанр Литературные
 Фольклорные Драматургические Публицистические Доступность
 Свободно доступные Коммерческие Закрытые Назначение
 Исследовательские Иллюстративные Динамичность Динамические (мониторные)
 Статические Разметка Размеченные Неразмеченные
 Характер разметки Морфологические Синтаксические Семантические
 Просодические и т.д. Объем текстов Полнотекстовые «Фрагментнотекстовые»
 Хронологический аспект Синхронические
 Диахронические «Общность» Общие
 Одного писателя.

ЛАБОРАТОРНАЯ РАБОТА 3. 2 час. Разметка корпуса.

Виды разметок. Экстралингвистическая метаразметка. Лингвистическая разметка. Автоматическая морфологическая и синтаксическая разметка.

Снятие неоднозначности. Лемматизация. Графематическая, дискурсивная, семантическая разметка. Жанровая разметка и другие виды разметки.

В ходе лабораторной работы студенты отвечают подготовленный теоретический материал и выполняют индивидуальные практические задания.

Задание 3. Познакомьтесь с материалом о становлении корпусной лингвистики и заполните таблицу.

Первое поколение корпусов 1. The Brown Corpus Точное название: Brown University Standard Corpus of Present Day American English. Составлялся с 1961 по 1964 год. Язык корпуса: американский английский, письменные тексты, 1 миллион словоупотреблений (это количество стало фактическим стандартом для всего первого поколения корпусов). В то время в лингвистике доминировала концепция Хомского, так что Nelson Francis и Henry Kucera (создатели Брауновского корпуса) делали свою работу в очень неблагоприятной атмосфере. Корпус состоит из 500 текстов по 2000 слов каждый. Фактически, он задал стандарт для корпусных исследований, поскольку была очень хорошо продумана структура и выбор категорий текстов. Этот же проект 4 установил традицию свободного доступа к корпусам для исследовательских нужд. На этом корпусе уже в 1969 году был основан словарь American Heritage Dictionary. В конце 1970-х гг. корпус был дополнен разметкой частей речи и морфологических признаков слов. Свободный доступ к корпусу возможен с сайта университета Лидс. Также свободный доступ к корпусу предоставляется через LDC (Linguistic Data Consortium) по адресу: <http://wave ldc.upenn.edu/cgi-bin/ldc/textcorpus?doc=yes&corpus=BROWN>

. Кроме того, Брауновский корпус распространяется на платной основе Международным компьютерным архивом современного английского языка (ICAME) (Берген, Норвегия). 2. Lancaster-Oslo/Bergen (LOB) Corpus 1970-78 год, проект университетов Ланкастера и Осло и научного центра в Бергене. Британский английский, 1 миллион словоупотреблений, структура похожа на Брауновский корпус. Учёные уже начали понимать, однако, что одного миллиона словоупотреблений недостаточно для анализа низкочастотных элементов языка (а их большинство). Тем не менее, на Брауновском и LOB корпусах основаны многие сотни качественных и интересных исследований. Сайт проекта - <http://khnt.hit.uib.no/icame/manuals/lobman/>

3. London-Lund Corpus (LLC) В 1975 году было завершено создание корпуса устной английской речи. Он содержал около 500 тысяч словоупотреблений с орфографической транскрипцией, фонетической и просодической разметкой. Эта грандиозная работа сначала была выполнена в

бумажном варианте сотрудниками UniversityCollegeLondon, а затем переведена в компьютерную форму лингвистами из шведского города Лунд. Сайт проекта - <http://www.ucl.ac.uk/english-usage>

4. Машинный Фонд русского языка Создание первого советского лингвистического корпуса началось в 1985 году в Институте русского языка Академии Наук СССР по инициативе академика А. П. Ершова. Успели только разработать концепцию и архитектуру корпуса и несколько программ, а также собрать какое-то количество текстов. В районе 5 1991 года финансирование прекратилось и работы заглохли. В 1985-90 в Машинном фонде русского языка накоплены на машинных носителях и частично в базах данных текстовые источники рус.лит-ры 19-20 вв., главнейшие словари рус. языка, «Краткая русская грамматика» (1989), нек-рые другие материалы справочного характера, созданы текстовые корпуса поэзии, художественной прозы, общественно-политических и технических текстов; разработан программно-источниковый пакет UNILEX для персональных компьютеров. Опыт МФРЯ используется при разработке, начиная с 2004 года, Национального корпуса русского языка. 5. Уппсальский корпус русского языка Одновременно (в 1980-е годы) в институте славистики университета Уппсалы (Швеция) был создан Уппсальский корпус современных русских текстов. 1 миллион словоупотреблений, около 600 текстов. Сайт проекта - <http://www.slaviska.uu.se/korpus.htm>

Второе поколение корпусов К 90-м годам технологии извлечения и хранения текстов позволили создавать корпуса из ста миллионов словоупотреблений и более. 1. TheCobuildProject / TheBankofEnglish Началось всё в 1980 году, когда издательство Collins принялось за составление корпуса для создания нового словаря. В 1990 году было объявлено об объединении усилий Collins и факультета английского языка университета Бирмингема в инициативу под названием TheBankofEnglish. TheBankofEnglish — это так называемый мониторинговый корпус. Вот слова руководителя проекта Джона Синклера: мониторинговый корпус это огромный, вечно изменяющийся поток языка, не имеющий чёткоопределённого размера. Этот поток проходит через фильтры, которые извлекают из него лингвистические данные. Около 300 миллионов словоупотреблений в 1997 году, а в 2005 уже 525 миллионов. Каждый месяц в корпус поступает два с половиной миллиона новых словоупотреблений. 25 процентов корпуса составляет устная речь, 75 процентов — письменная. По адресу <http://www.collins.co.uk/Corpus/CorpusSearch.aspx>

можно 6 использовать тестовую версию корпуса (56 миллионов словоупотреблений). 2. BritishNationalCorpus (BNC) 100 миллионов

словоупотреблений, представляет английский язык в целом, а не один жанр. Этот корпус имеет конечный размер, в отличие от CobuildProject. 90 процентов письменных текстов, 10 устных. В создании принимали участие многие организации, включая Британское правительство. Процесс завершился около 1995 года. Корпус состоит из 4124 текстов, из которых 863 транскрибированы из устных бесед или монологов. Каждый текст сегментирован на орфографические предложения, а внутри них каждому слову автоматически назначен код класса слова (части речи). Во всём корпусе 6,4 миллионов орфографических предложений. Сегментирование и классификация слов были выполнены программой стохастической разметки CLAWS, разработанной в университете Ланкастера. Классификационная схема предусматривает 65 частей речи, которые описаны в прилагающейся документации. Все тексты размечены с использованием наиболее стандартных способов — языка SGML и системы TEI. При создании корпуса были использованы новые подходы к отбору текстов, многоуровневая система контроля. Корпус доступен по адресу <http://www.natcorp.ox.ac.uk/>

3. TheInternationalCorpusofEnglish (ICE) Совместный проект нескольких десятков университетов. 20 параллельных подкорпусов, по миллиону словоупотреблений каждый, вместе 20 миллионов словоупотреблений. Можно изучать специфику стран, где английский – второй или официальный язык (Австралия, Канада, Новая Зеландия и т.п.). Разработано сложное программное обеспечение специально для анализа этого корпуса. Веб-сайт: <http://www.ucl.ac.uk/english-usage/ice/> 4. AmericanNationalCorpus Первый выпуск состоялся в 2003 году. Планируется 100 миллионов словоупотреблений, но пока только 11. Доступ исключительно на платной основе. Корпус в XML-формате. Вебсайт проекта: <http://www.americannationalcorpus.org/>

5. Gigawordcorpora 7 В 2000-е создаются мониторинговые корпуса английского, арабского, китайского и других языков. Спонсируются Европейским Союзом, создаёт их компания LinguisticDataConsortium. Уже 1 миллиард словоупотреблений. В основном тексты взяты из публицистики и новостей. Корпусы довольно дорогие. Посмотреть на список можно на сайте <http://www.ldc.upenn.edu>.

