

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИИ



Федеральное государственное бюджетное образовательное учреждение высшего образования

«РОССИЙСКИЙ ГОСУДАРСТВЕННЫЙ ГУМАНИТАРНЫЙ УНИВЕРСИТЕТ»

(ФГБОУ ВО «РГГУ»)

Институт лингвистики

УНЦ компьютерной лингвистики

Рабочая программа дисциплины

«Корпусная лингвистика: построение и использование корпусов»

Направление подготовки 45.04.03 Фундаментальная и прикладная лингвистика

Магистерская программа: Фундаментальная и компьютерная лингвистика

Квалификация выпускника: магистр

Форма обучения: очная

РПД адаптирована для лиц
с ограниченными возможностями
здоровья и инвалидов

Москва 2021

Корпусная лингвистика: построение и использование корпусов

Рабочая программа дисциплины

Составитель:

к.ф.н. А.Ч.Пиперски

Ответственный редактор:

д. филол. н., профессор В.И.Подлеская

УТВЕРЖДЕНО

Протокол заседания УНЦ компьютерной
лингвистики

№ 7 от «15» апреля 2021г.

ОГЛАВЛЕНИЕ

1. Пояснительная записка

1.1 Цель и задачи дисциплины

1.2. Перечень планируемых результатов обучения по дисциплине, соотнесенных с индикаторами достижения компетенций

1.3. Место дисциплины в структуре образовательной программы

2. Структура дисциплины

3. Содержание дисциплины

4. Образовательные технологии

5. Оценка планируемых результатов обучения

5.1. Система оценивания

5.2. Критерии выставления оценок

5.3. Оценочные средства (материалы) для текущего контроля успеваемости, промежуточной аттестации обучающихся по дисциплине

6. Учебно-методическое и информационное обеспечение дисциплины

6.1. Список источников и литературы

6.2. Перечень ресурсов информационно-телекоммуникационной сети «Интернет»

7. Материально-техническое обеспечение дисциплины

8. Обеспечение образовательного процесса для лиц с ограниченными возможностями здоровья и инвалидов

Приложения

Приложение 1. Аннотация дисциплины

Приложение 2. Лист изменений

1. Пояснительная записка

1.1 Цель и задачи дисциплины

Предметом дисциплины является изучение методов создания корпусов и других электронных, специальных методов сбора материала с использованием корпусов ресурсов для лингвистических исследований грамматики и лексики, методов корпусных исследований.

Курс направлен на решение следующих задач:

- познакомить магистрантов с последними трендами в области корпусной лингвистики, с основными типами лингвистических ресурсов, доступными в Интернете и используемыми в профессиональной деятельности: с наиболее актуальными лингвистическими корпусами (национальными и проблемными), специальными программами обработки текстов, лексикографическими ресурсами и т.п.;
- на примерах показать, какие новые возможности в исследовании грамматики и лексики языка дает использование корпусных, а также применение современных методов обработки этих данных;
- дать представления о проблемах создания корпусов, об основных принципах разработки данных ресурсов и об основных требованиях, предъявляемых к ним;
- познакомить с технологиями и проблемами разметки корпусов;
- научить работать с современными пакетами обработки собственных корпусов;
- научить применять специальные методы лингвистических исследований, использующие данные корпусов, в том числе и статистические методы исследования;
- обучить практическим навыкам по применению корпусных методов в своей исследовательской работе.

1.2 Перечень планируемых результатов обучения по дисциплине, соотнесенных с индикаторами достижения компетенций

Компетенция	Индикаторы компетенций	Результаты обучения
УК-2 Способен управлять проектом на всех этапах его жизненного цикла	УК-2.2	Умеет соотносить разнородные явления и систематизировать их в рамках избранных видов профессиональной деятельности
	УК-2.3	Имеет практический опыт работы с информационными объектами и сетью Интернет, опыт библиографического разыскания, создания научных текстов
ПК-2 Владеет принципами создания электронных языковых ресурсов	ПК-2.1	Знает основные принципы обработки информации; базовые принципы корпусной лингвистики, лексикографии,

(текстовых, речевых и мультимодальных корпусов; словарей, тезаурусов, онтологий; фонетических, лексических, грамматических и иных баз данных и баз знаний) и умением пользоваться такими ресурсами		математической статистики; базовые представления о языковом разнообразии; наиболее полные и значимые лингвистические корпуса, электронные словари и базы данных
--	--	---

1.3 Место дисциплины в структуре образовательной программы

Дисциплина «Корпусная лингвистика: построение и использование корпусов» относится к части, формируемой участниками образовательных отношений.

В результате освоения дисциплины формируются знания, умения и владения, необходимые для изучения следующих дисциплин и прохождения практик: Русская корпусная грамматика.

2. Структура дисциплины

Общая трудоёмкость дисциплины составляет 3з.е., 114ч., в том числе контактная работа обучающихся с преподавателем 30ч., промежуточная аттестация 18ч., самостоятельная работа обучающихся 66ч.

№ п/п	Раздел Дисциплины	Семестр	Неделя семестра	Виды учебной работы, включая самостоятельную работу студентов и трудоемкость (в часах)				Формы текущего контроля успеваемости (по неделям семестра) Форма промежуточной аттестации (по семестрам)
				лекции	семинары	самостоятельная работа	контроль	
1.	Введение. Общее представление о корпусах и корпусной лингвистике. Стандарты разметки. Типы разметки корпусов.	1	1	2		4		ДЗ1. Упр. по теме “Особенности поиска и управления выдачи в НКРЯ”.
2.	Проблемные корпуса (параллельные, диалектные, мультимедийные и др.)	1	2		2	4		ДЗ2. Практическая работа. Разметка мультимедийного / аудиокорпуса

3.	Поиск в корпусе. Использование языка SQL для поиска в корпусе. Составление сложных запросов к корпусу.	1	3		2	4		ДЗ3. Запросы для поиска сложных конструкций с использованием языка регулярных выражений: письменный отчет
4.	Особенности различных типов разметки. Морфологическая разметка	1	4		1	4		ДЗ4. Практическая работа. Тестирование морфологической или синтаксической разметки: письменный отчет
5.	Особенности различных типов разметки. Синтаксическая разметка	1	5		2	4		ДЗ5. Практическая работа. Тестирование морфологической или синтаксической разметки: письменный отчет
6.	Особенности разметки: другие типы разметки	1	6		1	4		ДЗ6. Практическая работа. Семантическая разметка / анафорическая разметка
7.	Промежуточная аттестация	1	7		1	12		Контрольная работа по теме: «Корпуса и лингвистические ресурсы»
8.	Методы корпусных исследований. Анализ примеров корпусных исследований	1	7	2		4		ДЗ7. Реферирование статьи из рекомендованного списка. Составление краткого ТЗ к собственному исследовательскому проекту.
9.	Инструменты разметки собственного исследовательского корпуса	1	8		1	4		ДЗ8. Разработка параметров и схемы разметки исследовательского корпуса, разметка корпуса с помощью специальной программы
10.	Составление конкордансов, частотных списков, списков коллокаций с использованием специальных программ	1	9		1	4		ДЗ9. Обработка данных корпуса с использованием специальных программ

