

МИНОБРНАУКИ РОССИИ



Федеральное государственное автономное образовательное учреждение
высшего образования
«Российский государственный гуманитарный университет»
(ФГАОУ ВО «РГГУ»)

ИНСТИТУТ ЛИНГВИСТИКИ

Кафедра теоретической и прикладной лингвистики

Информационные технологии и корпусные исследования в лингвистике

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Направление 45.04.02 «Лингвистика»

Направленность «Иностранные языки»

Уровень высшего образования: магистратура

Форма обучения: очная

РПД адаптирована для лиц
с ограниченными возможностями
здоровья и инвалидов

Москва 2026

Информационные технологии и корпусные исследования в лингвистике
Рабочая программа дисциплины

Составители:

профессор департамента филологии НИУ ВШЭ - Санкт-Петербург *М.В. Копотев*

доктор филологических наук, профессор кафедры ТиПЛ *С.А. Крылов*

кандидат филологических наук, доцент УНЦ компьютерной лингвистики *А.Ч. Пиперски*

Рецензенты:

к.ф.н., доцент кафедры ТиПЛ *С.Ю. Семенова*

УТВЕРЖДЕНО

Протокол заседания кафедры теоретической
и прикладной лингвистики

№ 12 от 15.12.2025

ОГЛАВЛЕНИЕ

1.	Пояснительная записка	4
1.1.	Цель и задачи дисциплины	4
1.2.	Перечень планируемых результатов обучения по дисциплине, соотнесенных с индикаторами достижения компетенций	4
1.3.	Место дисциплины в структуре образовательной программы	4
2.	Структура дисциплины	4
3.	Содержание дисциплины	5
4.	Образовательные технологии	6
5.	Оценка планируемых результатов обучения	6
5.1.	Система оценивания	6
5.2.	Критерии выставления оценки по дисциплине	7
5.3.	Оценочные средства (материалы) для текущего контроля успеваемости, промежуточной аттестации обучающихся по дисциплине	8
6.	Учебно-методическое и информационное обеспечение дисциплины	8
6.1.	Список источников и литературы	8
6.2.	Перечень ресурсов информационно-телекоммуникационной сети «Интернет»	8
6.3.	Профессиональные базы данных и информационно-справочные системы	9
7.	Материально-техническое обеспечение дисциплины	9
8.	Обеспечение образовательного процесса для лиц с ограниченными возможностями здоровья и инвалидов	10
9.	Методические материалы	11
9.1.	Планы семинарских/ практических/ лабораторных занятий	11
9.2.	Методические рекомендации по подготовке письменных работ	11
9.3.	Иные материалы	12
	Приложение 1. Аннотация рабочей программы дисциплины	13

1. Пояснительная записка

1.1. Цель и задачи дисциплины

Цель дисциплины – познакомить магистрантов с наиболее актуальными современными компьютерными корпусами текстов и лексикографическими ресурсами, программами обработки текста, с технологиями создания собственных исследовательских корпусов, научить применять методы создания собственных исследовательских корпусов, работы с корпусными данными, методы обработки этих данных в собственных научных исследованиях, а также познакомить с современными исследованиями и особенностями языка, выявленными на базе корпусных исследований.

Образовательными задачами дисциплины являются:

(а) ознакомление студентов с ключевыми аспектами современной корпусной лингвистики, а именно, с основными научными направлениями и с русскоязычной и англоязычной терминологией;

(б) изучение устройства корпусов разных языков;

(в) ознакомление с типами исследований, проводящихся на базе корпусов;

(г) изучение основ корпусной педагогики;

(д) конечная перспектива дисциплины — познакомить студентов с корпусами различных языков, научить их пользоваться корпусными ресурсами, показать, каким образом лингвисты и педагоги работают с корпусами, сформировать у студентов базовые навыки корпусной разметки.

Практические задачи дисциплины:

(а) ознакомление студентов с ключевыми аспектами современной корпусной лингвистики, а именно, с основными принципами аннотирования и методами ведения исследования;

(б) изучение на материале конкретных корпусов типов междисциплинарных корпусных исследований;

(в) ознакомление с интересными научно-исследовательскими задачами в каждой из рассмотренных областей;

(г) ознакомление с новыми возможностями в исследовании грамматики и лексики языка, которые дают использование корпусных методов, а также с применением современных методов обработки этих данных;

(д) ознакомление с технологиями и проблемами разметки корпусов;

(е) обучение практическим навыкам по применению корпусных методов в своей исследовательской работе.

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

ОПК 6 Способен применять современные технологии при осуществлении сбора, обработки и интерпретации данных эмпирического исследования; составлять и оформлять научную документацию;

ОПК 6.2 Готов применять в практической деятельности современные технологии сбора, обработки и интерпретации данных эмпирического исследования

ОПК 7 Способен работать с основными информационно-поисковыми и экспертными системами, системами представления знаний и обработки вербальной информации.

ОПК 7.1 Имеет навыки работы с основными информационно-поисковыми и экспертными системами, системами представления знаний и обработки вербальной информации.