История лингвистических корпусов.

Корпус	Период	Язык	Объем	Доступ	Дополнительная информация

ЛАБОРАТОРНАЯ РАБОТА 4. 2 час. Средства разметки корпуса.

Языковые средства представления размеченных текстов. Международные стандарты и проекты (TEI, EAGLES, CDIF, XCES). Проблемы, возникающие при разметке текстов. Многоязычные корпуса для перевода. Параллельные корпуса. Конкорданс. Многоязычие. Проблема выравнивания. Примеры параллельных корпусов.

В ходе лабораторной работы студенты отвечают подготовленный теоретический материал и выполняют индивидуальные практические задания.

Задание 4. На сайте Британского национального корпуса TheBritishNationalCorpus (BNC) (<http://www.natcorp.ox.ac.uk/>)

в строке поиска введите «see». Выберите 10 примеров и переведите их на русский язык. Отметьте, какие варианты перевода данной лексики вы выбрали, насколько они различны. Сравните результаты с другими студентами.

ЛАБОРАТОРНАЯ РАБОТА 5.2 час. Типология корпусов.

Создание и типология корпусов. Типы корпусов: устные и письменные, одноязычные и многоязычные, аннотированные и неаннотированные. Лингвистические исследования на базе корпусов. Использование корпусов в обучении языкам и гуманитарных исследованиях. Лингвистические исследования на базе корпуса: изучение лексики. Лингвистические исследования на базе корпуса: НКРЯ и BNC: изучение грамматики, исследование семантики (метафора, этимология).

В ходе лабораторной работы студенты отвечают подготовленный теоретический материал и выполняют индивидуальные практические задания.

Задание 5. Выберите концепт и определите его концептосферу. На сайтах Британского национального корпуса TheBritishNationalCorpus (BYU-BNC) (BNC) (<http://corpus.byu.edu/bnc/>)

и Корпуса современного американского английского CorpusofContemporaryAmericanEnglish (COCA) (<http://corpus.byu.edu/coca/>) сравните объем их 10 представлений. Выпишите для каждого элемента концептосферы пример употребления.

ЛАБОРАТОРНАЯ РАБОТА 6. 2 час. Обзор существующих корпусов различных типов.

Национальные корпусы. Национальный корпус русского языка. Корпусы английского языка. Практическое использование национальных корпусов. Проблемы и перспективы развития национальных корпусов. Автоматизация процессов обработки естественного языка с помощью корпусов. Использование корпусов в настройке и самообучении лингвистических автоматов. Методы извлечения информации из корпуса. Типы извлекаемой информации.

В ходе лабораторной работы студенты отвечают подготовленный теоретический материал и выполняют индивидуальные практические задания. Работа с лекционным материалом. Подготовка рефератов. Работа с электронными базами данных.

Задание 6. На сайте Корпуса русского литературного языка (<http://www.narusco.ru/>)

сравните частотность следующих лексем: Россия, родина, земля, страна, отечество.

На сайте Корпуса русского литературного языка (<http://www.narusco.ru/>) используйте функцию «создание конкорданса по текстам корпуса» и сравните объем представления всех словоформ определенного существительного.

На сайте Национального корпуса русского языка (<http://www.ruscorpora.ru/>)

в основном корпусе найдите примеры употребления 10 лексем. Составьте с ними упражнение с выбором правильного ответа для нахождения предложения с тем же смыслом.

На сайте Национального корпуса русского языка (<http://www.ruscorpora.ru/>)

на странице основного 11 корпуса выберите функцию «задать подкорпус», сравните объем выборки документов, предложений и слов 10 различных авторов.

ЛАБОРАТОРНАЯ РАБОТА 7. 2 час. Типология корпусов.

Создание и типология корпусов. Типы корпусов: устные и письменные, одноязычные и многоязычные, аннотированные и неаннотированные. Корпус как поисковая система. Языки запросов корпусных менеджеров. Выходные интерфейсы Сравнительный анализ корпусов и корпусных менеджеров. Поиск в корпусах: конкордансер и корпусный менеджер как поисковая система. Корпусные менеджеры. Сравнение корпусных менеджеров с архитектурой поисковых систем в

сети Интернет. Поисковые системы Интернет как корпусные менеджеры. Специализированные корпусные интерфейсы к индексам глобальных поисковых систем.

В ходе лабораторной работы студенты отвечают подготовленный теоретический материал и выполняют индивидуальные практические задания.

Задание 7.

Выберите 5 статей из различных газет, задайте для поиска 4 лексемы (2 глагола и 2 существительных или 2 местоимения), составьте для них конкордансы в формате Excel. Задание

Проанализируйте составленные конкордансы. Сравните объем выборки, частотность употребления, количество словоформ.

1. Изучите опыт создания зарубежных корпусов. 2. Познакомьтесь с возможностями русских корпусов.

ЛАБОРАТОРНАЯ РАБОТА 8. 2 час. Технологии разработки корпусов. Технология создания корпусов. Стадии работы Корпус как особый тип информационно-поисковой системы. Разбиение на подсистемы.

В ходе лабораторной работы студенты отвечают подготовленный теоретический материал и выполняют индивидуальные практические задания.

Задание 8. Произведите лемматизацию и стемминг следующих предложений: 1) У меня такая плохая память на лица. 2) Он не спал всю ночь, однако об отдыхе не помышлял. 3) Ответственных за связь со средствами массовой информации прошу немедленно вступить в переговоры. 4) He noticed that one of the frames held the one sheet for a film called The Art of the Cape. 5) This time he awoke to someone yelling, beating him, throwing him out of the freight car. 6) They stripped him of his powers; they stripped him of his lands; they plucked the weapons from the hands of his clansmen, that had borne arms for thirty centuries.

Произведите синтаксический анализ предложений, представив их в виде дерева зависимостей. 1) Она расплакалась и сунула бумаги в карман передника. 2) На него смотрело искаженное страданиями, но довольно юное лицо актера без ангажемента. 3) One morning in mid-December, Hogwarts woke to find itself covered in several feet of snow. 4) That same night he left for Anatolia and remembered, later on that trip, riding all day through fields of the poppies her beautiful face.

В программе WordCruncher на примере текстов Шекспира сравнить частотность и объем выборки следующих лексем: 15 А) “you”, “thy”, “thou”; Б) “man” и “woman”.

