Федеральное государственное бюджетное
образовательное учреждение высшего образования
«Государственный институт русского языка им. А.С. Пушкина»

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)

Наименование дисциплины (модуля):

Корпусная лингвистика

Б1.В.ДВ.05.02

Уровень высшего образования:

магистратура

Направление подготовки:

45.04.01 Филология

Направленность (профиль) образования:

Прикладная филология

Форма обучения:

очная

Рабочая программа дисциплины (модуля) разработана в соответствии с федеральным государственным образовательным стандартом высшего образования по направлению подготовки, утвержденным приказом Минобрнауки России № 980 от 12.08.2020.

1. Место дисциплины (модуля) в структуре ОПОП:

В соответствии с учебным планом магистратуры 45.04.01 "Филология" по специальности "Прикладная филология" дисциплина Б1.В.ДВ.05.02 «Корпусная лингвистика» относится к части, формируемой участниками образовательных отношений, и изучается во 2 семестре.

2. Цель и задачи освоения дисциплины (модуля)

Целью изучения курса «Корпусная лингвистика» развитие теоретических знаний студентов в области прикладной лингвистики и формировании навыков применения специализированных электронных корпусов текстов для решения научно-исследовательских задач. Эта глобальная цель конкретизируется в следующих задачах: формирование представлений о концепциях корпусной лингвистики, её месте в ряду лингвистических дисциплин и значении в рамках доказательной (эмпирической) лингвистики; изучение синтаксиса специализированных языков поисковых запросов, приемов алгоритмизации лингвистических задач и овладение навыками работы с корпусами русского языка; развитие навыков сбора, машинной обработки текстов, создания конкордансов и частотных словарей, а также применение корпусных данных в лексикографии и исследовании различных регистров языка. В результате магистранты должны овладеть новейшими технологиями получения языковых данных при учёте правильного формирования пользовательского запроса, что необходимо для их будущей научной и педагогической деятельности.

3. Планируемые результаты обучения по дисциплине (модулю), соотнесенные с требуемыми компетенциями выпускников

Компетенции выпускников (коды)	**Индикаторы достижения компетенций**	**Планируемые результаты обучения по дисциплине (модулю), соотнесенные с индикаторами достижения компетенций**
ПК-1	ПК-1.1 Выделяет и систематизирует значимые аспекты	Знать: основные понятия корпусной

Способен проводить разноаспектное описание лексических единиц для решения теоретических и прикладных задач	описания лексических единиц в соответствии с целями исследования и типом материала	лингвистики: лексическая единица (лемма, вокабула, словоформа), коллокация, валентность, семантические роли, прагматические маркеры, коннотации. Уметь: формулировать цели описания единиц применительно к конкретному языковому корпусу и типу исследования. Владеть: методами количественного и качественного анализа лексики, извлеченной из корпуса
	ПК-1.2. Применяет цифровые и корпусные технологии для работы с лексическими данными	Знать: принципы разметки (аннотирования) лексического уровня в мультимедийных корпусах, использование тегов для морфологических (*POS*-теггинг), семантических и просодических характеристик. Уметь: анализировать схемы разметки в различных подкорпусах. Владеть: Применять специализированное программное обеспечение для извлечения информации из языкового корпуса Знать: Методы извлечения и

	ПК-1.3. Представляет результаты разноаспектного описания лексических единиц в формате, пригодном для разных пользователей	представления языковой информации: конкорданс, частотные словари, тезаурусы, n-граммы. Уметь: проводить семантическую, прагматическую и мультимодальную разметку текстовых фрагментов, выделяя значимые аспекты. Владеть: Навыками проектирования и ведения глоссариев, тегсетов и руководств по разметке для коллективной работы над корпусами.

4. Объем дисциплины (модуля) _2_з.е., в том числе __32__академических часа на контактную работу обучающихся с преподавателем; ___38__академических часов на самостоятельную работу обучающихся.