ОПК 7.2 Готов использовать в профессиональной деятельности основные информационно-поисковые и экспертные системы, системы представления знаний и обработки вербальной информации

Компетенция (код и наименование)	Индикаторы компетенций (код и наименование)	Результаты обучения
<p><i>УК-6</i> Способен применять современные технологии при осуществлении сбора, обработки и интерпретации данных эмпирического исследования; составлять и оформлять научную документацию;</p>	<p><i>УК-6.2</i> Готов применять в практической деятельности современные технологии сбора, обработки и интерпретации данных эмпирического исследования</p>	<p>Знать</p> <ul style="list-style-type: none"> ▪ принципы создания собственных исследовательских корпусов; ▪ основные типы исследовательских задач, решаемых с использованием корпусов; ▪ основные применяемые в корпусных исследованиях лексикологии и грамматики методы ▪ требования, предъявляемые к верификации результатов ▪ основные методы статистического анализа корпусных данных. <p>Уметь:</p> <ul style="list-style-type: none"> ▪ осуществлять поиск в корпусах в соответствии с исследовательской гипотезой в области грамматики и лексикографических исследований; ▪ создавать и размечать собственные исследовательские и обучающие корпуса; ▪ работать с различными типами программ обработки текстов: конкордансерами, программами для поиска коллокаций, создания частотных списков и т.п., корпусными менеджерами; ▪ разрабатывать

		<p>методический материал по основным языковым дисциплинам с использованием корпусов;</p> <ul style="list-style-type: none"> ▪ осуществлять мониторинг и оценку различных типов современных корпусных ресурсов и выбирать ресурсы, подходящие для выполнения тех или иных исследовательских и производственных задач. ▪ Владеть: <ul style="list-style-type: none"> ▪ основными методами и средствами профессионального компьютерного инструментария для исследовательской и практической работы; ▪ методами сбора материала с использованием корпусов; ▪ методами анализа корпусных данных, включая статистические методы.
<p><i>ОПК 7</i> Способен работать с основными информационно-поисковыми и экспертными системами, системами представления знаний и обработки вербальной информации.</p>	<p><i>ОПК 7.1</i> Имеет навыки работы с основными информационно-поисковыми и экспертными системами, системами представления знаний и обработки вербальной информации.</p>	<p>Знать:</p> <ul style="list-style-type: none"> ▪ основные принципы создания корпусов и других компьютерных лингвистических ресурсов; ▪ характеристики и особенности современных доступных в Интернете национальных и проблемных корпусов, широко используемых в лингвистических исследованиях, включая недавно вошедшие в лингвистическую практику; <p>Уметь:</p> <ul style="list-style-type: none"> ▪ осуществлять поиск в корпусах в соответствии с исследовательской гипотезой в области грамматики и лексикографических исследований; ▪ создавать и размечать собственные исследовательские и обучающие корпуса;

		<ul style="list-style-type: none"> ▪ работать с различными типами программ обработки текстов: конкордансерами, программами для поиска коллокаций, создания частотных списков и т.п., корпусными менеджерами; <p>Владеть:</p> <ul style="list-style-type: none"> • основными методами и средствами профессионального компьютерного инструментария для исследовательской и практической работы.
	<p><i>ОПК 7.2</i> Готов использовать в профессиональной деятельности основные информационно-поисковые и экспертные системы, системы представления знаний и обработки вербальной информации</p>	<p>Знать:</p> <ul style="list-style-type: none"> ▪ основные принципы создания корпусов и других компьютерных лингвистических ресурсов; ▪ стандарты, типы и проблемы разметки корпусов, включая такие современные типы разметки, как дискурсивную разметку, интонационную разметку устных корпусов и т.п., применяемые в разметке технологии; <p>Уметь:</p> <ul style="list-style-type: none"> ▪ осуществлять поиск в корпусах в соответствии с исследовательской гипотезой в области грамматики и лексикографических исследований; ▪ создавать и размечать собственные исследовательские и обучающие корпуса; ▪ работать с различными типами программ обработки текстов: конкордансерами, программами для поиска коллокаций, создания частотных списков и т.п., корпусными менеджерами; ▪ разрабатывать методический материал по основным языковым дисциплинам

		использованием корпусов. Владеть: <ul style="list-style-type: none"> • основными методами и средствами профессионального компьютерного инструментария для исследовательской и практической работы.
--	--	---

1.3. Место дисциплины в структуре образовательной программы

Дисциплина «Информационные технологии и корпусные исследования в лингвистике» является частью Блока 1 учебного плана ОП ВО магистратуры «Иностранные языки» по направлению подготовки «45.04.02 – Лингвистика» и имеет статус дисциплины Обязательной части.

Для освоения дисциплины необходимы знания, умения и владения, сформированные в ходе изучения курсов «Общее языкознание и история лингвистических учений», «Методы лингвистического анализа» и «Семиотика».