1. Охарактеризуйте лексикографические исследования на базе корпусов.
2. Рассмотрите грамматические исследования на базе корпусов.
3. Изучите исследования дискурса, основанные на корпусах.

5. СТРУКТУРА, СОДЕРЖАНИЕ, УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ САМОСТОЯТЕЛЬНОЙ РАБОТЫ

Рекомендации по самостоятельной работе студентов

Самостоятельная работа студентов предполагает изучение тем. На основе изученного материала студентам предлагается ответить на вопросы, представленные после тем.

Методические рекомендации к самостоятельной работе: Систематизировать литературные источники, проанализировать и обобщить представленные в них концепции. Из всего материала выделить существующие точки зрения на проблему, проанализировать их, сравнить, дать им оценку. Итогом этой работы должна стать логически выстроенная система сведений по существу исследуемого вопроса.

Выполнить письменный анализ проблемных вопросов. Обосновать и применить к материалу одну из нескольких предложенных концепций – привести аргументы в пользу правильности выбранного варианта ответа и указать, в чем ошибочность других вариантов. Аргументировать выбор адекватного способа действий, ведущего к решению задачи.

Темы:

1. Лексикографические исследования, основанные на корпусах
2. Пример одного лексикографического исследования
3. Выделение коллокаций статистическими методами
4. Лингвистические исследования на базе корпуса: изучение грамматики.
5. Распределение и функции номинализаций.
6. Распределение грамматических категорий. Семантическое наполнение Британского национального корпуса.
7. Способы использования корпусов.
8. Семантическое наполнение Британского национального корпуса.
9. Методы извлечения информации из корпуса. Типы извлекаемой информации.
10. Программы для работы с корпусом: конкордансеры и другие.

11. Статистические меры вычисления совместной встречаемости

12. Использование корпусов в социологии, истории.

Вопросы для устного контроля самостоятельной работы студентов

Вариант 1.

1. Расскажите о становлении корпусной лингвистики как раздела компьютерной лингвистики, ее целях, задачах, предмете и объекте исследования.

2. Покажите двойственный характер корпусной лингвистики. 3. Какие лингвистические направления предвосхитили появление корпусной лингвистики.

3. Сопоставьте корпусную и традиционную лингвистику. В чем специфика корпусного подхода к языку

Вариант 2.

1. Каковы предпосылки создания и возможности использования корпусов в лингвистических исследованиях, лексикографии, переводческой деятельности и в обучении иностранным языкам. 4. Как развивались основные характеристики корпусов.

2. Классификации корпусов по разным основаниям.

3. Покажите основные процедуры обработки естественного языка в корпусе.

4. Какие программы производят процессы токенизации, лемматизации, стемминга и парсинга: специфика, цели, результаты.

5. Сопоставьте корпусные менеджеры.

Критерии оценки выполнения самостоятельной работы.

• Самостоятельная работа оценивается на «отлично», если студент демонстрирует глубокие знания теории в области корпусной лингвистики, излагает различные точки зрения и концепции с указанием авторов, владеет соответствующей научной терминологией, полно отвечает на дополнительные вопросы.

• Ответ оценивается на «хорошо», если студент, в целом владея теоретическим материалом, допускает отдельные неточности в интерпретации данных корпусов.

• Студент получает «удовлетворительно», если допускает значительные неточности и лакуны в изложении теоретического материала

• Оценка «неудовлетворительно» выставляется, если теоретические вопросы остаются нераскрытыми.

План-график выполнения самостоятельной работы по дисциплине

№	Вид самостоятельной	Дата/сроки	Примерные	Форма контроля
---	---------------------	------------	-----------	----------------

п/п	работы	выполнения	нормы времени на выполнение	
	Задания для самостоятельной работы к теме 1. Введение в корпусную лингвистику Корпусные лингвистические исследования	1-8 неделя	3 ч.	УО-1
	Задания для самостоятельной работы к теме 2. Корпусные лингвистические исследования	9-16 неделя	4 ч.	УО-1
	Задания для самостоятельной работы к лабораторным занятиям 1-8.	1-8 неделя	3 ч.	УО-1
	Задания для самостоятельной работы лабораторным занятиям 9-16.	9-16 неделя	3 ч.	УО-1

Задания для самостоятельной работы к теме 1.

- 1) Количественные методы в языкознании.
- 2) Понятие лингвистического корпуса.
- 3) Основные свойства лингвистических корпусов.
- 4) Репрезентативность и сбалансированность лингвистического корпуса.
- 5) Разметка в лингвистическом корпусе.
- 6) Виды корпусной разметки, метатекстовая и лингвистическая разметка.
- 7) Виды лингвистической разметки.
- 8) Корпус как информационно-поисковая система.
- 9) Критерии классификации корпусов.
- 10) Различные виды корпусов.
- 11) История появления корпусов языков мира.
- 12) Современные корпусы языков мира.
- 13) Национальный корпус русского языка.
- 14) Корпусы тюркских языков.
- 15) Корпус английского языка, его структурные, технологические и функциональные особенности.
- 16) Научные исследования на основе корпуса английского языка

Задания для самостоятельной работы к теме 2.

1. Основные понятия и определения корпусной лингвистики.
2. Лингвистические (языковые) и нелингвистические корпусы.
3. История лингвистических корпусов: от картотеки к корпусу.
4. Корпусная лингвистика: современное состояние.
5. Корпусная лингвистика в России.

6. Классификация (типология) корпусов по различным основаниям.
7. Типы корпусов по задачам. Типы корпусов по формальным признакам
8. Британский национальный корпус.
9. Чешский национальный корпус.
10. Польский национальный корпус.
11. Национальный корпус русского языка.
12. Мангеймский корпус немецкого языка.
13. Русско-английский корпус С. Шарова.
14. Корпус языка А.С. Грибоедова.
15. Корпус русского языка 19 века.
16. Зарубежные национальные корпуса.
17. Корпусы русского языка.
18. Специальные корпуса. Корпусные менеджеры.
19. Корпус как поисковая система. Языки запросов. Выходные интерфейсы.
20. Сравнительный анализ. Корпусные исследования.
21. Лексические исследования, базирующиеся на корпусах.
22. Грамматические исследования, базирующиеся на корпусах.
23. Семантические исследования, базирующиеся на корпусах.
24. Использование корпусов в социологии, исторической науке и др.
25. Предварительные работы по созданию корпуса: проблемы репрезентативности, отбор источников, внешние и внутренние критерии отбора.
26. Нормализация файлов. Графематический анализ. Автоматический морфологический и синтаксический анализ.
27. Средства создания и разметки корпусов: понятие разметки, типы разметки. Металингвистическая разметка.
28. Параллельные корпуса. Проблема выравнивания. Стандартизация в корпусной лингвистике.
29. Языковые средства представления размеченных текстов

Планирование и организация времени, отведённого на выполнение заданий самостоятельной работы: изучив график выполнения самостоятельной работы, следует правильно её организовать. Рекомендуется изучить структуру каждого задания, обратить внимание на график выполнения работ, отчётность по каждому заданию предоставляется в последнюю неделю согласно графику. Обратите внимание, что итоги самостоятельной работы влияют на окончательную оценку по итогам освоения учебной дисциплины.

Работа с литературой.

При выполнении ряда заданий требуется работать с литературой.