(В конце данного пункта следует отметить, если дисциплина или часть ее реализуется с применением электронного обучения и дистанционных образовательных технологий)

5. Содержание дисциплины (модуля), структурированное по темам (разделам) с указанием отведенного на них количества академических часов

Наименование и краткое содержание разделов и тем дисциплины (модуля). **Форма промежуточной аттестации по дисциплине (модулю)**	**Всего (часы)**	В том числе							
		Контактная работа (работа во взаимодействии с преподавателем) ***Виды контактной работы, часы***					**Самостоятельная работа обучающегося** ***Виды самостоятельной работы, часы***		
		Занятия лекционного типа	Занятия семинарского типа (практические/ лабораторные)	Групповые консультации	Индивидуальная работа с обучающимися	**Всего**			**Всего**
Тема 1. Корпусная лингвистика как научная дисциплина. Корпусы специальных текстов. Корпусная лингвистика и ее понятийный аппарат. История корпусной лингвистики. Виды языковых корпусов. Варьирование объема и баланса		2	2			4		4	8

корпуса.									
Тема 2. Создание специальных корпусов текстов. Разметка корпусов, стандартизация в корпусной лингвистике. Настройка пользовательских подкорпусов в НКРЯ для исследовательских целей: газетный, акцентологический, устный, диалектный, мультимедийный, параллельный, поэтический подкорпуса. Панхронический корпус.		2	2			4		4	8
Тема 3. Поиск по грамматической и метатекстовой		2	2			4		4	8

разметке в НКРЯ. Функции "График", "Статистика", "Частотность" в НКРЯ. Специальные корпуса текстов на русском языке.									
Тема 4. Корпусный анализ синтаксической вариативности и грамматикализации. Статистическое исследование конкурирующих конструкций (дательный самостоятельный vs. придаточное, дативы, генитивные конструкции отрицания).		2	2			4		4	8
Тема 5. Корпусные методы в обучении иностранным языкам и лексикографии.		2	2			4		4	8

Разработка DDL (Data-Driven Learning) модулей на базе корпусов. Создание учебных словарей на основе частотности и коллокаций, анализ межъязыковой интерференции через сопоставительные корпуса ошибок.									
Тема 6. Автоматическая разметка семантических ролей и дискурсивных маркеров. Оценка качества разметки и исправление ошибок автоматических парсеров.		2	2			4		6	10
Тема 7. Корпусные методы в		2	2			4		6	10

исторической лексикографии и этимологии. Сравнительный анализ диахронических корпусов (например, НКРЯ с историческим подкорпусом, Corpus of Historical American English). Реконструкция семантических сдвигов, датировка неологизмов, верификация этимологических гипотез на основе больших данных.									
Тема 8. Программа AntConc: функция REGEX: вариативные запросы поиска. Сервис NoSketchEngine.		2	2			4		6	10

Промежуточная аттестация *(указывается форма проведения)*	*зачет*					2
Итого		16	16	32	38	72

6. Фонд оценочных материалов по дисциплине (модулю)

Практические задания для текущего контроля

Программа курса «Корпусная лингвистика» для магистрантов предполагает баланс между теоретическим пониманием методов и практическими навыками работы с данными. Задания разделены на четыре тематических блока, от базовых операций до самостоятельного исследовательского проектирования.

Блок 1. Основы работы с корпусами и разметкой

1. **Сравнительный анализ интерфейсов.** Сравнить поисковые возможности и структуру метаданных двух корпусов (например, НКРЯ и Aranea для русского языка). Оценить, какие типы лингвистических и экстралингвистических запросов доступны в каждом из них.
2. **Оценка качества разметки.** Выбрать в корпусе фрагмент текста (100–150 токенов) со сложной синтаксической структурой. Сопоставить автоматическую разметку (морфологическую и синтаксическую) с ручным анализом, зафиксировать типы систематических ошибок парсера.
3. **Создание мини-корпуса.** Собрать из открытых источников (например, с сайта «Проза.ру» или из новостных архивов) 50 текстов по заданному критерию (жанр, год, автор). Очистить их от шума, привести к единому формату (TXT) и составить файл метаданных в формате TSV.
4. **Метаданные и репрезентативность.** Проанализировать состав подкорпуса устной речи НКРЯ. Определить, какие социальные и демографические группы в нем недопредставлены, и обосновать, как это искажает результаты исследования разговорных маркеров.
5. **Нормализация и лемматизация.** Написать скрипт (на Python) или использовать инструменты командной строки для лемматизации и очистки небольшого текстового файла от знаков препинания и стоп-слов. Сравнить результат работы двух разных лемматизаторов (например, PyMorphy2 и UDPipe) на омонимичных формах.