Дисциплина формирует компетенции, необходимые для прохождения практики «Научно-исследовательская работа», преддипломной практики и итоговой аттестации.

2. Структура дисциплины

Общая трудоёмкость дисциплины составляет 4 з.е., 144 ч., в том числе контактная работа обучающихся с преподавателем 40 ч., самостоятельная работа обучающихся 104 ч., включая 18ч. подготовку к экзамену.

Структура дисциплины для очной формы обучения

Объем дисциплины в форме контактной работы обучающихся с педагогическими работниками и (или) лицами, привлекаемыми к реализации образовательной программы на иных условиях, при проведении учебных занятий:

Семестр	Тип учебных занятий	Количество часов
3	Лекции	-
3	Семинары	40
Всего:		40

Объем дисциплины (модуля) в форме самостоятельной работы обучающихся составляет 104 академических часов, включая 18ч. подготовку к экзамену.

3. Содержание дисциплины

В соответствии с целями и задачами данного курса в программу включены следующие части и разделы.

Часть 1.

РАЗДЕЛ I. 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.

РАЗДЕЛ II. 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 2.

РАЗДЕЛ III. Основные методы использования корпусов в исследованиях грамматики и лексики

- 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.

РАЗДЕЛ IV. Примеры корпусных исследований

- 4.1. Примеры использования корпусов в обучении и в 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. Образовательные технологии

Для проведения учебных занятий по дисциплине используются различные образовательные технологии. Для организации учебного процесса может быть использовано электронное обучение и (или) дистанционные образовательные технологии.

5. Оценка планируемых результатов обучения

5.1 Система оценивания

Оценка за семестр складывается из следующих составляющих (максимальная сумма 100 баллов):

- 1) оценки за посещение семинаров (всего 10 баллов) и активную работу на них (до 10 баллов) – итого за работу на семинарах до 20 баллов;
- 2) оценка за текущую контрольную работу (до 10 баллов);
- 3) оценка за разработку проекта / доклада по теме (до 20 баллов);
- 4) оценка за презентацию проекта / выступление с докладом (до 10 баллов);
- 5) итоговая контрольная работа (до 20 баллов);
- 6) итоговое собеседование (до 20 баллов).

Для получения высокой оценки студенту необходимо систематически демонстрировать устойчивые результаты обучения.

Полученный совокупный результат конвертируется в традиционную шкалу оценок и в шкалу 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
0 – 19		F	

5.2 Критерии выставления оценки по дисциплине

Баллы/ Шкала ECTS	Оценка по дисциплине	Критерии оценки результатов обучения по дисциплине
100-83/ A,B	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>

Баллы/ Шкала ECTS	Оценка по дисциплине	Критерии оценки результатов обучения по дисциплине
82-68/ С	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>
67-50/ D,E	удовлетво- рительно/ 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	неудовлет- ворительно/ 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 (материалы) для текущего контроля успеваемости, промежуточной аттестации обучающихся по дисциплине

Ниже приводятся контрольные вопросы, образцы домашних работ, тестов и контрольных работ, которые могут использоваться для оценивания уровня усвоения материала по данной дисциплине. Посещаемость, работы на занятии, выполнение текущих домашних заданий.

Тематика заданий текущего контроля (УК 1.1, ОПК 7.1, 7.2):

Как устроен НКРЯ? Как устроены корпуса BNC, СОСА, СОНА?

Чем разметка в НКРЯ отличается от разметки в ОК НКРЯ?

Какой тип синтаксической разметки можно считать оптимальным?

Каким образом составлять исследовательские проекты школьников на базе корпусов?

Как использовать списки частотности в преподавании?

Вопросы для оценки качества освоения дисциплины (УК 1.1, ОПК 7.1, 7.2):

1. Целесообразность применения электронных технологий в преподавании языков.
2. Понятие «корпус».
3. История возникновения корпусной лингвистики.
4. Виды корпусов.
5. Принципы аннотирования.
6. Достоинства и недостатки разных видов исследований на базе корпусов.

Критерии оценки для промежуточной аттестации обучающихся (вопросы к зачету)

-результат, содержащий полный правильный ответ, полностью соответствующий требованиям критерия – 85 – 100 %;

-результат, содержащий неполный правильный ответ (степень полноты ответа – более 75%) или ответ, содержащий незначительные неточности, т.е. ответ, имеющий незначительные отступления от требований критерия, – 75 – 84% от максимального количества баллов;

-результат, содержащий неполный правильный ответ (степень полноты ответа – до 75%) или ответ, содержащий незначительные неточности, т.е. ответ, имеющий незначительные отступления от требований критерия – 60 -74 % от максимального количества баллов;

-результат, содержащий неполный правильный ответ, содержащий значительные неточности, ошибки (степень полноты ответа – менее 60%) – до 60 % от максимального количества баллов;

-неправильный ответ (ответ не по существу задания) или отсутствие ответа, т.е. ответ, не соответствующий полностью требованиям критерия, – 0 % от максимального количества баллов.