11.	Статистические методы исследований	1	10	2				
12.	Основные статистические характеристики данных. Меры средней тенденции и изменчивости, коэффициент вариации и др.	1	11		1	4		ДЗ10. Выполнение упр. по общему статистическому описанию данных. Первичный статистический анализ данных по проекту
13.	Статистические критерии зависимости	1	12		1	4		ДЗ11. Выполнение упр. по использованию статистических критериев. Проверка статистической гипотезы в рамках исследовательского проекта
14.	Методы корреляционного анализа	1	13		1	4		ДЗ12. Упр. по теме «Корреляционный анализ»
15.	Методы корпусной лексикографии.	1	14		1	4		ДЗ13. Практическое задание по теме «Статистические корпусные методы в лексикографии»
16.	Сравнение двух корпусов. Стилеметрия	1	15		1	4		ДЗ14. Анализ частотных словарей, сравнение словарей двух корпусов
17.	Корпусные исследования в грамматике. Пример. Корпусное исследование конструкции	1	16		1	4		ДЗ15. Реферирование статьи из рекомендованного списка.
18.	Обсуждение проекта	1	17		1			
19.	Экзамен							Контрольные вопросы
				6	24	66	18	

3. Содержание дисциплины

Содержание программы учебной дисциплины:

Часть 1. Введение. Общее представление о корпусах и корпусной лингвистике.

1.1. Краткая история. Предмет и задачи курса.

Краткая история корпусной лингвистики. Преимущества современных корпусных исследований. Возможность объединения формального и эмпирического подхода в

современной корпусной лингвистике. Компьютерные ресурсы, необходимые лингвистам для решения различных задач. Задачи, решаемые с помощью компьютерных ресурсов.

1.2. Основные понятия корпусной лингвистики.

Корпус. Национальный корпус. Проблемный корпус. Основные единицы. Основные требования, предъявляемые к корпусу. Поиск в корпусе. Основные требования и параметры поиска.

1.3. Стандарты и типы разметки. Основные принципы и методы разметки корпусов. Современные технологии разметки корпусов.

Часть 2. Корпуса и инструментарий работы с корпусами

2.1. Типы программ обработки текста, методы работы с программами обработки текста.

2.2. Программы разметки собственных исследовательских корпусов

Часть 3. Основные методы использования корпусов в грамматике и лексике

3.1. Области использования корпусных данных

3.2. Методы сбора и статистической обработки корпусных данных. Общие статистические характеристики: меры средней тенденции и изменчивости. Проверка статистических критериев, исследование зависимостей. Корреляционный анализ.

Часть 4. Примеры корпусных исследований

4.1. Примеры использования корпусов в обучении и научных исследованиях: методология создания дидактических материалов с использованием корпусов; методология создания исследовательского корпуса с использованием корпусов общего назначения. Примеры корпусных диалектологических, диахронических, социолингвистических и гендерных исследований, исследований стиля

4.3. Использование корпусов в лексикографической работе. Статистические методы в лексикографии

4.4. Сравнение корпусов. Стилеметрия.

4.5. Примеры применения корпусного анализа в грамматических исследованиях

4.6. Специальные лексикографические и грамматические ресурсы, созданные на основе корпусов, включая FrameNet, PropBank и т.д.

4. Образовательные технологии

Интерактивные формы обучения в данном курсе предполагают:

1. систематическое использование компьютерных презентаций (как преподавателем в установочной части занятия, так и студентом, выступающим с критическим разбором реферируемого научного сочинения);
2. он-лайн демонстрации работы с лингвистическими базами данных и энциклопедическими интернет-ресурсами (информационно-справочный ресурс по языкам мира Etnolog <http://www.ethnologue.com> (частично-платный ресурс, социолингвистическая информация, ареал распространения, карты); типологическая база данных WALS, The World Atlas of Language Structures <http://wals.info>; информационно-справочный ресурс по языкам мира <http://glottolog.org> (генеалогическая аффилиация и обширная библиография); многоязычная платформа для составления конкордансов и исследования

грамматической и лексической дистрибуции WordScetchEngine <https://www.sketchengine.co.uk/> ; лексикологическая база данных WordNet (<http://wordnet.princeton.edu/>; электронная энциклопедия по лингвистике серии Oxford Research Encyclopedia <http://linguistics.oxfordre.com/> ; и др.;

3. использование открытых он-лайн аудио- и видео обучающих ресурсов (для академического английского: www.cambridge.org/elt/english-for-academics);
4. практическую работу с электронными анализаторами речи и платформами лингвистического документирования (платформы E-Language Archiving Technology, ELAN <http://www.lat-mpi.eu/tools/elan>; коллекция электронных ресурсов на сайте Summer Institute of Linguistics (www.sil.org), в том числе, программа для акустического анализа речи Speech analyzer, программа IPA Help для обучения и пользования Международной фонетической транскрипцией, комплекс программ для полевой работы; программа для акустического анализа речи PRAAT (<http://www.fon.hum.uva.nl/praat>); открытое он-лайн издание по документированию и консервации языков Language Documentation & Conservation, LD&C <http://www.nflrc.hawaii.edu/ldc>).
5. В период временного приостановления посещения обучающимися помещений и территории РГГУ. для организации учебного процесса с применением электронного обучения и дистанционных образовательных технологий могут быть использованы следующие образовательные технологии:
 - видео-лекции;
 - онлайн-лекции в режиме реального времени;
 - электронные учебники, учебные пособия, научные издания в электронном виде и доступ к иным электронным образовательным ресурсам;
 - системы для электронного тестирования;
 - консультации с использованием телекоммуникационных средств.

