

УТВЕРЖДАЮ
Проректор по учебной работе и
качеству образования

_____ И. А. Долгова

30 сентября 2026

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Б1.О.07

Корпусная лингвистика

Направление подготовки:	45.04.04 Интеллектуальные системы в гуманитарной среде
Профиль подготовки:	Цифровая лингвистика
Квалификация:	магистр
Форма обучения:	очная
Год начала подготовки:	2027

Рабочая программа составлена в соответствии с:

- федеральным государственным образовательным стандартом высшего образования – магистратура по направлению подготовки 45.04.04 Интеллектуальные системы в гуманитарной среде, утверждённым приказом Министерства науки и высшего образования Российской Федерации от 13 августа 2020 года № 1003.

Разработчик программы: Кириллов А. Г., кандидат филологических наук, доцент

Рабочая программа согласована с руководителем образовательной программы 45.04.04 Интеллектуальные системы в гуманитарной среде. Рабочая программа рассмотрена и одобрена на заседании кафедры перевода и межкультурной коммуникации 28 сентября 2026, протокол № 2.

1. ОБЩАЯ ХАРАКТЕРИСТИКА

1.1. Цель и задачи дисциплины

Цель изучения дисциплины — формирование у обучающихся системы фундаментальных знаний о принципах корпусной организации языковых данных, методологии проектирования и разметки лингвистических корпусов, а также практических навыков использования корпусных технологий и менеджеров в научно-исследовательской и прикладной деятельности.

Задачи изучения дисциплины:

- изучение теоретических основ современной корпусной лингвистики, типологии корпусов и стандартов представления текстовых данных в цифровой среде;
- освоение методов автоматического и полуавтоматического аннотирования (разметки) текстовых массивов (морфологического, синтаксического, семантического);
- формирование практических навыков работы с ведущими отечественными и зарубежными корпусными менеджерами (НКРЯ, BNC, Sketch Engine, AntConc и др.) для извлечения лингвистических и статистических характеристик текста;
- развитие умений проектировать, собирать (включая технологии Web-as-Corpus) и валидировать собственные специализированные корпуса текстов под конкретные исследовательские задачи.

1.2. Требования подготовленности обучающегося к освоению содержания дисциплины

Освоение дисциплины «Корпусная лингвистика» во 2 семестре опирается на компетенции, полученные студентами в ходе освоения дисциплин 1 семестра, в первую очередь — «Теоретическая лингвистика» и «Математика в гуманитарных науках».

Обучающийся перед началом освоения дисциплины должен:

Знать:

- уровни языковой системы (фонетический, морфологический, лексический, синтаксический) и базовые единицы анализа;
- базовые понятия математической логики, теории множеств и основы лингвостатистики (частота, распределение);
- принципы классификации лексикографических источников.

Уметь:

- проводить традиционный лингвистический анализ текста на различных уровнях языка;
- использовать базовые количественные и математические методы для описания гуманитарных данных.

Владеть:

- навыками работы со специализированными информационными ресурсами и электронными библиотеками;
- понятийно-терминологическим аппаратом фундаментального языкознания.

1.3. Планируемые результаты обучения по дисциплине

Код и наименование компетенции	Код и наименование индикатора достижения компетенции	Результаты обучения по дисциплине
ОПК-5. Способен применять новые информационные технологии в гуманитарных областях знаний с использованием средств интеллектуального анализа данных, машинного обучения, компьютерной лингвистики и представления знаний	ОПК-5.1. Выбирает методы интеллектуального анализа данных, машинного обучения и компьютерной лингвистики для автоматизированной обработки текстовой информации	ОПК-5.1.1. Знает архитектуру, типы и форматы разметки лингвистических корпусов, а также методы автоматизированного сбора и систематизации больших текстовых массивов
	ОПК-5.2. Реализует алгоритмы предварительной обработки, анализа и представления знаний на основе больших массивов языковых данных	ОПК-5.2.1. Умеет использовать корпусные менеджеры (Sketch Engine, AntConc, NoSketch Engine) и инструменты разметки для извлечения, фильтрации и лемматизации текстовых данных
	ОПК-5.3. Применяет специализированные программные библиотеки, инструменты ИИ и технологии интеллектуального анализа для решения прикладных лингвистических задач	ОПК-5.3.1. Владеет практическим опытом создания, аннотирования и статистического анализа собственных специализированных корпусов текстов для решения исследовательских задач

2. ОБЪЁМ, СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

2.1. Объём дисциплины

Семестр	Контактная работа			СР	Форма ПА	Итоговый объём, часов/з. е.
	Л	П	ПА			
2	–	36	9	135	зачёт/оценка	180/5
Итого:	–	36	9	135		180/5

Л — лекции, П — все виды занятий семинарского типа, ПА — промежуточная аттестация, СР — самостоятельная работа обучающегося.