Образцы домашних заданий (ОПК 6.2, 7.1, 7.2)

1. Укажите, сколько раз в Основном корпусе НКРЯ встречается слово *сверхпроводимость* во всех формах. Выразите его частотность в ipm.
2. Сравните частотность употребления двусложных сравнительных союзов *точно, будто и словно* у нескольких русских поэтов по Поэтическому корпусу НКРЯ. Как изменяется частотность этих союзов во времени? С какими трудностями вы столкнулись при поиске?
3. Кто из русских писателей по данным НКРЯ реже всего употребляет в своей прозе союз *и* — Ф. М. Достоевский, М. А. Булгаков или М. А. Шолохов? Опишите, как вы получили ответ.
4. Сравните частотность притяжательных местоимений различных лиц и чисел в русских поэтических и прозаических текстах по НКРЯ.
5. Используя НКРЯ, сравните свойства русских конструкций *только и знает / делает / умеет что*. В какой форме в каждой из них чаще употребляется смысловой глагол: в инфинитиве или в форме, дублирующей форму вспомогательного глагола? В каком времени чаще используется каждая из этих конструкций: в прошедшем или в настоящем?

Образцы самостоятельных работ (ОПК 6.2, 7.1, 7.2)

1. Какой из перечисленных корпусов русского языка не имеет морфологической разметки?
(А) Уппсальский корпус; (Б) НКРЯ; (В) ХАНКО; (Г) ruTenTen

2. Если мы ищем формы глагола «стоять» и находим фразу «Этот дом стоит миллион евро», это — ...
(A) True Positive; (B) True Negative; (B) False Positive; (Г) False Negative
3. Сколько раз в корпусе встречаются *dis legomena*?
(A) 1; (B) 2; (B) 3; (Г) 4
4. В каком из этих тэгсетов число возможных тэгов наибольшее?
(A) CLAWS7; (B) Penn Treebank Tagset; (B) Brown Corpus Tagset; (Г) MULTEXT-East (Russian)
5. Если вычислить количество вхождений каждого из типов слов в корпус C, медиана этого значения скорее всего будет составлять ...
(A) 2; (B) 3; (B) 10; (Г) 100
6. Если изобразить на координатной плоскости с обычными (линейными) шкалами график, в котором по оси x отложен ранг слова, а по оси y — его частота, согласно закону Ципфа мы получим ...
(A) окружность; (B) ветвь гиперболы; (B) параболу; (Г) прямую
7. Средняя частотность отдельного слова (mean word frequency, MWF) обычно ... по мере роста корпуса
(A) увеличивается; (B) уменьшается; (B) остаётся неизменным; (Г) невозможно сказать
8. Первым корпусом, в названии которого употреблено слово «национальный», стал ...
(A) Болгарский национальный корпус; (B) Чешский национальный корпус; (B) Британский национальный корпус; (Г) Национальный корпус русского языка
9. Оцените по основному подкорпусу НКРЯ вероятности перехода между тэгами:
 $P(A|S)$, $P(S|S)$ и $P(V|S)$
10. Какая вероятность перехода больше: $P(S|взмахнул)$ или $P(взмахнул|S)$? Подтвердите свой ответ с помощью НКРЯ.
11. Рассчитайте индекс Герфиндаля—Гиршмана для романа Чарльза Диккенса «Great Expectations» (в оригинале, без лемматизации)
12. Оцените с помощью *Araneum Russicum Minus*, в каком числе доля творительного падежа от общего количества форм существительных выше: в единственном или во множественном? Опишите сделанные запросы.
13. Сколько типов и сколько токенов насчитывается в следующем мини-корпусе? Кратко опишите возможные проблемы при подсчёте.
We did send your invitation on, but she's travelling so she may not have got it. She sent us a rather vague address in Ibiza, but we haven't heard from her since she was in Paris.

14. Найдите моду и медиану рангово-частотного профиля, а также TTR для романа Джейн Остин «Pride and Prejudice» (в оригинале, без лемматизации).

Образцы проверочных (контрольных) заданий (ОПК 6.2, 7.1, 7.2)

1. Установите, какое ударение чаще используется в прилагательном *допризывной* / *допризывный*. Какой корпус лучше подходит для такого исследования: НКРЯ или ruTenTen — и почему?
2. Постройте в Excel график распределения частотностей для 1000 наиболее частотных слов по одному из подкорпусов, представленных в словаре [Ляшевская, Шаров 2009], откладывая по оси *x* ранг слова, а по оси *y* — его частотность. Аппроксимируйте полученное распределение при помощи степенной функции и оцените, насколько хорошо оно описывается при помощи закона Ципфа.
3. Сравните частотность сочетаний *A and B* и *B and A* для всех пар английских цветообозначений из множества {*red, black, green, blue, white, pink*}, используя любой достаточно большой корпус английского языка (BNC, COCA и т. п.). Какие цветообозначения тяготеют к первому месту в словосочетании, а какие — ко второму и с какими фонетическими особенностями это может быть связано?
4. Используя подкорпус CHES в Birkbeck Spelling Error Corpus, оцените, в каких словах 10-летние англоязычные дети чаще всего допускают орфографические ошибки. Попробуйте обобщить полученные результаты.
5. Используя COCA, укажите, какие существительные чаще всего сочетаются с прилагательными *independent, free* и *autonomous*. Попробуйте обобщить различия в сочетаемости этих прилагательных.
6. Используя корпус Google Books через интерфейс Brigham Young University, сравните частотность и сочетаемость слов со значением приблизительности (*almost, approximately, nearly* и т. д.) в разные десятилетия XIX и XX веков.