Рекомендуется использовать различные возможности работы с литературой: фонды научной библиотеки ДВФУ (<http://www.dvfu.ru/library/>) и других ведущих вузов страны, а также доступных для использования научно-библиотечных систем.

В процессе выполнения самостоятельной работы рекомендуется работать со следующими видами изданий:

а) научные издания, предназначенные для научной работы и содержащие теоретические, экспериментальные сведения об исследованиях. Они могут публиковаться в форме монографий, научных статей в журналах или в научных сборниках;

б) учебная литература подразделяется на:

- учебные издания (учебники, учебные пособия, тексты лекций), в которых содержится наиболее полное системное изложение дисциплины или какого-то ее раздела;

- справочники, словари и энциклопедии – издания, содержащие краткие сведения научного или прикладного характера, не предназначенные для сплошного чтения. Их цель – возможность быстрого получения самых общих представлений о предмете.

Существуют два метода работы над источниками:

– сплошное чтение обязательно при изучении учебника, глав монографии или статьи, то есть того, что имеет учебное значение. Как правило, здесь требуется повторное чтение, для того чтобы понять написанное. Старайтесь при сплошном чтении не пропускать комментарии, сноски, справочные материалы, так как они предназначены для пояснений и помощи. Анализируйте рисунки (карты, диаграммы, графики), старайтесь понять, какие тенденции и закономерности они отражают;

– метод выборочного чтения дополняет сплошное чтение; он применяется для поисков дополнительных, уточняющих необходимых сведений в словарях, энциклопедиях, иных справочных изданиях. Этот метод крайне важен для повторения изученного и его закрепления, особенно при подготовке к зачету.

Для того чтобы каждый метод принёс наибольший эффект, необходимо фиксировать все важные моменты, связанные с интересующей Вас темой.

Тезисы – это основные положения научного труда, статьи или другого произведения, а возможно, и устного выступления; они несут в себе большой объем информации, нежели план. Простые тезисы лаконичны по форме; сложные – помимо главной авторской мысли содержат краткое ее обоснование и доказательства, придающие тезисам более весомый и убедительный характер. Тезисы прочитанного позволяют глубже раскрыть

его содержание; обучаясь излагать суть прочитанного в тезисной форме, вы сумеете выделять из множества мыслей авторов самые главные и ценные и делать обобщения.

Конспект – это способ самостоятельно изложить содержание книги или статьи в логической последовательности. Конспектируя какой-либо источник, надо стремиться к тому, чтобы немногими словами сказать о многом. В тексте конспекта желательно поместить не только выводы или положения, но и их аргументированные доказательства (факты, цифры, цитаты).

Писать конспект можно и по мере изучения произведения, например, если прорабатывается монография или несколько журнальных статей.

Составляя тезисы или конспект, всегда делайте ссылки на страницы, с которых вы взяли конспектируемое положение или факт, – это поможет вам сократить время на поиск нужного места в книге, если возникает потребность глубже разобраться с излагаемым вопросом или что-то уточнить при написании письменных работ.

Методические рекомендации по выполнению заданий для самостоятельной работы и критерии оценки.

УО-1. От обучающегося требуется:

– отчёт по теме осуществляется в форме собеседования/устного опроса (УО-1);

– свободно ориентироваться в теории речевой коммуникации и уметь применять её положения при анализе речевых ситуаций.

Собеседование (устный опрос) позволяет оценить знания и кругозор студента, умение логически построить ответ, владение монологической речью и иные коммуникативные навыки.

Опрос – важнейшее средство развития мышления и речи. Обучающая функция опроса состоит в выявлении деталей, которые по каким-то причинам оказались недостаточно осмысленными в ходе учебных занятий и при подготовке задания по самостоятельной работе.

Критерии оценки. Используется зачетная система. Во время собеседования/опроса допускается не более 3-х ошибок.

6. КОНТРОЛЬ ДОСТИЖЕНИЯ ЦЕЛЕЙ КУРСА

№ п/п	Контролируемые разделы / темы дисциплины	Код и наименование индикатора достижения компетенции	Результаты обучения	Оценочные средства	
				текущий контроль	промежуточная аттестация
1.	Раздел 1 №1, Введение в корпусную лингвистику	ПК -9.1 Знает основные принципы обработки информации; базовые принципы корпусной лингвистики, лексикографии, математической статистики; базовые представления о языковом разнообразии; наиболее полные и значимые лингвистические корпуса, электронные словари и базы данных.	Знает основные корпуса и положения корпусной лингвистики.	УО-1	Вопросы № 1,2 экзаменационного билета, билеты 1-12
			Умеет пользоваться корпусами и корпусными менеджерами.		
			Владеет навыками составления корпусов.		
2.	Раздел 1 №1, Введение в корпусную лингвистику	ПК -9.2 Умеет пользоваться основными методами, способами и средствами получения, хранения, переработки информации; пользоваться лингвистически ориентированными программными продуктами	Знает основные принципы устройства корпуса.	УО-1	Вопросы № 1,2 экзаменационного билета, билеты 1-12
			Умеет искать информацию в корпусе.		
			Владеет навыками алгоритмизации составления корпуса.		
3.	Раздел 2 №2, Корпусные лингвистические исследования	ПК -9.3 Имеет практический опыт разработки электронных языковых ресурсов; опыт применения основных методов, способов и средств получения, хранения,	Знает основные принципы разработки электронных языковых ресурсов.	УО-1	Вопросы № 1,2 экзаменационного билета, билеты 1-12
			Умеет применять основные методы, способы и средства получения, хранения, переработки информации в лингвистическом корпусе.		
			Владеет навыками переработки языковых		

		переработки информации	данных для составления лингвистического корпуса.		
--	--	------------------------	--	--	--

Типовые контрольные задания, методические материалы, определяющие процедуры оценивания знаний, умений и навыков и (или) опыта деятельности, а также критерии и показатели, необходимые для оценки знаний, умений, навыков и характеризующие результаты обучения, представлены в Приложении.

7. СПИСОК УЧЕБНОЙ ЛИТЕРАТУРЫ И ИНФОРМАЦИОННО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Основная литература

(электронные и печатные издания)

1. Информатика и лингвистика : учеб. пособие / Т.М. Волосатова, Н.В. Чичварин. — М. : ИНФРА-М, 2018. — 196 с. + Доп. материалы [Электронный ресурс; Режим доступа <http://www.znanium.com>]. — (Высшее образование: Бакалавриат). — www.dx.doi.org/10.12737/16175. - Режим доступа: <http://znanium.com/catalog/product/938009>
2. Чернявская, В. Е. Лингвистика текста. Лингвистика дискурса : учеб. пособие / В. Е. Чернявская. — М. : Флинта : Наука, 2019. — 208 с. - ISBN 978-5-9765-1454-6 (Флинта), ISBN 978-5-02-037800-1 (Наука) - Режим доступа: <http://znanium.com/catalog/product/457232>
3. Информатика и лингвистика: Учебное пособие / Волосатова Т.М., Чичварин Н.В. - М.: НИЦ ИНФРА-М, 2019. - 196 с.: 60x90 1/16. - (Высшее образование: Бакалавриат) (Переплёт 7БЦ) ISBN 978-5-16-010977-0 - Режим доступа: <http://znanium.com/catalog/product/508097>
4. Щипицина, Л. Ю. Информационные технологии в лингвистике [Электронный ресурс] : учеб. пособие / Л. Ю. Щипицина. — М. : ФЛИНТА, 2019. — 128 с. - ISBN 978-5-9765-1431-7 - Режим доступа: <http://znanium.com/catalog/product/462989>
4. Лексикология английского языка: Учебное пособие / Кругликова Е.А. - Краснояр.: СФУ, 2019. - 162 с.: ISBN 978-5-7638-3479-6 - Режим доступа: <http://znanium.com/catalog/product/978599>