5. Оценка планируемых результатов обучения

5.1. Система оценивания

При выставлении оценки в ведомость и в зачетную книжку преподаватель должен указать результат в соответствии с традиционной шкалой оценок и со шкалой оценок Европейской системы переноса и накопления кредитов (European Credit Transfer System; далее – ECTS) в соответствии с таблицей:

100-балльная шкала	Традиционная шкала		Шкала ECTS
95 – 100	отлично	зачтено	A
83 – 94			B
68 – 82	хорошо		C
56 – 67	удовлетворительно		D
50 – 55			E
20 – 49	неудовлетворительно	не зачтено	FX
0 – 19			F

Распределение баллов по видам учебной деятельности таково:

- посещение семинарских занятий – до 8 баллов,
- уровень активности студента при подготовке к занятиям (конспектирование специальной литературы, готовность отвечать на вопросы по анализу кейсов, активное участие в дискуссиях, коллоквиумах и мозговом штурме и проч.) и во

время проведения занятий (участие в обсуждениях и выполнении коллективных заданий) – всего до 32 баллов,

- качество выполнения контрольной работы (текущая аттестация) – до 20 баллов,
- успешность выполнения итогового творческого задания – до 40 баллов.

Оценка «зачтено» выставляется, если студент набрал в сумме не менее 50 баллов. Магистрант, не набравший в сумме 50 баллов, сдаёт зачёт по всему курсу и предъявляет преподавателю собственноручно написанные конспекты специальной литературы и выполненные домашние задания ко всем семинарам.

5.2. Критерии выставления оценок

Баллы/ Шкала ECTS	Оценка по дисциплине	Критерии оценки результатов обучения по дисциплине
100-83/ А,В	«отлично»/ «зачтено (отлично)»/ «зачтено»	<p>Выставляется обучающемуся, если он глубоко и прочно усвоил теоретический и практический материал, может продемонстрировать это на занятиях и в ходе промежуточной аттестации.</p> <p>Обучающийся исчерпывающе и логически стройно излагает учебный материал, умеет увязывать теорию с практикой, справляется с решением задач профессиональной направленности высокого уровня сложности, правильно обосновывает принятые решения.</p> <p>Свободно ориентируется в учебной и профессиональной литературе.</p> <p>Оценка по дисциплине выставляется обучающемуся с учётом результатов текущей и промежуточной аттестации.</p> <p>Компетенции, закреплённые за дисциплиной, сформированы на уровне – «высокий».</p>
82-68/ С	«хорошо»/ «зачтено (хорошо)»/ «зачтено»	<p>Выставляется обучающемуся, если он знает теоретический и практический материал, грамотно и по существу излагает его на занятиях и в ходе промежуточной аттестации, не допуская существенных неточностей.</p> <p>Обучающийся правильно применяет теоретические положения при решении практических задач профессиональной направленности разного уровня сложности, владеет необходимыми для этого навыками и приёмами.</p> <p>Достаточно хорошо ориентируется в учебной и профессиональной литературе.</p> <p>Оценка по дисциплине выставляется обучающемуся с учётом результатов текущей и промежуточной аттестации.</p> <p>Компетенции, закреплённые за дисциплиной, сформированы на уровне – «хороший».</p>

Баллы/ Шкала ECTS	Оценка по дисциплине	Критерии оценки результатов обучения по дисциплине
67-50/ D,E	«удовлетворительно»/ «зачтено (удовлетворительно)»/ «зачтено»	<p>Выставляется обучающемуся, если он знает на базовом уровне теоретический и практический материал, допускает отдельные ошибки при его изложении на занятиях и в ходе промежуточной аттестации.</p> <p>Обучающийся испытывает определённые затруднения в применении теоретических положений при решении практических задач профессиональной направленности стандартного уровня сложности, владеет необходимыми для этого базовыми навыками и приёмами.</p> <p>Демонстрирует достаточный уровень знания учебной литературы по дисциплине.</p> <p>Оценка по дисциплине выставляется обучающемуся с учётом результатов текущей и промежуточной аттестации.</p> <p>Компетенции, закреплённые за дисциплиной, сформированы на уровне – «достаточный».</p>
49-0/ F,FX	«неудовлетворительно»/ не зачтено	<p>Выставляется обучающемуся, если он не знает на базовом уровне теоретический и практический материал, допускает грубые ошибки при его изложении на занятиях и в ходе промежуточной аттестации.</p> <p>Обучающийся испытывает серьёзные затруднения в применении теоретических положений при решении практических задач профессиональной направленности стандартного уровня сложности, не владеет необходимыми для этого навыками и приёмами.</p> <p>Демонстрирует фрагментарные знания учебной литературы по дисциплине.</p> <p>Оценка по дисциплине выставляется обучающемуся с учётом результатов текущей и промежуточной аттестации.</p> <p>Компетенции на уровне «достаточный», закреплённые за дисциплиной, не сформированы.</p>

5.3. Оценочные средства для текущего контроля успеваемости и промежуточной аттестации

Текущий контроль успеваемости студентов проводится в следующих формах: выполнение домашних заданий (15 заданий – 50 баллов максимум); тестовое задание (максимально 10 баллов); защита исследовательского проекта (максимально - 40 баллов). Для получения удовлетворительной оценки необходимо набрать минимум 60 баллов. Самостоятельная работа: по темам семинарских заданий студенты выполняют 15 домашних заданий, из них 6 выполняется в группах по 2-3 человека. Кроме того, каждая группа студентов (2-3 человека) по согласованию с преподавателем выбирают одну из тем небольшого корпусного исследования грамматики или лексики из списка, предложенного

преподавателем, и защищает свое исследование в виде презентации, краткого отчета и обсуждения на зачетном занятии. При подготовке к тесту предполагается знакомство с полным списком обязательной литературы.

В качестве домашних заданий предлагаются задания следующих типов

- Д31. Упр. по теме “Особенности поиска и управления выдачи в НКРЯ”.
- Д32. Практическая работа. Разметка мультимедийного / аудиокорпуса.
- Д33. Запросы для поиска сложных конструкций с использованием языка регулярных выражений: письменный отчет.
- Д34. Практическая работа. Тестирование морфологической или синтаксической разметки: письменный отчет.
- Д35. Практическая работа. Тестирование морфологической или синтаксической разметки: письменный отчет.
- Д36. Практическая работа. Семантическая разметка / анафорическая разметка.
Контрольная работа по теме: «Корпуса и лингвистические ресурсы».
- Д37. Реферирование статьи из рекомендованного списка. Составление краткого ТЗ к собственному исследовательскому проекту.
- Д38. Разработка параметров и схемы разметки исследовательского корпуса, разметка корпуса с помощью специальной программы.
- Д39. Обработка данных корпуса с использованием специальных программ.
- Д310. Выполнение упр. по общему статистическому описанию данных. Первичный статистический анализ данных по проекту.
- Д311. Выполнение упр. по использованию статистических критериев. Проверка статистической гипотезы в рамках исследовательского проекта.
- Д312. Упр. по теме «Корреляционный анализ».
- Д313. Практическое задание по теме «Статистические корпусные методы в лексикографии».
- Д314. Анализ частотных словарей, сравнение словарей двух корпусов.
- Д315. Реферирование статьи из рекомендованного списка.