Блок 2. Частотный анализ и коллокации

6. **Ключевые слова (Keywords).** Используя Sketch Engine или AntConc, составить список ключевых слов (keywords) для корпуса текстов одного автора по сравнению с эталонным референтным корпусом. Проинтерпретировать топ-10 слов с точки зрения идиостиля.
7. **Динамика частотности.** Построить график частотности трех общественно-политических лексем (например, *суверенитет*, *цифровизация*, *токсичный*) в основном корпусе НКРЯ по десятилетиям с 1950 по 2020 год. Объяснить пики и спады частотности историческим контекстом.
8. **Анализ коллокаций.** Выявить коллокации для многозначного слова (например, *камень* или *ключ*) в основном корпусе. Разделить их на смысловые группы, соответствующие разным лексико-семантическим вариантам слова.
9. **Статистическая значимость.** Рассчитать для выбранной коллокации (например, *круглый дурак*) статистические показатели ассоциации (MI, LogDice, T-score). Объяснить, почему для разных типов лингвистических задач (поиск устойчивых выражений vs. анализ словосочетаний) предпочтительны разные меры.
10. **Семантические примитивы.** Составить частотный список глаголов движения в корпусе русской прозы XIX века. Сопоставить его со списком из корпуса текстов XXI века и описать произошедшие семантические сдвиги.

Блок 3. Синтаксис, семантика и прагматика

11. **Синтаксические конструкции.** В синтаксически размеченном корпусе (например, SynTagRus) найти все реализации дативного субъекта. Описать их частотность, лексическое наполнение и типичные предикаты.
12. **Диахронический сдвиг.** Отследить изменение управления глагола (например, *управлять* + творительный падеж vs. *управлять* + винительный падеж) на материале корпуса за последние 100 лет.
13. **Прагматические маркеры.** Составить список контекстных фильтров для поиска дискурсивных маркеров (*ну*, *типа*, *как бы*) в устном подкорпусе. Оценить их распределение в речи разных возрастных групп.
14. **Аспектуальная семантика.** Проанализировать конкуренцию видовременных форм глагола в контексте обстоятельств типа *вдруг*, *сразу*, *мгновенно*. Построить распределение частот совершенного и несовершенного вида в этих позициях.
15. **Корпусный анализ метафор.** Найти контексты со словом *яд* в прямом и метафорическом значениях. Описать семантические поля, в которые встраивается слово при метафоризации, используя инструмент Word Sketch.

Блок 4. Прикладные задачи и исследовательское проектирование

16. **Авторизация текста.** Разработать и протестировать на мини-корпусе (10 авторов, по 5 текстов на каждого) алгоритм атрибуции текста по 5–7 дистинктивным признакам (длина предложения, коэффициент лексического разнообразия TTR, частотность служебных слов).
17. **Сентимент-анализ.** Провести лексикографическую разметку полярности (позитив/негатив/нейтрально) для 200 слов, характерных для жанра кинорецензий. Оценить точность автоматической классификации тональности на тестовой выборке корпуса.
18. **Лингвистическая экспертиза.** Сформулировать корпусную стратегию для решения прикладной задачи (например, доказательство плагиата или установление авторства анонимного письма). Описать, какие подкорпуса и статистические критерии будут использованы.
19. **Проектирование исследовательского корпуса.** Написать развернутый паспорт исследовательского корпуса (2–3 тысячи словоупотреблений) для курсовой работы. Обосновать выбор текстов, принципы разметки (аннотации речевых актов, эмоций или синтаксиса), структуру XML/TEI и репрезентативность выборки.
20. **Визуализация данных.** На основе данных, полученных в задании №7 (динамика частотности), построить интерактивную визуализацию (например, в Python с использованием Plotly или библиотеки D3.js), добавив на временную шкалу маркеры ключевых исторических событий, повлиявших на употребление лексем.

Рекомендации по формату сдачи:

- Теоретические и аналитические задания (1–5, 11–15, 18, 19) оформляются в виде отчета объемом 2–4 страницы с приложением списков примеров.
- Практические задания (6–10, 16, 17, 20) сдаются в виде Jupyter-ноутбука или R-скрипта с комментариями к коду, итоговыми таблицами и графиками, а также текстовым выводом на 1 страницу.