Дополнительная литература

(печатные и электронные издания)

1. Баранов, А. Н. Лингвистическая экспертиза текста : теория и практика [Электронный ресурс] : учеб. пособие / А. Н. Баранов. - 4-е изд., стереотип. - М. : Флинта : Наука, 2018. - ISBN 978-5-9765-0083-9 (Флинта), ISBN 978-5-02-034561-4 (Наука) - Режим доступа: <http://znanium.com/catalog/product/454832>
2. Моисеева И.Ю. Квантитативная лингвистика и новые информационные технологии [Электронный ресурс]: учебное пособие/ Моисеева И.Ю.— Электрон.текстовые данные.— Оренбург: Оренбургский государственный университет, ЭБС АСВ, 2018.— 103 с.— Режим доступа: <http://www.iprbookshop.ru/71281.html>.

3. Беляева Л.Н. Лингвистические технологии в современном сетевом пространстве. *Languageworker* в индустрии локализации [Электронный ресурс]: монография/ Беляева Л.Н.— Электрон.текстовые данные.— СПб.: Книжный дом, 2018.— 133 с.— Режим доступа: <http://www.iprbookshop.ru/71523.html>.
4. Кайсарова Д.В. Математическая лингвистика. Практикум [Электронный ресурс]/ Кайсарова Д.В., Коцюба И.Ю.— Электрон.текстовые данные.— СПб.: Университет ИТМО, 2018.— 67 с.— Режим доступа: <http://www.iprbookshop.ru/67815.html>.

Дополнительная литература
(печатные и электронные издания)

1. Лайонз, Дж. Введение в теоретическую лингвистику / Дж. Лайонз ; пер. с англ. под ред. и с предисл. В. А. Звегинцева. – М. : Прогресс, 2017. – 543 с. <http://lib.dvfu.ru:8080/lib/item?id=chamo:63721&theme=FEFU>
2. Мечковская, Н. Б. Общее языкознание: структурная и социальная типология языков: учебн. пособие / Н. Б. Мечковская. – М. : Флинта : Наука , 2017. – 312 с. <http://lib.dvfu.ru:8080/lib/item?id=chamo:249150&theme=FEFU>
3. Соссюр, Ф. Курс общей лингвистики / Ф. де Соссюр. – Екатеринбург : Изд-во Уральского университета, 2017. – 426 с. <http://lib.dvfu.ru:8080/lib/item?id=chamo:366054&theme=FEFU>
4. Степанов, Ю. С. Основы языкознания : учебное пособие для аспирантов и студентов филологических факультетов педагогических институтов / Ю. С. Степанов. – М. : Просвещение, 2018. – 271 с. <http://lib.dvfu.ru:8080/lib/item?id=chamo:59469&theme=FEFU>

**Перечень ресурсов информационно-телекоммуникационной сети
«Интернет»**

- Студенты могут получить доступ к электронным образовательным ресурсам через сайт ДВФУ *научно-информационные электронные ресурсы* (доступ с сайта Научной библиотеки ДВФУ) URL: http://www.dvfu.ru/web/library/rus_res
1. Научная электронная библиотека eLIBRARY.RU URL: <http://www.elibrary.ru>
 2. Базы данных компании East View Publications URL: <http://www.ebiblioteka.ru/>
 3. Базы данных компании "Интегрум"
 4. Электронная библиотека диссертаций РГБ URL: <http://diss.rsl.ru>

5. Единое окно доступа к образовательным ресурсам URL: <http://window.edu.ru/>
6. Электронно-библиотечная система Издательства "Лань" URL: <http://e.lanbook.com>
7. Электронно-библиотечная система znanium.com НИЦ "ИНФРА-М" URL: <http://znanium.com>
8. Информационная система - ЕДИНОЕ ОКНО доступа к образовательным ресурсам <http://window.edu.ru/>

а так же в свободном доступе в Интернет:

1. Gaudeamus [Электронная библиотека учебной PDF-литературы и учебников для вузов. (бесплатные полнотекстовые учебники)] URL: <http://www.gaudeamus.omskcity.com/>

Перечень информационных технологий и программного обеспечения:

1. Microsoft Word
2. Microsoft PowerPoint
3. Microsoft Internet Explorer/ Mozilla Firefox/ Opera
4. SketchEngine, AntConc (в свободном доступе)

8.МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ

Методические рекомендации к освоению дисциплины

Описание последовательности действий обучающихся, алгоритм изучения дисциплины:

- Систематизировать литературные источники,
- проанализировать и обобщить представленные в них концепции
- Из всего материала выделить существующие точки зрения на проблему,
- Проанализировать их, сравнить, дать им оценку.
- Итогом этой работы должна стать логически выстроенная система сведений по существу исследуемого вопроса.

9. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Учебные занятия по дисциплине могут проводиться в следующих помещениях, оснащенных соответствующим оборудованием и программным обеспечением, расположенных по адресу Приморский край, г. Владивосток, Фрунзенский р-н г., Русский Остров, ул. Аякс, п, д. 10:

Перечень материально-технического и программного обеспечения дисциплины приведен в таблице.

№ помещения по плану БТИ	Наименование специальных помещений и помещений для самостоятельной работы ¹	Оснащенность специальных помещений и помещений для самостоятельной работы	Перечень лицензионного программного обеспечения. Реквизиты подтверждающего документа
Учебные аудитории для проведения учебных занятий:			
	D208/347, D303, D313а, D401, D453, D461, D518, D708, D709, D758, D761, D762, D765, D766, D771, D917, D918, D920, D925, D576, D807	Лекционная аудитория оборудована маркерной доской, аудиопроигрывателем	ЗДЕСЬ ДОПОЛНИТСЯ ЛИЦЕНЗИОННЫМ ПО
	D229, D304, D306, D349, D350, D351, D352, D353, D403, D404, D405, D414, D434, D435, D453, D503, D504, D517, D522, D577, D578, D579, D580, D602, D603, D657, D658, D702, D704, D705, D707, D721, D722, D723, D735, D736, D764, D769, D770, D773, D810, D811, D906, D914, D921, D922, D923, D924, D926	Мультимедийная аудитория: Проектор Mitsubishi EW330U, Экран проекционный ScreenLineTrimWhiteIce, профессиональная ЖК-панель 47", 500 Кд/м2, Full HD M4716CCBA LG, подсистема видеоисточников документ-камера CP355AF Avervision; подсистема видеокоммутации; подсистема аудиокоммутации и звукоусиления; подсистема интерактивного управления	
	D207/346	Мультимедийная аудитория: Проектор 3-chip DLP, 10 600 ANSI-лм, WUXGA 1 920x1 200 (16:10) PT-DZ110XE Panasonic; экран 316x500 см, 16:10 с эл. приводом; крепление настенно-потолочное ElproLargeElectrolProjecta; профессиональная ЖК-панель 47", 500 Кд/м2, Full HD M4716CCBA LG; подсистема видеоисточников документ-камера CP355AF Avervision; подсистема видеокоммутации; подсистема аудиокоммутации и звукоусиления; подсистема интерактивного управления),	
	D226	Мультимедийная аудитория: Проектор Mitsubishi EW330U, Экран проекционный ScreenLineTrimWhiteIce, профессиональная ЖК-панель 47", 500 Кд/м2, Full HD M4716CCBA LG, подсистема видеокоммутации; подсистема	