Экзамен ориентирован на следующие контрольные вопросы

- Определение корпуса. Задачи, решаемые с использованием корпусов.
- Параметры поиска в корпусе.
- Требования к корпусу, предъявляемые пользователем.
- Характеристики основных корпусов.
- Типы проблемных корпусов.
- Характеристика лингвистических ресурсов одного из типов: словари, тезаурусы, специальные лексикографические ресурсы: фреймбанки и т.п., лингвистические базы данных.
- Особенности проблемных корпусов.
- Виды разметки в корпусе.
- Корпуса в психо- и социолингвистических исследований

6. Учебно-методическое и информационное обеспечение дисциплины

6.1 Список источников и литературы

Основная литература

1. Баранов А.Н. Корпусная лингвистика // Баранов А.Н. Введение в прикладную лингвистику. М., 2003. С. 112–137.
2. Ермолаев О.Ю. Математическая статистика для психологов. 2-е изд., испр. - М.: МПСИ, Флинта, 2003 - 336 с. С. 10-64, с. 101-175, с. 202-222
3. Захаров В.П. Корпусная лингвистика: Учебно-метод. пособие. – СПб., 2005. – 48 с. Доступно для скачивания по адресу: <http://vp-zakharov.narod.ru/publications.html>
4. Захаров В.П., Хохлова М.В. Анализ эффективности статистических методов выявления коллокаций в текстах на русском языке // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог» 2010. Выпуск 9 (16). <http://www.dialog-21.ru/dialog2010/materials/pdf/22.pdf>
5. Кустова Г. И., Ляшевская О. Н., Падучева Е. В., Рахилина Е. В. Семантическая разметка лексики в Национальном корпусе русского языка: принципы, проблемы, перспективы. // Национальный корпус русского языка: 2003-2005. Результаты и перспективы. - М., 2005. С. 155-174.
6. Ляшевская О. Н., Плунгян В. А., Сичинава Д. В. О морфологическом стандарте Национального корпуса русского языка. // Национальный корпус русского языка: 2003-2005. Результаты и перспективы. - М., 2005. С. 111-135.
7. Ляшевская О.Н., Кузнецова Ю.Л. Русский фреймнет: к задаче создания корпусного словаря конструкций // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог 2009» (Бекасово, 27-31 мая 2009 г.). Вып. 8 (15). М.: РГГУ, 2009. С. 306-312. [pdf](#)
8. Ляшевская, О. Н. , Шаров С. А. Частотный словарь современного русского языка : на материалах Национального корпуса русского языка. М. : Азбуковник, 2009. XV, 1074 с. (1)
9. Резникова Т. И. Обзор общедоступных корпусов русского языка в Интернете. // НТИ, 2005, серия 2. №3. С. 2-9. (доступна с http://catalog.viniti.ru/search_extra.asp)
10. Савчук С.О. Метатекстовая разметка в Национальном корпусе русского языка: базовые принципы и основные функции // Национальный корпус русского языка: 2003-2005. Результаты и перспективы. - М., 2005. С. 62-88
11. Chris Manning and Hinrich Schütze, Chapter 5. Collocations. Foundations of Statistical Natural Language Processing, MIT Press. Cambridge, MA: May 1999. Доступна с сайта http://reslib.com/book/Foundations_of_Statistical_Natural_Language_Processing или <http://nlp.stanford.edu/fsnlp/promo/colloc.pdf>

Рекомендованная литература

1. Андрющенко В.М. Концепция и архитектура машинного фонда русского языка // Отв. ред.
2. Баранов А.Н. Корпусная лингвистика // Баранов А.Н. Введение в прикладную лингвистику. М., 2003. С. 112–137.
3. Гмурман В. Е. Руководство к решению задач по теории вероятностей и математической статистике. 9-е изд., стер.—М.: Высшая школа, 2004.— 404 с.
4. Добрушина Н. Р. [Как использовать Национальный корпус русского языка в образовании?](#) // Национальный корпус русского языка: 2003—2005. М.: Индрик, 2005. <http://ruscorpora.ru/sbornik2005/18dobrushina.pdf>
5. Доклады научной конференции «Корпусная лингвистика и лингвистические базы данных» / Под ред. А.С. Герда. СПб., 2002.

6. *Ермолаев О.Ю.* Математическая статистика для психологов. 2-е изд., испр. - М.: МПСИ, Флинта, 2003 - 336 с. С. 10-64, с. 101-175, с. 202-222
7. *Захаров В.П.* Корпусная лингвистика. Слайды к курсу лекций. (доступны по адресам <http://company.yandex.ru/academic/class/courses/zakharov.xml>, <http://teormin.ifmo.ru/education/corpus-linguistics/>)
8. *Захаров В.П.* Корпусная лингвистика: Учебно-метод. пособие. – СПб., 2005. – 48 с. Доступно для скачивания по адресу: <http://vp-zakharov.narod.ru/publications.htm>
9. *Захаров В.П., Хохлова М.В.* Анализ эффективности статистических методов выявления коллокаций в текстах на русском языке // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог» 2010. Выпуск 9 (16). <http://www.dialog-21.ru/dialog2010/materials/pdf/22.pdf>
10. *Зубов А.В.* Информационные технологии в лингвистике: Учебное пособие. - М.: Издательский центр «Академия», 2004. - 208 с.
11. Инструментарий русистики: корпусные подходы. Slavica Helsingiensia Series. Вып. 34. Helsinki 2008.
12. *Коваль С.А.* Роль корпуса в создании реалистичных моделей словоизменительной морфологии. URL: http://skowal.narod.ru/research/corpora2006/Koval_Corpora.2006.htm
13. *Кустова Г. И., Ляшевская О. Н., Падучева Е. В., Рахилина Е. В.* Семантическая разметка лексики в Национальном корпусе русского языка: принципы, проблемы, перспективы. // Национальный корпус русского языка: 2003-2005. Результаты и перспективы. - М., 2005. С. 155-174.
14. *Ляшевская О. Н., Плунгян В. А., Сичинава Д. В.* О морфологическом стандарте Национального корпуса русского языка. // Национальный корпус русского языка: 2003-2005. Результаты и перспективы. - М., 2005. С. 111-135.
15. *Ляшевская О.Н., Кузнецова Ю.Л.* Русский фреймнет: к задаче создания корпусного словаря конструкций // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог 2009» (Бекасово, 27-31 мая 2009 г.). Вып. 8 (15). М.: РГГУ, 2009. С. 306-312. [pdf](#)
16. *Ляшевская, О. Н., Шаров С. А.* Частотный словарь современного русского языка: на материалах Национального корпуса русского языка. М.: Азбуковник, 2009. XV, 1074 с. (1)
17. [Научно-техническая информация. Сер. 2. 2003. № 6. http://www.viniti.ru/cgi-bin/nti/nti.pl?action=show&year=2_2003&issue=6](http://www.viniti.ru/cgi-bin/nti/nti.pl?action=show&year=2_2003&issue=6)
18. Научно-техническая информация. Сер. 2. 2005. № 3.
19. Национальный корпус русского языка: 2003–2005. М.: Индрик, 2005.
20. Национальный корпус русского языка: 2006–2008. Новые результаты и перспективы. С.-Петербург 2009
21. *Носенко И.А.* Начала статистики для лингвистов. М., 1981.
22. *Овчинникова, И. Г., Угланова И. А.* Компьютерное моделирование вербальной коммуникации: учеб.-метод. пособие. М.: Флинта: Наука, 2009. 134с. (2)
23. *Резникова Т. И.* Обзор общедоступных корпусов русского языка в Интернете. // *НТИ*, 2005, серия 2. №3. С. 2-9. (доступна с http://catalog.viniti.ru/search_extra.asp)
24. *Савчук С.О.* Метатекстовая разметка в Национальном корпусе русского языка: базовые принципы и основные функции // Национальный корпус русского языка: 2003-2005. Результаты и перспективы. - М., 2005. С. 62-88