Критерии оценивания самостоятельных и контрольных работ

- результат, содержащий полный правильный ответ, полностью соответствующий требованиям критерия – 85 – 100 %;
- результат, содержащий неполный правильный ответ (степень полноты ответа – более 75%) или ответ, содержащий незначительные неточности, т.е. ответ, имеющий незначительные отступления от требований критерия, – 75 – 84% от максимального количества баллов;
- результат, содержащий неполный правильный ответ (степень полноты ответа – до 75%) или ответ, содержащий незначительные неточности, т.е. ответ, имеющий незначительные отступления от требований критерия – 60 -74 % от максимального количества баллов;
- результат, содержащий неполный правильный ответ, содержащий значительные неточности, ошибки (степень полноты ответа – менее 60%) – до 60 % от максимального количества баллов;
- неправильный ответ (ответ не по существу задания) или отсутствие ответа, т.е. ответ, не соответствующий полностью требованиям критерия, – 0 % от максимального количества баллов.

Тематика рефератов по дисциплине (ОПК 6.2, 7.1, 7.2)

Студентам предлагаются рефераты по статьям из сборника Biber, Douglas & Randi Reppen (eds.). 2011. *Corpus linguistics*. 4 vols. London: Sage.

Вопросы для промежуточной аттестации (ОПК 6.2, 7.1, 7.2)

1. Основные методы лингвистического исследования: интроспекция, эксперимент и наблюдение над реальностью. Место корпусной лингвистики в этом противопоставлении.
2. Лингвистические корпуса: определение и примеры применения в лингвистических исследованиях.
3. Корпуса русского языка (обзор):
 1. Национальный корпус русского языка (НКРЯ)
 2. ruWac
 3. ruTenTen
 4. Хельсинкский аннотированный корпус (ХАНКО)
 5. Интегрум
 6. Открытый корпус (OpenCorpora)
 7. Генеральный Интернет-корпус русского языка (ГИКРЯ)
 - ...
4. Корпуса английского языка (обзор):
 1. British National Corpus (BNC)
 2. Corpus of Contemporary American English (COCA)
 3. Corpus of Global Web-Based English (GloWbe)
 4. Brown Corpus
 5. Google Books: Google Ngrams Viewer и поисковый интерфейс на сайте Brigham Young University.0
 - ...
5. Типы разметки в корпусах:
 1. Морфологическая разметка
 2. Синтаксическая разметка
 3. Прочие виды лингвистической разметки
 4. Метаразметка.
6. Стандарты морфологической разметки для русского и английского языка. Омонимия и её разрешение.
7. Количественные исследования на корпусном материале. Базовые методы статистики в корпусных исследованиях.
8. Нормирование частотности языковых единиц в корпусах различного объёма.
9. Частотные словари. Закон Ципфа.
10. Исследование сочетаемости слов при помощи корпусов. Коллокации и меры их оценки. Лексические функции и их корпусное исследование.
11. Дифференциальные исследования на корпусном материале и приспособленность различных корпусов русского и английского языка для их проведения.
12. Проблема отбора текстов в корпус, достижения репрезентативности и сбалансированности корпуса.
13. Многоязычные корпуса и их использование в лексикографии и в преподавании иностранных языков.
14. Интернет как корпус. Поисковые системы как заменитель корпусов («Googleology»), Яндекс.Блоги.
15. Создание пользовательских корпусов:
 1. Корпусные менеджеры: WordSmith, AntConc и т. д.
 2. Создание пользовательских корпусов в системе SketchEngine и возможности их исследования.
 3. Создание мультимодальных корпусов в программе ELAN.
16. Применение корпусных методов в различных областях лингвистики:
 1. Грамматика
 2. Лексикография

3. Социолингвистика и др.

Критерии оценивания для промежуточной аттестации обучающихся (вопросы к зачету)

- результат, содержащий полный правильный ответ, полностью соответствующий требованиям критерия – 85 – 100 %;
- результат, содержащий неполный правильный ответ (степень полноты ответа – более 75%) или ответ, содержащий незначительные неточности, т.е. ответ, имеющий незначительные отступления от требований критерия, – 75 – 84% от максимального количества баллов;
- результат, содержащий неполный правильный ответ (степень полноты ответа – до 75%) или ответ, содержащий незначительные неточности, т.е. ответ, имеющий незначительные отступления от требований критерия – 60 -74 % от максимального количества баллов;
- результат, содержащий неполный правильный ответ, содержащий значительные неточности, ошибки (степень полноты ответа – менее 60%) – до 60 % от максимального количества баллов;
- неправильный ответ (ответ не по существу задания) или отсутствие ответа, т.е. ответ, не соответствующий полностью требованиям критерия, – 0 % от максимального количества баллов.