¹ В соответствии с п.4.3.1 ФГОС

		аудиокоммутации и звукоусиления; подсистема интерактивного управления), D362 (профессиональная ЖК-панель 47", 500 Кд/м2, Full HD M4716CCBA LG, подсистема аудиокоммутации и звукоусиления; Компьютерный класс на 15 посадочных мест	
	D447, D448, D449, D450, D451, D452, D502, D575	Мультимедийная аудитория: Проектор Mitsubishi EW330U, Экран проекционный ScreenLineTrimWhiteIce, подсистема видеоисточников документ-камера CP355AF Avervision; подсистема видеокоммутации; подсистема аудиокоммутации и звукоусиления; подсистема интерактивного управления	
	D446, D604, D656, D659, D737, D808, D809, D812	Мультимедийная аудитория: Проектор Mitsubishi EW330U, Экран проекционный ScreenLineTrimWhiteIce, профессиональная ЖК-панель 47", 500 Кд/м2, Full HD M4716CCBA LG, подсистема видеоисточников документ-камера CP355AF Avervision; подсистема видеокоммутации; подсистема аудиокоммутации и звукоусиления; подсистема интерактивного управления; Компьютерный класс; Рабочее место: Компьютеры (Твердотельный диск - объемом 128 ГБ; Жесткий диск - объем 1000 ГБ; Форм-фактор – Tower); комплектуется клавиатурой, мышью. Монитором АОС i2757Fm; комплектом шнуров эл. питания) Модель - M93p 1; Лингафонный класс, компьютеры оснащены программным комплексом Sanakostudy 1200	
	D501, D601	Мультимедийная аудитория: Проектор Mitsubishi EW330U, Экран проекционный ScreenLineTrimWhiteIce, профессиональная ЖК-панель 47", 500 Кд/м2, Full HD M4716CCBA LG, подсистема видеоисточников документ-камера CP355AF Avervision; подсистема видеокоммутации; подсистема аудиокоммутации и звукоусиления; подсистема интерактивного управления; Компьютерный класс на 26 рабочих мест. Рабочее место: Моноблок Lenovo C360G-i34164G500UDK	

Помещения для самостоятельной работы:			
	<p>А1042 аудитория для самостоятельной работы студентов</p>	<p>Моноблок Lenovo C360G-i34164G500UDK – 115 шт.; Интегрированный сенсорный дисплей PolymediaFlipBox; Копир-принтер-цветной сканер в e-mail с 4 лотками XeroxWorkCentre 5330 (WC5330C; Полноцветный копир-принтер-сканер XeroxWorkCentre 7530 (WC7530CPS Оборудование для инвалидов и лиц с ограниченными возможностями здоровья: Дисплей Брайля Focus-40 Blue – 3 шт.; Дисплей Брайля Focus-80 Blue; Рабочая станция LenovoThinkCentre E73z – 3 шт.; Видео увеличитель ONYX Swing-Arm PC edition; Маркер-диктофон TouchMemo цифровой; Устройство портативное для чтения плоскочечатных текстов PEarl; Сканирующая и читающая машина для незрячих и слабовидящих пользователей SARA; Принтер Брайля EmprintSpotDot - 2 шт.; Принтер Брайля Everest - D V4; Видео увеличитель ONYX Swing-Arm PC edition; Видео увеличитель Topaz 24" XL стационарный электронный; Обучающая система для детей тактильно-речевая, либо для людей с ограниченными возможностями здоровья; Увеличитель ручной видео RUBY портативный – 2 шт.; Экран Samsung S23C200B; Маркер-диктофон TouchMemo цифровой.</p>	<p>MicrosoftWindows 7 ProMAGic 12.0 Pro, JawsforWindows 15.0 Pro, Openbook 9.0, DuxburyBrailleTranslator, DolphinGuide (контракт № А238-14/2); Неисключительные права на использование ПО Microsoft рабочих станций пользователей (контракт ЭА-261-18 от 02.08.2018): - лицензия на клиентскую операционную систему; - лицензия на пакет офисных продуктов для работы с документами включая формат.docx , .xlsx , .vsd , .ppt.; - лицензия на право подключения пользователя к серверным операционным системам , используемым в ДВФУ : MicrosoftWindowsServer 2008/2012; - лицензия на право подключения к серверу MicrosoftExchangeServerEnterprise; - лицензия на право подключения к внутренней информационной системе документооборота и порталу с возможностью поиска информации во множестве удаленных и локальных хранилищах, ресурсах, библиотеках информации, включая порталные хранилища, используемой в ДВФУ: MicrosoftSharePoint; - лицензия на право подключения к системе централизованного управления рабочими станциями, используемой в ДВФУ: MicrosoftSystemCenter.</p>

Х. ФОНДЫ ОЦЕНОЧНЫХ СРЕДСТВ

Фонды оценочных средств представлены в Приложении.



МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное автономное образовательное учреждение высшего образования
«Дальневосточный федеральный университет»
(ДВФУ)

**ВОСТОЧНЫЙ ИНСТИТУТ – ШКОЛА РЕГИОНАЛЬНЫХ И МЕЖДУНАРОДНЫХ
ИССЛЕДОВАНИЙ**

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ
по дисциплине «Технологии корпусной лингвистики»
Направление подготовки 45.03.03 Фундаментальная и прикладная
лингвистика
профиль «Лингвистика в информационно-коммуникационной цифровой
среде»
Форма подготовки очная

Владивосток
2021

**Перечень форм оценивания, применяемых на различных этапах формирования компетенций в ходе освоения
дисциплины**

№ п/п	Контролируемые разделы / темы дисциплины	Код и наименование индикатора компетенции	Результаты обучения	Оценочные средства	
				текущий контроль	промежуточная аттестация
1	Раздел 1 №1, Введение в корпусную лингвистику	ПК -9.1 Знает основные принципы обработки информации; базовые принципы корпусной лингвистики, лексикографии, математической статистики; базовые представления о языковом разнообразии; наиболее полные и значимые лингвистические корпуса, электронные словари и базы данных.	Знает основные корпуса и положения корпусной лингвистики.	УО-1	Вопросы № 1,2 экзаменационного билета, билеты 1- 12
			Умеет пользоваться корпусами и корпусными менеджерами.		
			Владеет навыками составления корпусов.		
1	Раздел 1 №1, Введение в корпусную лингвистику	ПК -9.2 Умеет пользоваться основными методами, способами и средствами получения, хранения, переработки информации; пользоваться лингвистически ориентированными программными продуктами	Знает основные принципы устройства корпуса.	УО-1	Вопросы № 1,2 экзаменационного билета, билеты 1- 12
			Умеет искать информацию в корпусе.		
			Владеет навыками алгоритмизации составления корпуса.		