25. [Труды международного семинара по компьютерной лингвистике и ее приложениям «Диалог 2000», «Диалог 2001», «Диалог 2002», «Диалог 2003», «Диалог 2004», «Диалог 2005».](#) (<http://www.dialog-21.ru/materials/>)
26. Хэндауты к занятиям по теме «Статистическим методам в лингвистических исследованиях»
27. Шаров С.А. Представительный корпус русского языка в контексте мирового опыта. *НТИ*, 2003, серия 2. (<http://corpus.leeds.ac.uk/serge/publications/nti03-draft.pdf>)
28. A (very) brief description of the query syntax. <http://www.ims.uni-stuttgart.de/projekte/CorpusWorkbench/COPSyntax.html>
29. Biber D., S Conrad, R Reppen. *Corpus Linguistics: Investigating Language Structure and Use*. - Cambridge University Press, 1998. URL: <http://books.google.com/books?id=2h5F7TXa6psC>
30. Chris Manning and Hinrich Schütze, Chapter 5. Collocations. *Foundations of Statistical Natural Language Processing*, MIT Press. Cambridge, MA: May 1999. Доступна с сайта http://reslib.com/book/Foundations_of_Statistical_Natural_Language_Processing или <http://nlp.stanford.edu/fsnlp/promo/colloc.pdf>
31. Chris Manning and Hinrich Schütze, Chapter 7. Word sense disambiguation. Chapter 8. Lexical acquisition. *Foundations of Statistical Natural Language Processing*, MIT Press. Cambridge, MA: May 1999. Доступна с сайта http://reslib.com/book/Foundations_of_Statistical_Natural_Language_Processing
32. *Developing linguistic corpora: a guide to good practice*. Edited by Martin Wynne. URL: <http://www.ahds.ac.uk/creating/guides/linguistic-corpora/index.htm>
33. EAGLES: *Recommendations for the morphosyntactic annotation of corpora*, EAGTCWGMAC/ R. 1996. Available from <ftp://ftp.ilc.pi.cnr.it/pub/eagles/corpora/annotate.ps.gz>
34. Kennedy, Graeme. *An Introduction to Corpus Linguistics / Graeme Kennedy*. - London: Longman, 1998. - 315 p
35. Leech G. *Corpus Annotation Schemes // Literary and Linguistics Computing*. 1993. Vol. 8, N. 4. P. 275–281.
36. McEnery A., Wilson A. *Corpus Linguistics*. Edinburgh, 1996. (<http://www.lancs.ac.uk/fss/courses/ling/corpus/>)
37. Sinclair J. *Corpus, Concordance, Collocation*. Oxford University Press, 1991.
38. Woods, Anthony; Paul Fletcher & Arthur Hughes. 1986. *Statistics in Language Studies*. Cambridge: Cambridge University Press.

Программное обеспечение и Интернет-ресурсы

Учебная дисциплина должна быть обеспечена учебно-методической документацией и материалами. Обязательная литература должна быть представлена в библиотеке ВУЗа, сети Интернет или локальной сети ВУЗа (факультета). Для обучающихся должен быть обеспечен доступ к современным профессиональным базам данных, информационным справочным и поисковым системам. В частности, в процессе обучения используются следующие Интернет-ресурсы:

А. Корпуса и порталы по корпусной лингвистике

- Национальный корпус русского языка (<http://www.ruscorpora.ru>)
- Электронное издание Русской грамматики 1980 (<http://www.rusgram.narod.ru>)
- Сервис научной литературы JSTOR (<http://www.jstor.org/>)
- BNC <http://www.natcorp.ox.ac.uk/>

- COCA – Corpus of Contemporary American English (COCA)
<http://www.americancorpus.org/>
- Корпуса современного немецкого языка и оболочка для их исследования COSMAS II
<http://www.ids-mannheim.de/cosmas2/>
- <http://corpus.byu.edu/>

Б. Специальные программы обработки текстов

Б.1. Конкордансеры

AntConc - скачивается с http://www.antlab.sci.waseda.ac.jp/antconc_index.html

TextStat - <http://neon.niederlandistik.fu-berlin.de/en/textstat/>

SCP – Simple Concordance Program - <http://www.textworld.com/scp/>

<http://www.concordancesoftware.co.uk/>

Б.2. Инструменты для сбора данных по N-граммам

<http://www.textworld.com/>

kfNgram - <http://www.kwicfinder.com/kfNgram/kfNgramHelp.html>

В. Ресурсы на основе корпусов

<http://dict.ruslang.ru/> - словари на основе Национального корпуса русского языка

<http://www.sketchengine.co.uk/>

FrameBank

Г. Программы морфологической, синтаксической, семантической разметки

<http://aot.ru/>

<http://www.connexor.eu/technology/machinese/>

<http://beta.visl.sdu.dk/>

Д. Программы статистической обработки данных

Microsoft Excel, пакет языка R <http://www.r-project.org/>

Е. Программы разметки собственного исследовательского корпуса

- UAMStools – доступна с сайта [Programms\UAMCorpusToolInstaller1_3_2.exe](#)

Инструментарий и документация

Комплект необходимых материалов: раздаточного материала к семинарам, текстов заданий и подробных инструкций по их выполнению, необходимых программных средств, презентаций лекций, электронного учебника по корпусной лингвистике В.П.Захарова находится в электронной библиотеке кафедры.