6. Учебно-методическое и информационное обеспечение дисциплины

5.4 Список источников и литературы

Литература

Основная

1. Корпусные исследования по русской грамматике : сб. ст. / Рос. акад. наук, Ин-т языкознания ; [ред.-сост. К. Л. Киселева и др.]. - М. : Пробел-2000, 2009. - 513 с.; 22 см. - Библиогр. в конце ст. - ISBN 978-5-98604-148-3 : 330.00. Ссылка на ресурс: <http://text.lib.rsuh.ru/macro/256.txt>
2. Плунгян Владимир Александрович. Введение в грамматическую семантику: грамматические значения и грамматические системы языков мира : учеб. пособие / В. А. Плунгян ; [М-во образования и науки Рос. Федерации, Гос. образоват. учреждение высш. проф. образования "Рос. гос. гуманитарный ун-т"]. - М. : РГГУ, 2011. - 669 с.; 22 см. - Библиогр.: с. 498-581. - Указ.: с. 582-664. - ISBN 978-5-7281-1122-1 : 435.00.

5.5 Перечень ресурсов информационно-телекоммуникационной сети «Интернет».

№п/п	Наименование	Условия доступа/скачивания
1	British National Corpus	http://www.natcorp.ox.ac.uk/
2	Corpus of Contemporary American	http://corpus.byu.edu/coca/
3	Национальный корпус русского языка	www.ruscorpora.ru

5.6 Профессиональные базы данных и информационно-справочные системы

Доступ к профессиональным базам данных: <https://liber.rsuh.ru/ru/bases>

Информационные справочные системы:

1. Национальный корпус русского языка
2. British National Corpus
3. Corpus of Contemporary American English
4. Sketch Engine
5. Aranea

6. Материально-техническое обеспечение дисциплины

Занятия по курсу можно проводить с максимальной эффективностью в компьютерном классе или аудитории с доступом в Интернет, проектором и экраном для презентаций, CD-проигрыватель, DVD-проигрыватель. Необходимо также наличие доски или флипчарта, чтобы преподаватель мог разбирать примеры по ходу объяснения и записывать задания.

Операционная система: Microsoft Windows 2000, Microsoft Windows XP, Microsoft Windows Vista;

- Не менее 256 МБ оперативной памяти, рекомендуемый объём - 512 МБ;
- Видеокарта и монитор с разрешением не менее 1024x768 точек.

Состав программного обеспечения (ПО), современных профессиональных баз данных (БД) и информационно-справочных систем (ИСС)

Перечень ПО

1. Windows
2. Microsoft Office
3. Kaspersky Endpoint Security

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

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.

8. Методические материалы

8.1 Планы семинарских/ практических/ лабораторных занятий

Основные темы курса:

1. Введение. Общее представление о корпусах и корпусной лингвистике. Стандарты разметки. Типы разметки корпусов.
2. Особенности различных типов разметки. Морфологическая разметка
3. Особенности различных типов разметки. Синтаксическая разметка
4. Особенности разметки: другие типы разметки
5. Методы корпусных исследований. Анализ примеров корпусных исследований

План семинарских занятий и самостоятельной работы студентов

В соответствии с учебным планом предусмотрены семинарские занятия. Некоторые из них строго обязательны, а другие допускают рассмотрение той или иной темы с разной степенью подробности: разворачивание и уточнение темы или, напротив, объединение нескольких тем.

Часть 1

№ занятия	Тема семинара	Вопросы для подготовки к семинару и самостоятельной работы
1	Введение. Общее представление о корпусах и корпусной лингвистике.	1. Стандарты разметки. 2. Типы разметки корпусов.
2	Проблемные корпуса (параллельные, диалектные, мультимедийные и др.)	1. Параллельные корпуса. 2. Диалектные корпуса. 3. Мультимедийные корпуса.
3	Поиск в корпусе. Использование языка SQL для поиска в корпусе. Составление сложных запросов к корпусу.	1. Визуальные интерфейсы корпусов. 2. Языки запросов к корпусу.
4-6	Особенности различных типов разметки.	1. Особенности различных типов разметки. Морфологическая разметка. 2. Особенности различных типов разметки. Синтаксическая разметка. 3. Особенности разметки: другие типы разметки. 4. Методы корпусных исследований. Анализ примеров корпусных исследований.
7	Промежуточная аттестация: Контрольная работа по теме: «Корпуса и лингвистические ресурсы»	Обобщение пройденного материала
8	Инструменты разметки собственного исследовательского корпуса	1. Система SketchEngine.. 2. Программа WebBootCaT.
9-10	Составление конкордансов, частотных списков, списков коллокаций с использованием специальных программ	1. Анализ корпусов с помощью AntConc. 2. Система SketchEngine для анализа корпусов.
	Итого 20ч.	