Критерии оценки практических заданий

Общие принципы начисления баллов (для всех заданий)

- **Своевременность сдачи:** 0 баллов штрафа при сдаче в срок; -10 баллов при задержке до 3 дней; -20 баллов при задержке более 3 дней.
- **Академическая добросовестность:** обнаружение плагиата или автоматической генерации текста без авторской переработки — 0 баллов за задание.

- **Формат сдачи:** несоблюдение требований к оформлению (отсутствие комментариев в коде, неверный формат метаданных) — штраф от -5 до -15 баллов.

Блок 1. Основы работы с корпусами и разметкой (Задания 1–5) *Базовый уровень (0–60 баллов):* задание выполнено, файлы или скриншоты интерфейсов приложены, даны поверхностные описания. *Продвинутый уровень (61–85 баллов):* проведен корректный анализ, выявлены 3–5 ключевых различий или типов ошибок, приведены точные количественные примеры (5–10 контекстов). *Экспертный уровень (86–100 баллов):* глубокая аналитика с опорой на теорию корпусной лингвистики (репрезентативность, баланс), предложены способы устранения выявленных проблем разметки, для задания №3 составлен валидный файл метаданных в формате TSV или JSON, полностью готовый к загрузке в корпус-менеджер.

Блок 2. Частотный анализ и коллокации (Задания 6–10) *Базовый уровень (0–60 баллов):* списки слов или графиков извлечены, но в отчете отсутствует статистическая интерпретация и проверка значимости. *Продвинутый уровень (61–85 баллов):* приведены верные статистические показатели (MI, LogDice), выборка релевантна, дано лингвистически обоснованное объяснение результатов, графики наглядны и снабжены осями. *Экспертный уровень (86–100 баллов):* проведен сравнительный анализ разных статистических мер (например, почему для задания №9 выбрана именно мера LogDice), выявлены скрытые факторы (например, влияние омонимии на коллокации), для задания №8 предложена собственная семантическая классификация, подтвержденная количественными данными.

Блок 3. Синтаксис, семантика и прагматика (Задания 11–15) *Базовый уровень (0–60 баллов):* найдены единичные примеры конструкций, фильтрация контекстов неполная, выводы носят характер общих рассуждений. *Продвинутый уровень (61–85 баллов):* выборка составляет не менее 50 релевантных примеров, фильтры для поиска (CQL или интерфейсные) описаны точно, распределение данных представлено в виде таблиц. *Экспертный уровень (86–100 баллов):* проведен количественный анализ распределения (проценты, абсолютные частоты), выявлена статистически значимая корреляция (например, между возрастом говорящих и частотой маркеров в задании №13), для задания №12 построена диахроническая модель изменения конструкции.

Блок 4. Прикладные задачи и исследовательское проектирование (Задания 16–20) *Базовый уровень (0–60 баллов):* алгоритм или стратегия описаны только на вербальном уровне, код не предоставлен или не запускается, признаки для анализа выбраны произвольно. *Продвинутый уровень (61–85 баллов):* код рабочий и прокомментированный, проведена оценка точности (accuracy/precision/recall), для задания №16 точность классификации превышает 70%, паспорт корпуса в задании №19 содержит все обязательные разделы. *Экспертный уровень (86–100 баллов):* для заданий №16 и №17 проведена кросс-валидация, признаки для классификации статистически обоснованы; в задании №18 предложенная

корпусная стратегия полностью исключает ложноположительные результаты; в задании №19 паспорт корпуса соответствует стандарту TEI, выборка строго сбалансирована по жанрам и времени, а в задании №20 визуализация интерактивна и позволяет фильтровать данные по дополнительным параметрам.

Итоговая шкала перевода баллов в оценку за отдельное задание

- 90–100 баллов: Отлично
- 75–89 баллов: Хорошо
- 60–74 балла: Удовлетворительно
- 0–59 баллов: Неудовлетворительно (требуется пересдача)