Теоретический материал: презентации лекций ([Lect_materials](#)), дополнительный лекционный материал ([Lect_materials\Other_useful_lections](#)), учебник по корпусной лингвистике ([Lect_materials\Testbook_VZ-corpora.doc](#)), глава из учебника А.Н.Баранова: [Lect_materials\baranov_112-137.pdf](#), список основной литературы по курсу можно найти в папке: [Lect_materials](#) в файле [Lect_materials\Литература.doc](#). А также архив папки в письме «Материалы по корпусной лингвистике: теория».

Планы семинарских занятий

Занятие 1

Проблемные корпуса (параллельные, диалектные, мультимедийные и др.)

Цель занятия: познакомить студентов с основными типами проблемных корпусов, особенностями их организации и использования.

Основные ресурсы для аудиторной работы:

<http://ruscorpora.ru/search-para.html> - параллельный корпус

<http://ruscorpora.ru/search-murco.html> - мультимедийный корпус

<http://childes.psy.cmu.edu/> - Child Language Data Exchange System

Занятие 2

Поиск в корпусе. Использование языка SQP для поиска в корпусе. Составление сложных запросов к корпусу.

Цель занятия: научить студентов строить запросы к корпусам с использованием специального языка запросов CQP.

Основные ресурсы для аудиторной работы:

Текст задания и комментарии к выполнению находятся на диске с комплектом материалов по курсу в папке Tasks_and_Seminars\Regular_expressions\

Поиск производится в корпусах, доступных с сайта <http://corpus.leeds.ac.uk/ruscorpora.html> и <http://corpus.leeds.ac.uk/internet.html>

(с системой грамматических тэгов, используемых в разметке русских корпусов, можно ознакомиться: <http://corpus.leeds.ac.uk/mocky/msd-ru.html> *The current draft specification of the tagset is available from [msd-ru.html](http://corpus.leeds.ac.uk/mocky/msd-ru.html)*)

Дополнительные материалы:

<http://corpus.leeds.ac.uk/help.html>

<http://www.ims.uni-stuttgart.de/projekte/CorpusWorkbench/>

Занятие 3

Особенности различных типов разметки. Морфологическая разметка

Цель занятия: познакомить студентов с основными типами морфологической аннотации в корпусе, методами разметки и методами тестирования

Основная литература и источники для аудиторной работы:

Текст задания на диске «Материалы к курсу «Корпусные методы исследования грамматики и лексики»» в папке Tasks_and_Seminars\Tasks_and_Seminars\Test_pos_taggers/

Морфологические анализаторы для работы на семинаре:

Treetagger: <http://www.lsi.upc.edu/~nlp/SVMTool/demo.php> .

<http://www.connexor.eu/technology/machinese/machinesesyntax/>

Для русского языка можно протестировать разметчик Trigram, основанный на скрытых марковских моделях. Разметчик находится в папке \Programms\Trigram\RusCorpora.

MyStem: <http://company.yandex.ru/technology/mystem/?ncrnd=6499>

Морфологический разметчик для изучаемого студентом языка, если для него нет ссылки на странице <http://nlp.stanford.edu/links/statnlp.html>, если таковой существует и к нему есть некоммерческий доступ.

(для этого в поисковике надо набрать <Xxxx> pos-tagger on-line, Xxxx – интересующий студента язык)

Занятие 4

Особенности различных типов разметки. Синтаксическая разметка

Цель занятия: познакомить студентов с организацией синтаксически размеченных корпусов, с методами синтаксической разметки и методами ее тестирования.

Основная литература и источники для аудиторной работы:

Текст задания на диске «Материалы к курсу «Корпусные методы исследования грамматики и лексики»» в папке [Tasks and Seminars\Test synt parsing\](#)
Синтаксический корпус на сайте НКРЯ: <http://ruscorpora.ru/search-syntax.html>
Синтаксический анализатор на сайте aot.ru: <http://aot.ru/demo/synt.html>
Синтаксические анализаторы для других языков на сайтах:
<http://nlp.stanford.edu:8080/parser/>
<http://www.link.cs.cmu.edu/link/index.html>
<http://www2.lingsoft.fi/cgi-bin/engcg>

Занятие 5

Особенности различных типов разметки. Семантическая и анафорическая разметка

Цель занятия: познакомить студентов с видами семантической разметки, со специализированными лексикографическими ресурсами FrameNet, Фрейс Банк

Основная литература и источники для аудиторной работы:

Текст задания на диске «Материалы к курсу «Корпусные методы исследования грамматики и лексики»» в папке Tasks_and_Seminars\Tasks_and_Seminars\Sem_filer\
Литература:

1. *Кустова Г. И., Ляшевская О. Н., Падучева Е. В., Рахилина Е. В.* Семантическая разметка лексики в Национальном корпусе русского языка: принципы, проблемы, перспективы. // Национальный корпус русского языка: 2003-2005. Результаты и перспективы. - М., 2005. С. 155-174.
2. *Ляшевская О.Н., Кузнецова Ю.Л.* Русский фреймнет: к задаче создания корпусного словаря конструкций // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог 2009» (Бекасово, 27-31 мая 2009 г.). Вып. 8 (15). М.: РГГУ, 2009. С. 306-312. [pdf](#)

Интернет ресурсы:

<http://aot.ru/demo/graph.html>
<http://verbs.colorado.edu/~mpalmer/projects/verbnet.html>
<http://www.senseval.org/>

Занятие 6

Инструменты разметки собственного исследовательского корпуса

Цель занятия: познакомить студентов с технологиями разметки собственного исследовательского корпуса, научить размечать корпус с использованием специализированных программ

Основная литература и источники для аудиторной работы:

Текст задания на диске «Материалы к курсу «Корпусные методы исследования грамматики и лексики»» в папке Текст задания [Tasks_and_Seminars\Coder\Task_Coder.doc](#)
Программа разметки собственного корпуса UAMCorpusTool ([Programms\UAMCorpusToolInstaller1_3_2.exe](#)) и другие необходимые материалы находятся в папке [Tasks_and_Seminars\Coder](#). Инструкцию по работе с программой находится в файле [Tasks_and_Seminars\Coder\UAMCorpusToolManualv2.0.pdf](#)
Программу также можно скачать с сайта <http://www.wagsoft.com/CorpusTool/>