Часть 2.

Интернет и развитие гуманитарного знания: понятие digital humanities, анализ big data	1 часа
Понятие «корпус»: лингвистические корпуса в доцифровую эпоху, конкордансы, репрезентативность корпуса, сбалансированность корпуса, аннотирование корпуса	1 часа
Классификация корпусов: одноязычные/многоязычные/параллельные; письменные/устные; синхронические/диахронические; общие/ специфические; открытые/закрытые.	2 часа
Типы лингвистического аннотирования; аннотация автоматическая, полуавтоматическая, ручная;	2 часа

лингвистическая и экстралингвистическая разметка.	
Морфологическая аннотация. Проблемы морфологического аннотирования, снятие омонимии.	2 часа
Синтаксическая аннотация. Проблемы синтаксического аннотирования. Понятие синтаксического дерева, синтаксического отношения, лексической функции	2 часа
Другие виды разметок: семантическая, мультимодальная, просодическая и др.	2 часа
Многоязычные корпуса и корпуса второго языка	1 часа
Интернет как корпус и корпус «Сделай сам!»	2 часа
Типы исследований на базе корпусов: статистические методы; corpus-based, corpus-informed, corpus-driven research	1 часа
Количественные методы: базовая статистика (среднее, мода, медиана)	1 часа
Количественные методы: статистика распределений (критерий Стьюдента, стандартное отклонение и др.), методы извлечения коллокаций	1 часа
Презентация проекта	2 часа
Итого 20 ч.	

9.2. Методические рекомендации по подготовке письменных работ

Самостоятельная работа студентов предполагает:

1. подготовку письменных и устных домашних заданий;
2. внеаудиторную работу студентов (самостоятельное освоение теоретического материала, подготовка домашних заданий).

Самостоятельная работа студента играет большую роль в освоении материала, поскольку она делает восприятие информации не пассивным, а активным процессом. Здесь важны и самостоятельный поиск материала в научной литературе и на соответствующих сайтах, и его конспектирование, и при необходимости его трансформация в схемы и алгоритмы.

Общие принципы самостоятельной работы. За редким исключением, СРС нацелена не на запоминание материала, а на его понимание, осмысление, упорядочение и активное практическое владение им. Критерием адекватного понимания является способность объяснить материал своими словами непрофессионалу, умение приводить иллюстративные примеры и практически разрабатывать электронные обучающие материалы, используя предлагаемые инструменты. Необходимо уметь формулировать вопросы и находить ответы на них самостоятельно, в ходе консультации с преподавателем и другими членами группы. Только после этого нужно приступить к выполнению задания.

9.3. Иные материалы

Рекомендуемая дополнительная литература:

- Копотев М. Введение в корпусную лингвистику. Praha: Animedia.
- E. T. Meyer, R. Shroeder *The Oxford Handbook of Internet Studies*
- T. McEnergy and A. Hardie *The Oxford Handbook of the History of Linguistics*

Дополнительная:

Груздева, Елена В. 2012. *Корпусная лингвистика*. 2-е изд. М.: ФЛИНТА..

Киселёва, Ксения Л., Владимир А. Плунгян, Екатерина В. Рахилина, Сергей Г. Татевосов (ред.). *Корпусные исследования по русской грамматике*. М: Пробел–2009.

Национальный корпус русского языка: 2003—2005. Сборник статей. М.: Индрик, 2005.

Ляшевская, Ольга Н. & Шаров, Сергей А. *Частотный словарь современного русского языка (на материалах Национального корпуса русского языка)*. М.: Азбуковник. 2009.

Плунгян, Владимир А., Екатерина В. Рахилина, Татьяна И. Резникова (ред.). *Национальный корпус русского языка: 2006—2008. Новые результаты и перспективы*. СПб.: Нестор-История, 2009.

- Плунгян, Владимир А., Шестакова Лариса Л. (ред.). 2013. *Корпусный анализ русского стиха : Сборник научных статей*. М.: Издательский центр «Азбуковник».
- Aijmer, Karin & Christoph Rühlemann. 2014. *Corpus pragmatics: A handbook*. Cambridge: Cambridge University Press.
- Baker, Paul. 2010. *Sociolinguistics and corpus linguistics*. Edinburgh: Edinburgh University Press.
- Biber, Douglas & Randi Reppen (eds.). 2011. *Corpus linguistics*. 4 vols. London: Sage.
- Cheng, Winnie. 2012. *Exploring corpus linguistics: Language in action*. London & New York: Routledge.
- Gatto, Maristella. 2014. *The Web as corpus: Theory and practice*. New York : Bloomsbury Academic.
- Gries, Stefan Th. 2009. *Quantitative corpus linguistics with R: A practical introduction*. London & New York: Routledge.
- Lüdeling, Anke & Merja Kytö. 2008–2009. *Corpus linguistics: An international handbook*. 2 vols. HSK 29.1–2. Berlin & New York: Walter de Gruyter.
- Meyer, Charles F. 2002. *English corpus linguistics: An introduction*. Cambridge & New York: Cambridge University Press.
- McEnery, Tony & Andrew Hardie. 2012. *Corpus linguistics: Method, theory and practice*. Cambridge & New York: Cambridge University Press.
- O’Keefe, Anne & Michael McCarthy (eds.). 2010. *The Routledge handbook of corpus linguistics*. London & New York: Routledge.
- Sinclair, John. 1991. *Corpus, concordance, collocation*. Oxford: Oxford University Press.