2	Раздел 2 №2, Корпусные лингвистические исследования	ПК -9.3Имеет практический опыт разработки электронных языковых ресурсов; опыт применения основных методов, способов и средств получения, хранения, переработки информации	Знает основные принципы разработки электронных языковых ресурсов.	УО-1	Вопросы № 1,2 экзаменационного билета, билеты 1-12
			Умеет применять основные методы, способы и средства получения, хранения, переработки информации в лингвистическом корпусе.		
			Владеет навыками переработки языковых данных для составления лингвистического корпуса.		

Оценочные средства для текущего контроля

Текущая аттестация студентов по дисциплине проводится в соответствии с локальными нормативными актами ДВФУ и является обязательной. Текущая аттестация по дисциплине проводится в форме устного опроса и ведущим преподавателем.

Объектами оценивания выступают:

- активность на занятиях, своевременность выполнения различных видов заданий, посещаемость всех видов занятий по аттестуемой дисциплине;
- степень усвоения теоретических знаний;
- уровень овладения практическими умениями и навыками по всем видам учебной работы;
- результаты самостоятельной работы.

Оценочные средства для текущей аттестации

Вопросы для устного опроса (УО-1):

Тема 1.

1. Лексикографические исследования, основанные на корпусах
2. Пример одного лексикографического исследования
3. Выделение коллокаций статистическими методами
4. Лингвистические исследования на базе корпуса: изучение грамматики.
5. Распределение и функции номинализаций.
6. Распределение грамматических категорий. Семантическое наполнение Британского национального корпуса.
7. Способы использования корпусов.
8. Семантическое наполнение Британского национального корпуса.
9. Методы извлечения информации из корпуса. Типы извлекаемой информации.
10. Программы для работы с корпусом: конкордансеры и другие.
11. Статистические меры вычисления совместной встречаемости
12. Использование корпусов в социологии, истории.

Тема 2.

1. Чем занимается корпусная лингвистика? В чем заключается двойственность объекта ее исследования?
2. Какие направления послужили основой для развития корпусной лингвистики?
3. В чем состоят принципиальные различия между традиционной и корпусной лингвистикой?
4. Что такое языковой (лингвистический) корпус?
5. В чем состоят цели создания и использования языковых корпусов?

6. Что понимается под репрезентативностью корпуса?
7. Охарактеризуйте существующие классификации корпусов по разным критериям.
8. Что такое «токен» и в чем состоит процесс токенизации?
9. Что такое «лемма» и в чем состоит процесс лемматизации?
10. Чем отличается процесс лемматизации от стемминга?
11. Что такое парсинг и для чего он проводится?
12. Что такое разметка и для чего она проводится? Назовите типы лингвистических и экстралингвистических тэгов.
13. Дайте характеристику следующим типам разметки: морфологической, синтаксической, семантической, анафорической, просодической и дискурсной.
14. В чем заключается проблема морфологической неоднозначности и как она решается?
15. Назовите функции корпусных менеджеров.
16. Что такое конкорданс и для чего он используется?
17. Опишите общую модель языка запросов.
18. Дайте краткую характеристику Национального корпуса русского языка (состав и структура, виды разметки и т.д.)
19. Дайте краткую характеристику Корпуса современного американского варианта английского языка (Corpus of Contemporary American English).
20. Обобщите основные возможности использования языковых корпусов в научно-исследовательской и практической переводческой деятельности, а также в области обучения иностранным языкам.

Дополнительные вопросы:

1. Ручное аннотирование текста
2. Сравнение результатов аннотирования текста
3. Использование цепочки обрабатываемых ресурсов
4. Создание и применение словаря
5. Применение JARE для автоматического аннотирования
6. Использование SearchableDataStore
7. Аннотирование текста с использованием JAVA
8. Извлечение информации из коллекции гомогенных текстов

Критерии оценки устного опроса:

– 100-86 баллов выставляется студенту, если студент полно и грамотно освещает проблему; проблема раскрыта в связях и с обоснованиями, с корректным использованием лингвистических терминов и понятий в контексте ответа; представлена аргументация.

– 85-76 баллов выставляется студенту, если студент грамотно освещает проблему, но аргументация недостаточна; проблема раскрыта с корректным использованием большинства лингвистических понятий и терминов, однако студент допускает некоторые неточности в интерпретации понятий, приводит недостаточное количество примеров.

– 75-61 балл выставляется студенту, если проблема освещена неполно, аргументация отсутствует или слабо связана с раскрытием проблемы; использование лингвистических терминов некорректно, количество примеров ограничено.

60-0 баллов выставляется студенту, если студентом проблема не раскрыта или обнаруживается существенное непонимание проблемы; собеседование студентом не пройдено.

Оценочные средства для промежуточной аттестации

Код и наименование индикатора компетенции	Результаты обучения	Шкала оценивания промежуточной аттестации			
		Неудовлетворительно	Удовлетворительно	Хорошо	Отлично
ПК -9.13 Знает основные принципы обработки информации; базовые принципы корпусной лингвистики, лексикографии, математической статистики; базовые представления о языковом разнообразии; наиболее полные и значимые лингвистические корпуса, электронные словари и базы данных.	Знает основные корпуса и положения корпусной лингвистики.	Знает названия основных корпусов и положения корпусной лингвистики.	Знает базовые методы сбора и документации лингвистических данных при помощи корпусных менеджеров	Знает основные понятия по методам сбора и документации лингвистических данных;	Способность перечислить и раскрыть суть методов сбора и документации лингвистических данных;
	Умеет пользоваться корпусами и корпусными менеджерами.	Умеет включить корпусный менеджер.	Умеет использовать методику сбора и документации лингвистических данных для анализа и оценивания различных фактов и явлений на основе корпусного менеджера AntConc.	Умение работать с информацией из электронных и библиотечных источников, умение применять известные методы разработки проектных идей, умение представлять результаты существующих проектных решений по изучаемой проблеме и собственных исследований на их основе, умение применять творческий подход к решению задачи сбора и документации	способность работать с информацией из электронных и библиотечных источников; способность найти проектные разработки по похожей проблематике и обосновать объективность применения их результатов; способность изучить основные этапы проектирования; способность применять методы концептуальных решений при решении проектных задач.