Занятие 7

Составление конкордансов, частотных списков, списков коллокаций с использованием специальных программ

Цель занятия: познакомить студентов со специальными программами обработки собственных корпусов: конкордансерами, их возможностями и параметрами, научить работать с программой обработки собственного корпуса – составлению конкордансов, частотных списков, списков n-грамм

Основная литература и источники для аудиторной работы:

Основные конкордансеры для работы находятся в папке Concord_Progr (а также скачиваются с сайтов: AntConc - скачивается с

http://www.antlab.sci.waseda.ac.jp/antconc_index.html

TextStat - <http://neon.niederlandistik.fu-berlin.de/en/textstat/>

SCP – Simple Concordance Program - <http://www.textworld.com/scp/>

<http://www.concordancesoftware.co.uk/>, kfNgram -

<http://www.kwicfinder.com/kfNgram/kfNgramHelp.html>)

Дополнительные сайты:

<http://aot.ru/search1.html>

<http://www.lex tutor.ca/concordancers/>

Занятие 8

Основные статистические характеристики данных. Меры средней тенденции и изменчивости, коэффициент вариации и др.

Цель занятия: познакомить студентов с основными понятиями описательной статистики, дать представление об основных статистических характеристиках лингвистических данных

Основная литература и источники для аудиторной работы:

Раздаточный материал по теме

Файл Excel с лингвистическими данными для обработки

Ермолаев О.Ю. Математическая статистика для психологов. 2-е изд., испр. - М.: МПСИ, Флинта, 2003 - 336 с. С. 10-64

Занятие 9

Статистические критерии зависимости. Проверка гипотез. Критерий хи-квадрат. Критерий Стьюдента. Критерий Манн-Уитни. Критерий Фишера.

Цель занятия: познакомить студентов с основными понятиями проверки статистических гипотез, научить применять различные статистические критерии для проверки гипотез при исследовании языковых данных

Основная литература и источники для аудиторной работы:

Раздаточный материал по теме

Файл Excel с лингвистическими данными для обработки

Ермолаев О.Ю. Математическая статистика для психологов. 2-е изд., испр. - М.: МПСИ, Флинта, 2003 - 336 с. С. 56-109, 125-178

Занятие 10

Корреляционный анализ

Цель занятия: познакомить студентов с основными понятиями корреляционного анализа, научить применять корреляционный анализ для исследования взаимозависимости различных факторов, влияющих на распределение некоторого языкового явления в корпусе / языке

Основная литература и источники для аудиторной работы:

Раздаточный материал по теме

Файл Excel с лингвистическими данными для обработки

Ермолаев О.Ю. Математическая статистика для психологов. 2-е изд., испр. - М.: МПСИ, Флинта, 2003 - 336 с. С. 202-217

Занятие 11

Методы корпусной лексикографии

Цель занятия: познакомить студентов с основными методами корпусных методов в лексикографии, научить использовать корпусной материал при создании различных лексикографических ресурсов (словарей, специальных словарей, тезаурусов), познакомить с основными методами статистической обработки данных, используемых при создании лексикографических ресурсов: методами выделения коллокаций, исследования синонимов, разрешения семантической неоднозначности, составления лексикографического портрета лексемы

Основная литература и источники для аудиторной работы:

Раздаточный материал по теме: методы выделения устойчивых словосочетаний

1. Chris Manning and Hinrich Schütze, Chapter 5. Collocations. Foundations of Statistical Natural Language Processing, MIT Press. Cambridge, MA: May 1999. Доступна с сайта http://reslib.com/book/Foundations_of_Statistical_Natural_Language_Processing или <http://nlp.stanford.edu/fsnlp/promo/colloc.pdf>
2. *Захаров В.П., Хохлова М.В.* Анализ эффективности статистических методов выявления коллокаций в текстах на русском языке // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог» 2010. Выпуск 9 (16). <http://www.dialog-21.ru/dialog2010/materials/pdf/22.pdf>

<http://webascorpus.org/searchwc.html>

<http://corpus.leeds.ac.uk/ruscorpora.html>

таблица excel с данными для составления лексикографического портрета лексемы

<http://www.textology.ru/razdel.aspx?id=33>

Занятие 12

Сравнение двух корпусов. Стилеметрия

Цель занятия: познакомить студентов с основными сравнения лексического состава корпусов, а также с параметрами и методами стилеметрии.

Основная литература и источники для аудиторной работы:

Раздаточный материал по теме

Гринбаум О.Н. Структуризация художественной прозы с использованием ЭВМ (II): детализация структурированного текста // Квантитативная лингвистика и автоматический анализ текстов. 872. Тарту, 1989. С. 12-24.

Занятие 13

Корпусные исследования в грамматике

Цель занятия: познакомить студентов с примерами использования корпусных методов в грамматических исследованиях, основными положениями грамматики конструкций и корпусными методами, используемыми в рамках данного направления.

Основная литература и источники для аудиторной работы:

Раздаточный материал по теме

<http://rusgram.ru/>

Статьи для докладов на семинаре (из сборников и доступные с сайта

<http://studiorum.ruscorpora.ru/>):

1. Инструментарий русистики: корпусные подходы. Slavica Helsingiensia Series. Вып. 34. Helsinki 2008.
2. Национальный корпус русского языка: 2006–2008. Новые результаты и перспективы. С.-Петербург 2009
3. Корпусные исследования по русской грамматике. Москва 2009.

Занятие 14

Обсуждение исследовательских проектов

Цель занятия: ознакомить студентов с различными примерами корпусных исследований на основе обсуждений, выполненных в рамках курса проектов, выработать навыки анализа корпусных исследований различных языковых явлений

Материалы для аудиторной работы:

Собственный исследовательский проект группы: презентация

Исследовательский проект другой группы для анализа и рецензирования

6.2 Перечень ресурсов информационно-телекоммуникационной сети «Интернет»

№п /п	Наименование
1	Международные реферативные наукометрические БД, доступные в рамках национальной подписки в 2020 г. Web of Science Scopus
2	Профессиональные полнотекстовые БД, доступные в рамках национальной подписки в 2020 г. Журналы Cambridge University Press ProQuest Dissertation & Theses Global SAGE Journals Журналы Taylor and Francis
3	Профессиональные полнотекстовые БД JSTOR Издания по общественным и гуманитарным наукам Электронная библиотека Grebennikon.ru
4	Компьютерные справочные правовые системы Консультант Плюс, Гарант

7. Материально-техническое обеспечение дисциплины

Занятия по курсу можно проводить с максимальной эффективностью, если проводить их в компьютерном классе с доступом в Интернет, проектором и экраном для презентаций. Необходимые программы:

№п /п	Наименование ПО	Производитель	Способ распространения
1	Adobe Master Collection CS4	Adobe	лицензионное
2	Microsoft Office 2010	Microsoft	лицензионное
3	Windows 7 Pro	Microsoft	лицензионное
7	Microsoft Share Point 2010	Microsoft	лицензионное
12	Windows 10 Pro	Microsoft	лицензионное
13	Kaspersky Endpoint Security	Kaspersky	лицензионное
14	Microsoft Office 2016	Microsoft	лицензионное
15	Visual Studio 2019	Microsoft	лицензионное
16	Adobe Creative Cloud	Adobe	лицензионное
17	Zoom	Zoom	лицензионное

8. Обеспечение образовательного процесса для лиц с ограниченными возможностями здоровья

В ходе реализации дисциплины используются следующие дополнительные методы обучения, текущего контроля успеваемости и промежуточной аттестации обучающихся в зависимости от их индивидуальных особенностей:

- для слепых и слабовидящих:
 - лекции оформляются в виде электронного документа, доступного с помощью компьютера со специализированным программным обеспечением;
 - письменные задания выполняются на компьютере со специализированным программным обеспечением, или могут быть заменены устным ответом;
 - обеспечивается индивидуальное равномерное освещение не менее 300 люкс;
 - для выполнения задания при необходимости предоставляется увеличивающее устройство; возможно также использование собственных увеличивающих устройств;
 - письменные задания оформляются увеличенным шрифтом;
 - экзамен и зачёт проводятся в устной форме или выполняются в письменной форме на компьютере.
- для глухих и слабослышащих:
 - лекции оформляются в виде электронного документа, либо предоставляется звукоусиливающая аппаратура индивидуального пользования;
 - письменные задания выполняются на компьютере в письменной форме;

- экзамен и зачёт проводятся в письменной форме на компьютере; возможно проведение в форме тестирования.

- для лиц с нарушениями опорно-двигательного аппарата:
 - лекции оформляются в виде электронного документа, доступного с помощью компьютера со специализированным программным обеспечением;
 - письменные задания выполняются на компьютере со специализированным программным обеспечением;
 - экзамен и зачёт проводятся в устной форме или выполняются в письменной форме на компьютере.

При необходимости предусматривается увеличение времени для подготовки ответа.

Процедура проведения промежуточной аттестации для обучающихся устанавливается с учётом их индивидуальных психофизических особенностей. Промежуточная аттестация может проводиться в несколько этапов.

При проведении процедуры оценивания результатов обучения предусматривается использование технических средств, необходимых в связи с индивидуальными особенностями обучающихся. Эти средства могут быть предоставлены университетом, или могут использоваться собственные технические средства.

Проведение процедуры оценивания результатов обучения допускается с использованием дистанционных образовательных технологий.

Обеспечивается доступ к информационным и библиографическим ресурсам в сети Интернет для каждого обучающегося в формах, адаптированных к ограничениям их здоровья и восприятия информации:

- для слепых и слабовидящих:
 - в печатной форме увеличенным шрифтом;
 - в форме электронного документа;
 - в форме аудиофайла.
- для глухих и слабослышащих:
 - в печатной форме;
 - в форме электронного документа.
- для обучающихся с нарушениями опорно-двигательного аппарата:
 - в печатной форме;
 - в форме электронного документа;
 - в форме аудиофайла.

Учебные аудитории для всех видов контактной и самостоятельной работы, научная библиотека и иные помещения для обучения оснащены специальным оборудованием и учебными местами с техническими средствами обучения:

- для слепых и слабовидящих:
 - устройством для сканирования и чтения с камерой SARA CE;
 - дисплеем Брайля PAC Mate 20;
 - принтером Брайля EmBraille ViewPlus;
- для глухих и слабослышащих:
 - автоматизированным рабочим местом для людей с нарушением слуха и слабослышащих;
 - акустический усилитель и колонки;
- для обучающихся с нарушениями опорно-двигательного аппарата:
 - передвижными, регулируемые эргономическими партами СИ-1;
 - компьютерной техникой со специальным программным обеспечением.

АННОТАЦИЯ ДИСЦИПЛИНЫ

Дисциплина реализуется в Институте лингвистики УНЦ лингвистической типологии.

Цель дисциплины:

Предметом дисциплины является изучение методов создания корпусов и других электронных, специальных методов сбора материала с использованием корпусов ресурсов для лингвистических исследований грамматики и лексики, методов корпусных исследований.

Задачи дисциплины:

- познакомить магистрантов с последними трендами в области корпусной лингвистики, с основными типами лингвистических ресурсов, доступными в Интернете и используемыми в профессиональной деятельности: с наиболее актуальными лингвистическими корпусами (национальными и проблемными), специальными программами обработки текстов, лексикографическими ресурсами и т.п.;
- на примерах показать, какие новые возможности в исследовании грамматики и лексики языка дает использование корпусных, а также применение современных методов обработки этих данных;
- дать представления о проблемах создания корпусов, об основных принципах разработки данных ресурсов и об основных требованиях, предъявляемых к ним;
- познакомить с технологиями и проблемами разметки корпусов;
- научить работать с современными пакетами обработки собственных корпусов;
- научить применять специальные методы лингвистических исследований, использующие данные корпусов, в том числе и статистические методы исследования;
- обучить практическим навыкам по применению корпусных методов в своей исследовательской работе.

Дисциплина направлена на формирование следующих компетенций:

Компетенция	Индикаторы компетенций	Результаты обучения
УК-2 Способен управлять проектом на всех этапах его жизненного цикла	УК-2.2	Умеет соотносить разнородные явления и систематизировать их в рамках избранных видов профессиональной деятельности
	УК-2.3	Имеет практический опыт работы с информационными объектами и сетью Интернет, опыт библиографического разыскания, создания научных текстов

<p>ПК-2 Владеет принципами создания электронных языковых ресурсов (текстовых, речевых и мультимодальных корпусов; словарей, тезаурусов, онтологий; фонетических, лексических, грамматических и иных баз данных и баз знаний) и умением пользоваться такими ресурсами</p>	<p>ПК-2.1</p>	<p>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</p>
--	---------------	---

По дисциплине предусмотрена промежуточная аттестация в форме экзамена.
Общая трудоемкость освоения дисциплины составляет 3 зачетных единиц.

ЛИСТ ИЗМЕНЕНИЙ

№	Текст актуализации или прилагаемый к РПД документ, содержащий изменения	Дата	№ протокола
1	Приложение №		