Задания по курсу для самостоятельной работы студентов

1. **Критика архитектуры корпуса.** Выбрать любой национальный корпус (например, НКРЯ) и проанализировать его архитектуру. Оценить принципы балансировки, разметки и репрезентативности. Выявить три методологические лакуны (например, смещение в сторону письменного стандарта) и предложить способы их устранения в новой версии корпуса.
2. **Сравнительный анализ теггеров.** Взять один сложный для разметки фрагмент текста (около 500 слов, содержащий неологизмы, имена собственные и синтаксические эллипсисы). Разметить его вручную, затем пропустить через два разных автоматических теггера (например, TreeTagger и spaCy). Составить таблицу расхождений и рассчитать метрики точности (precision) и полноты (recall).
3. **Эволюция понятия «репрезентативность».** Написать аналитический обзор (2–3 страницы) того, как менялось понимание репрезентативности корпуса: от ранних частотных словарей и корпусов Brown/LOB до современных веб-корпусов (Araneum, TenTen) и корпусов, сбалансированных по данным социальных сетей.
4. **Этические границы.** Подготовить аналитическую записку об этических и юридических проблемах при включении в корпус данных из закрытых Telegram-каналов, личной переписки или медицинских форумов. Разработать проект информированного согласия для донора текстовых данных.

Практические задания (анализ данных)

5. **Семантические примитивы.** С помощью конкордансера (AntConc или встроенный в Sketch Engine) извлечь все контексты употребления лексем *«правда»* и *«истина»* в основном корпусе НКРЯ. Провести семантико-

синтаксический анализ: определить коллокаты, типичные актанты и дискурсивные функции каждого слова. Доказать или опровергнуть их синонимию.

6. **Диахронический сдвиг.** Используя исторический подкорпус НКРЯ, отследить изменение сочетаемости слова *«государь»* с XV по XIX век. Построить график частотности и проанализировать смену ключевых коллокатов (например, от *«государь-батюшка»* к *«государь-император»*).
7. **Корпусный аудит перевода.** Взять оригинальный текст (например, рассказ А.П. Чехова) и два его перевода на английский язык. С помощью выравнивания параллельных корпусов проанализировать, как переводятся русские видо-временные формы (например, совершенный вид прошедшего времени) и какие стратегии перевода (нейтрализация, экспликация) преобладают.
8. **Метаданные и прагматика.** Сформулировать и проверить гипотезу о связи возраста автора и частоты использования дискурсивных маркеров (*«типа»*, *«как бы»*, *«на самом деле»*) на материале НКРЯ. Оценить статистическую значимость различий (например, с помощью логит-регрессии или критерия хи-квадрат).
9. **Фразеологизмы в реальном употреблении.** Выбрать 5 фразеологизмов (например, *«бить баклуши»*, *«спустя рукава»*). Найти в НКРЯ 100 примеров их употребления. Классифицировать случаи по степени идиоматичности: от полностью идиоматичных до тех, где компоненты употреблены в прямом смысле, и определить синтаксические позиции, в которых фразеологизм теряет устойчивость.
10. **Корпусная лексикография.** Составить черновик словарной статьи для современного неологизма (например, *«зумер»*, *«кринж»* или *«донат»*) исключительно на основе данных из газетного подкорпуса НКРЯ за последние 3 года. Включить в статью дефиницию, грамматические пометы, типичные коллокаты и иллюстративные примеры.

Программирование и автоматическая обработка (NLP)

11. **Скрипт для извлечения именных групп.** Написать скрипт на Python (с использованием spaCy или NLTK), который автоматически извлекает из произвольного текста все именные группы (NP) со структурой «прилагательное + существительное». Оценить точность работы скрипта на корпусе новостей.
12. **Визуализация семантической сети.** Написать скрипт, который строит граф коллокаций для заданного узла (например, *«кризис»*) на основе данных корпуса. Визуализировать сеть с помощью NetworkX, выделив цветом сильные и слабые коллокаты, а размером узлов — их частотность.
13. **Создание миниатюрного корпуса.** Собрать собственный корпус объемом 100 тысяч слов (например, из постов конкретного Telegram-канала или ветки форума). Очистить данные от шума, провести базовую лемматизацию и разметить метаданные (дата, источник, пол автора, если он указан) в формате TEI или JSON.

14. **Регулярки для нестандартной разметки.** Написать набор регулярных выражений для извлечения из сырого текстового дампа (например, логов чата) метаданных: времени отправки, ID пользователя и статуса сообщения. Преобразовать результат в формат CSV для последующей загрузки в корпус-менеджер.
15. **Оценка качества OCR.** Взять сканированный текст дореволюционной книги, пропущенный через OCR. Написать скрипт, который находит типичные ошибки распознавания (например, путаница *«ять»* и *«е»*, *«і»* и *«н»*) и автоматически составляет отчет о качестве текста в процентах.