				лингвистических данных;	
	Владеет навыками составления корпусов.	Слабо владеет навыком составления корпуса;	Владеет навыками анализа проблем, возникающих при решении исследовательских и практических задач на основе сбора и документации лингвистических данных в SketchEngine (WebBootCat).	Владение терминологией проблемного проектирования, владение способностью сформулировать задание сбора и документации лингвистических данных;	Способность бегло и точно применять терминологический аппарат при обосновании своих предложений при разработке сбора и документации лингвистических данных;
ПК -9.2 Умеет пользоваться основными методами, способами и средствами получения, хранения, переработки информации; пользоваться лингвистически ориентированными программными продуктами	Знает основные принципы устройства корпуса.	Путает принципы устройства корпуса;	Знает методы описания языка через создание собственных учебных корпусов;	Может дать определения основных понятий и категорий современной лингвистики, знает, где найти информацию;	Способен перечислить и раскрыть суть основных понятий и категорий современной лингвистики, знает, где найти информацию;
	Умеет искать информацию в корпусе.		Способен проводить несложные научные эксперименты и экспертизу с поиском информации в корпусе;	Умение работать с электронными базами данных и библиотечными каталогами, умение применять известные методы научных исследований, умение представлять результаты проектных исследований по	Способность работать с данными, каталогов для исследования путем поиска информации в корпусе;

				изучаемой проблеме и собственных исследований, умение применять методы научных исследований учитывая новизну собственных концептуальных решений с использованием корпуса.	
	Владеет навыками алгоритмизации составления корпуса.	Практически не владеет навыками алгоритмизации составления корпуса.	Владеет навыками алгоритмизации составления простейшего корпуса без разметки.	Владеет навыками алгоритмизации составления корпуса с морфологической разметкой.	Владеет навыками алгоритмизации составления корпуса со всеми видами разметки.
ПК -9.3Имеет практический опыт разработки электронных языковых ресурсов; опыт применения основных методов, способов и средств получения, хранения, переработки информации	Знает основные принципы разработки электронных языковых ресурсов.	Знает только некоторые основные принципы разработки электронных языковых ресурсов.	Знает простейшие принципы разработки электронных языковых ресурсов.	Знает все принципы разработки электронных языковых ресурсов.	Знает основные принципы разработки электронных языковых ресурсов и умеет использовать их при решении лингвистического проекта.
	Умеет применять основные методы, способы и средства получения, хранения, переработки	Умеет применять некоторые методы, способы и средства получения, хранения, переработки информации в лингвистическом	Умеет применять базовые методы, способы и средства получения, хранения, переработки информации в	Умеет применять основные методы, способы и средства получения, хранения, переработки информации в лингвистическом	Умеет применять основные методы, способы и средства получения, хранения, переработки информации в лингвистическом корпусе на примере собственного лингвистического корпуса для решения проекта.

	информации в лингвистическом корпусе.	корпусе в неполном объеме.	лингвистическом корпусе.	корпусе.	
	Владеет навыками переработки языковых данных для составления лингвистического корпуса.	Владеет некоторыми навыками переработки языковых данных для составления лингвистического корпуса.	Владеет базовыми навыками переработки языковых данных для составления лингвистического корпуса	Владеет лингвистическими базами данных для применения навыка переработки языковых данных для составления лингвистического корпуса	Владеет навыками переработки языковых данных для составления лингвистического корпуса с использованием своих баз данных уникального характера.

Промежуточная аттестация студентов по дисциплине проводится в соответствии с локальными нормативными актами ДВФУ и является обязательной, в 5 семестре предусмотрен экзамен

Содержание экзамена

Вопросы к экзамену

1. Британский национальный корпус.
2. История лингвистических корпусов: от картотеки к корпусу.
3. Классификация (типология) корпусов.
4. Корпусная лингвистика: современное состояние.
5. Основные понятия корпусной лингвистики.
6. Направления в лингвистике, предвосхитившие появление корпусной лингвистики: от картотеки к корпусу
7. История создания лингвистических корпусов.
8. Основные характеристики корпусов.
9. Репрезентативность корпусов.
10. Классификация корпусов по различным основаниям.
11. Особые типы корпусов.
12. Понятие разметки.
13. Разметка. Средства разметки корпусов
14. Лингвистическая разметка.
15. Экстралингвистическая разметка.
16. Стандартизация в корпусной лингвистике.
17. Обзор существующих корпусов различных типов.
18. Зарубежные национальные корпусы.
19. Специальные корпусы.
20. Корпусные исследования.
21. Пользователи корпусов.
22. Способы использования корпусов.
23. Лексикографические исследования, основанные на корпусах.
24. Понятие гранулярности.

Критерии выставления оценки студенту на экзамене по дисциплине «Технологии корпусной лингвистики»

Баллы (рейтингов ой оценки)	Оценка зачета/ экзамена	Требования к сформированным компетенциям
	(стандартная)	

86-100	<i>«отлично»</i>	Оценка «отлично» выставляется студенту, если он глубоко и прочно усвоил программный материал, исчерпывающе, последовательно, четко и логически стройно его излагает, умеет тесно увязывать теорию с практикой, свободно справляется с задачами, вопросами и другими видами применения знаний, причем не затрудняется с ответом при видоизменении заданий, правильно обосновывает принятое решение, владеет разносторонними навыками и приемами выполнения практических задач.
76-85	<i>«хорошо»</i>	Оценка «хорошо» выставляется студенту, если он твердо знает материал, грамотно и по существу излагает его, не допуская существенных неточностей в ответе на вопрос, правильно применяет теоретические положения при решении практических вопросов и задач, владеет необходимыми навыками и приемами их выполнения.
61-75	<i>«удовлетворительно»</i>	Оценка «удовлетворительно» выставляется студенту, если он имеет знания только основного материала, но не усвоил его деталей, допускает неточности, недостаточно правильные формулировки, нарушения логической последовательности в изложении программного материала, испытывает затруднения при выполнении практических работ.
< 61	<i>«неудовлетворительно»</i>	Оценка «неудовлетворительно» выставляется студенту, который не знает значительной части программного материала, допускает существенные ошибки, неуверенно, с большими затруднениями выполняет практические работы. Как правило, оценка «неудовлетворительно» ставится студентам, которые не могут продолжить обучение без дополнительных занятий по соответствующей дисциплине.

Методические указания по сдаче экзамена

Экзамен принимается ведущим преподавателем. При большом количестве групп у одного преподавателя или при большой численности потока по распоряжению заведующего кафедрой (заместителя директора по учебной и воспитательной работе) допускается привлечение в помощь ведущему преподавателю других преподавателей. В первую очередь привлекаются преподаватели, которые проводили лабораторные занятия по дисциплине в группах.

В исключительных случаях, по согласованию с заместителем директора Школы по учебной и воспитательной работе, заведующий кафедрой имеет право принять экзамен в отсутствие ведущего преподавателя.

Форма проведения экзамена (устная, письменная и др.) утверждается на заседании кафедры по согласованию с руководителем в соответствии с рабочей программой дисциплины.

Во время проведения экзамена студенты могут пользоваться рабочей программой дисциплины, а также с разрешения преподавателя, проводящего экзамен, справочной литературой и другими пособиями (учебниками, учебными пособиями, рекомендованной литературой и т.п.).

Время, предоставляемое студенту на подготовку к ответу на экзамене, должно составлять не более 40 минут. По истечении данного времени студент должен быть готов к ответу.

Присутствие на зачете посторонних лиц (кроме лиц, осуществляющих проверку) без разрешения соответствующих лиц (ректора либо проректора по учебной и воспитательной работе, директора Школы, руководителя ОПОП или заведующего кафедрой), не допускается. Инвалиды и лица с ограниченными возможностями здоровья, не имеющие возможности самостоятельного передвижения, допускаются зачет с сопровождающими.

При промежуточной аттестации обучающимся устанавливается оценка.

В зачетную книжку студента вносится запись и вносится в экзаменационную ведомость. При неявке студента на зачёт в ведомости делается запись «не явился».