Терминологический словарь-минимум:

Baker, Paul, Andrew Hardie & Tony McEnery. 2006. *A glossary of corpus linguistics*. Edinburgh: Edinburgh University Press.

Рекомендуемое программное обеспечение

п/п	Наименование	Условия доступа/скачивания
1.	AntConc	Свободное лицензионное соглашение
2.	SketchEngine	Платная подписка

АННОТАЦИЯ РАБОЧЕЙ ПРОГРАММЫ ДИСЦИПЛИНЫ

Дисциплина «**Информационные технологии и корпусные исследования в лингвистике**» реализуется совместно УНЦ компьютерной лингвистики и кафедрой теоретической и прикладной лингвистики факультета теоретической и прикладной лингвистики Института лингвистики РГГУ.

Цель дисциплины – познакомить магистрантов с наиболее актуальными современными компьютерными корпусами текстов и лексикографическими ресурсами, программами обработки текста, с технологиями создания собственных исследовательских корпусов, научить применять методы создания собственных исследовательских корпусов, работы с корпусными данными, методы обработки этих данных в собственных научных исследованиях, а также познакомить с современными исследованиями и особенностями языка, выявленными на базе корпусных исследований.

Образовательными задачами дисциплины являются:

(а) ознакомление студентов с ключевыми аспектами современной корпусной лингвистики, а именно, с основными научными направлениями и с русскоязычной и англоязычной терминологией;

(б) изучение устройства корпусов разных языков;

(в) ознакомление с типами исследований, проводящихся на базе корпусов;

(г) изучение основ корпусной педагогики;

(д) конечная перспектива дисциплины — познакомить студентов с корпусами различных языков, научить их пользоваться корпусными ресурсами, показать, каким образом лингвисты и педагоги работают с корпусами, сформировать у студентов базовые навыки корпусной разметки.

Практические задачи дисциплины:

(а) ознакомление студентов с ключевыми аспектами современной корпусной лингвистики, а именно, с основными принципами аннотирования и методами ведения исследования;

(б) изучение на материале конкретных корпусов типов междисциплинарных корпусных исследований;

(в) ознакомление с интересными научно-исследовательскими задачами в каждой из рассмотренных областей;

(г) ознакомление с новыми возможностями в исследовании грамматики и лексики языка, которые дают использование корпусных методов, а также с применением современных методов обработки этих данных;

(д) ознакомление с технологиями и проблемами разметки корпусов;

(е) обучение практическим навыкам по применению корпусных методов в своей исследовательской работе.

В результате освоения дисциплины студент должен

Знать:

- основные принципы создания корпусов и других компьютерных лингвистических ресурсов;
- характеристики и особенности современных доступных в Интернете национальных и проблемных корпусов, широко используемых в лингвистических исследованиях, включая недавно вошедшие в лингвистическую практику;
- стандарты, типы и проблемы разметки корпусов, включая такие современные типы разметки, как дискурсивную разметку, интонационную разметку устных корпусов и т.п., применяемые в разметке технологии;
- принципы создания собственных исследовательских корпусов;
- основные типы исследовательских задач, решаемых с использованием корпусов;
- основные применяемые в корпусных исследованиях лексики и грамматики методы;
- требования, предъявляемые к верификации результатов;
- основные методы статистического анализа корпусных данных.

Уметь:

- применять полученные знания в области корпусной лингвистики в научно-исследовательской и других видах практической деятельности;
- осуществлять мониторинг и оценку различных типов современных корпусных ресурсов и выбирать ресурсы, подходящие для выполнения тех или иных исследовательских и производственных задач;
- осуществлять поиск в корпусах в соответствии с исследовательской гипотезой в области грамматики и лексикографических исследований;
- создавать и размечать собственные исследовательские и обучающие корпуса;
- работать с различными типами программ обработки текстов: конкордансерами, программами для поиска коллокаций, создания частотных списков и т.п., корпусными менеджерами;
- разрабатывать методический материал по основным языковым дисциплинам с использованием корпусов.

Владеть:

- основными методами и средствами профессионального компьютерного инструментария для исследовательской и практической работы;
- методами сбора материала с использованием корпусов;
- методами анализа корпусных данных, включая статистические методы.

Программой дисциплины предусмотрена промежуточная аттестация в форме экзамена. Общая трудоемкость освоения дисциплины составляет 4 зачетные единицы.