Проектные и исследовательские задания

16. **Проектирование специализированного корпуса.** Разработать детальную документацию (прототип) узкопрофильного корпуса (например, «Корпус судебной речи в России» или «Корпус экологического дискурса в блогосфере»). Описать критерии отбора текстов, балансировку по жанрам, схему аннотации (включая разметку судебных ролей или эко-лексики) и интерфейс поиска.
17. **Синтаксические параллели.** На материале Parallel Universal Dependencies (UD) исследовать, какими синтаксическими конструкциями (деепричастный оборот, придаточное предложение, номинализация) передается английское причастие *-ing* в русских переводах. Построить иерархию частотности трансформаций.
18. **Корпусное исследование неологии.** Выявить 20 неологизмов 2020–2024 годов (по данным СМИ и соцсетей). Отследить динамику их частотности в основном корпусе НКРЯ (если они туда попали) или в веб-подкорпусе. Построить кривую «жизненного цикла» неологизма (от резкого взлета до стабилизации или угасания).
19. **Сравнительное исследование дискурсивов.** Сформулировать и проверить гипотезу о национальной специфике вежливости. Сравнить частотность и прагматические функции маркеров *«пожалуйста»* и *«спасибо»* в НКРЯ с их аналогами (*please*, *thank you*) в British National Corpus (BNC) на материале диалогических текстов.
20. **Магистерская исследовательская заявка.** На основе любого из выполненных практических заданий (например, №6, 8 или 18) подготовить развернутый проект статьи: сформулировать исследовательский вопрос, уточнить корпусную методологию, описать дизайн эксперимента, определить критерии оценки результатов и список литературы (10–15 источников).

Темы докладов

- Параметризация метаданных в корпусе: от базовых тегов (автор, дата, жанр) к многоуровневым аннотациям (социолингвистические переменные, прагматические пресуппозиции) и их влиянию на результаты поиска.

- Автоматическое снятие омонимии (WSD) на материале национального корпуса: сравнение алгоритмов, основанных на словарных определениях, и нейросетевых подходов.
- Коллокации и коллигации в корпусном анализе: статистические меры ассоциации (MI-score, T-score, LogDice) и их применимость для выявления устойчивых сочетаний в русском языке.
- Диахронические корпуса и отслеживание языковых изменений: корпусный анализ неологизмов, архаизмов и грамматикализации конструкций на материале исторического подкорпуса.
- Корпусные методы в лексикографии: от словарных помет к данным реального употребления (на примере создания или верификации статей толкового или двуязычного словаря).
- Синтаксические парсеры и дерево зависимостей: оценка качества автоматического синтаксического анализа на материале размеченного корпуса сложных предложений.
- Корпусный анализ дискурсивных маркеров: выявление и количественное описание прагматических функций частиц и коннекторов (например, «ведь», «же», «мол») в устной и письменной речи.
- Создание узкоспециализированных корпусов: методология сбора, очистки и аннотирования текстов профессиональной или субкультурной коммуникации (например, корпус медицинской выписки, геймерского сленга или судебной речи).
- Корпусные исследования перевода: анализ переводческих стратегий, явлений эксплицитации и имплицитации на материале параллельных корпусов (англо-русского, русско-китайского).
- Субкорпусы устной речи: проблемы транскрибирования, разметки паралингвистических явлений (паузы, вздохи, смех) и их учет в синтаксическом и просодическом анализе.
- Применение дистрибутивной семантики и векторных моделей к корпусным данным: выявление семантических полей и скрытых семантических сдвигов.
- Корпусный анализ медиадискурса: отслеживание динамики частотности идеологем, метафор и фреймов в новостных текстах за определенный период.
- Методы оценки репрезентативности и баланса корпуса: статистические критерии и практические проблемы при формировании сбалансированного подкорпуса.
- Корпусные исследования усвоения второго языка: анализ межъязыковой интерференции и типичных ошибок на материале learner corpora (например, RICLE — Russian Interlanguage Corpus of Learners of English).
- Критическая корпусная лингвистика: этические аспекты, проблема репрезентативности маргинальных социолектов и предвзятость (bias) в обучающих данных.

Шкала и критерии оценивания *(шкала и критерии оценивания могут быть едиными (типовыми) для всех дисциплин (модулей), входящих в основную образовательную программу)*

Оценка качества учебных достижений обучающихся по дисциплине осуществляется в виде текущего контроля успеваемости и промежуточной аттестации.

Требования к структуре и содержанию доклада:

1) регламент выступления - 10 минут; наличие текста доклада и компьютерной презентации (презентация и текст доклада не должны дублировать друг друга);

2) использование самостоятельно подобранного текстового материала и его анализа в аспекте, определяемом темой доклада;

3) наличие учебного задания для студентов группы по теме доклада.

Критерии оценки доклада

Критерии	Показатели
Наличие необходимых элементов	Есть письменный текст доклада, компьютерная презентация, не дублирующие друг друга, студент выступил с докладом
Соблюдение требований к структуре и содержанию доклада	В докладе используется самостоятельно подобранный текстовый материал, есть его анализ и учебное задание

Шкала оценивания доклада (в баллах):

- все требования соблюдены, необходимые элементы есть, выполнено качественно: 15 б.

- все требования соблюдены, необходимые элементы есть, выполнено качественно, но презентация и текст доклада дублируют друга друга – 10 б

- не все требования соблюдены, не все необходимые элементы в наличии (нет текста доклада/презентации/нет самостоятельно подобранного текстового материала/задания – 5-7 баллов).

- доклад не представлен – 0 баллов.

7. Методические материалы

7.1. Перечень основной и дополнительной учебной литературы и иные материалы

Основная литература:

1. Шунейко А. А. Корпусная лингвистика : учебник для вузов / А. А. Шунейко. — Москва : Издательство Юрайт, 2026. — 222 с. — (Высшее образование). — ISBN 978-5-534-13603-6. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: https://urait.ru/bcode/588309 (дата обращения: 10.10.2026).

Дополнительная литература:

1. Захаров В. И., Богданова С. Ю. Корпусная лингвистика: учебник. 3-е изд., перераб. СПб.: Изд-во С.-Петерб.ун-та, 2020. - 234с. ISBN: 978-5-288-05997-1

2. Моделирование в корпусной лингвистике: специализированные корпусы русского языка / В.П. Захаров, И.В. Азарова, О.А.Митрофанова, А.М.Попов, М.В. Хохлова; отв.ред. В.П.Захаров. – СПб.: Изд-во С.-Петерб.ун-та, 2019. - 208с. – ISBN: 978-5-288-05902-5

Периодические издания:

1. Вопросы языкознания: Научный журнал. - https://vja.ruslang.ru/ru/archive (Архив журнала с открытым доступом в Интернет).

7.2. Перечень лицензионного программного обеспечения

8. Ресурсное обеспечение

8.1. Перечень профессиональных баз данных и информационных справочных систем:

Наименование	Ссылка на ресурс	Доступность (свободный доступ / ограниченный

		доступ)
1	***2***	***3***
Информационно-справочные системы		
Библиографические базы данных по общественным наукам ИНИОН	http://inion.ru/	Свободный доступ
Методология: словарь системы основных понятий. Сайт академика РАО Новикова А.М.	http://www.anovikov.ru/dict/met_sl.htm	Свободный доступ

8.2. Перечень ресурсов информационно-телекоммуникационной сети «Интернет»:

Наименование	Ссылка на ресурс	Доступность (свободный доступ / ограниченный доступ)
1	***2***	***3***
Электронно-библиотечные системы		
ЭБС Юрайт	https://urait.ru	После регистрации с компьютера Института –

		доступ с любого компьютера, подключенного к сети Интернет.
Профессиональные базы данных		
Справочно-информационный портал ГРАМОТА.РУ – русский язык для всех	http://gramota.ru/	Свободный доступ
«ИНФРА–М»	http://znanium.com	Свободный доступ
BOOK.ru	https://www.book.ru	После регистрации с компьютера Института – доступ с любого компьютера, подключенного к сети Интернет
Научная электронная библиотека РИНЦ	https://elibrary.ru/	Свободный доступ
Научная электронная библиотека «КиберЛенинка»	https://cyberleninka.ru/	Свободный доступ

9. Язык преподавания: русский

10. Разработчики программы

Лунькова Е. С., д.филол. наук, профессор кафедры ОиРЯ ФГБОУ ВО «Государственный институт русского языка им. А. С. Пушкина»