2.2. Структура дисциплины

Наименование тематического раздела дисциплины	Количество часов		
	Л	П	СР
Теоретические основы и стандартизация в корпусной лингвистике		6	22
Типология лингвистических корпусов		6	22
Технологии аннотирования (разметки) корпусов		8	30
Технологии проектирования и создания корпусов		8	30
Инструментарий корпусного анализа: корпусные менеджеры		8	31
Всего:	–	36	135

2.3. Содержание тематических разделов дисциплины

Раздел 1. Теоретические основы и стандартизация в корпусной лингвистике
Предмет, история и современное состояние корпусной лингвистики. Понятие лингвистического корпуса, его отличие от электронной библиотеки, полнотекстовой базы данных и текстового массива. Основные характеристики корпуса: репрезентативность, конечность/открытость, структурированность. Проблема стандартизации корпусных данных. Международные консорциумы и стандарты (TEI — Text Encoding Initiative, EAGLES).
Раздел 2. Типология лингвистических корпусов
Классификация корпусов по типу языкового материала (письменные, устные, мультимедийные). Классификация по составу языков (одноязычные, многоязычные, параллельные, сопоставимые). Специализированные корпуса: учебные, диалектные, исторические, поэтические, корпуса терминов. Обзор ведущих национальных корпусов (Национальный корпус русского языка — НКРЯ, British National Corpus — BNC, COCA).
Раздел 3. Технологии аннотирования (разметки) корпусов
Понятие метатекстовой и лингвистической разметки (аннотации). Уровни лингвистической разметки: графическая (токенизация, сегментация), морфологическая (лемматизация, POS-tagging), синтаксическая (деревя зависимости), семантическая. Проблема многозначности и автоматическое снятие омонимии (disambiguation). Практические инструменты автоматического теггинга и парсинга текстов.

Раздел 4. Технологии проектирования и создания корпусов

Этапы разработки корпуса текстов. Определение параметров репрезентативности и принципов отбора текстового материала. Методы автоматизированного сбора данных из сети Интернет. Концепция Web-as-Corpus (Интернет как корпус). Очистка веб-текстов от разметки (web scraping, boilerplate removal) и метатекстовое описание документов.

Раздел 5. Инструментарий корпусного анализа: корпусные менеджеры

Архитектура современных корпусных систем. Архитектура и функционал локальных (AntConc, WordSmith Tools) и веб-ориентированных (Sketch Engine, NoSketch Engine) корпусных менеджеров. Языки корпусных запросов (CQL — Corpus Query Language). Построение и анализ конкордансов. Извлечение частотных списков (лексем, словоформ, n-грамм). Понятие и вычисление мер связи для выделения коллокаций (MI, T-score, LogDice). Применение корпусных методов в лексикографии, грамматических исследованиях и анализе дискурса.

2.4. Организация учебных занятий

Дисциплина может реализовываться с применением электронного обучения и дистанционных образовательных технологий.

3. ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

3.1. Учебная литература

3.1.1. Основная

1. Беляева, Л. Н. Сетевые лингвистические технологии / Л. Н. Беляева, О. Н. Камшилова, К. Р. Пиотровская ; Российский государственный педагогический университет им. А. И. Герцена. – Санкт-Петербург : Российский государственный педагогический университет им. А.И. Герцена (РГПУ), 2019. – 112 с. : ил., табл., схем. – URL: <https://biblioclub.ru/index.php?page=book&id=577480>.
2. Информационные технологии в лингвистике: учебное пособие (практикум) : [16+] / сост. Ю. А. Пирвердиева ; Северо-Кавказский федеральный университет. – Ставрополь : Северо-Кавказский Федеральный университет (СКФУ), 2019. – 182 с. – URL: <https://biblioclub.ru/index.php?page=book&id=596213>.
3. Калугян, К. Х. Информационные технологии в лингвистике : учебное пособие : [16+] / К. Х. Калугян ; Ростовский государственный экономический университет (РИНХ). – Ростов-на-Дону : Издательско-полиграфический комплекс РГЭУ (РИНХ), 2021. – 52 с. : схем., табл. – URL: <https://biblioclub.ru/index.php?page=book&id=686245>.
4. Копотев, М. Введение в корпусную лингвистику : учебное пособие : [12+] / М. Копотев. – Прага : Animedia Company, 2014. – 195 с. : ил., табл. – URL: <https://biblioclub.ru/index.php?page=book&id=375463>.
5. Моисеева, И. Ю. Квантитативная лингвистика и новые информационные технологии : учебное пособие / И. Ю. Моисеева ; Оренбургский государственный университет. – Оренбург: Оренбургский государственный университет, 2017. – 103 с. – URL: <https://biblioclub.ru/index.php?page=book&id=481797>.
6. Хроленко, А. Т. Современные информационные технологии для гуманитария : [16+] / А. Т. Хроленко, А. В. Денисов. – 5-е изд., стер. – Москва : ФЛИНТА, 2018. – 129 с. : ил. – URL: <https://biblioclub.ru/index.php?page=book&id=363413>.

3.1.2. Дополнительная

1. Антопольский, А. Б. Лингвистические информационные ресурсы / А. Б. Антопольский ; науч. ред. Д. В. Ефременко ; Институт научной информации по общественным наукам. – Москва : Институт научной информации по общественным наукам (ИНИОН) РАН, 2022. – 466 с. : табл. – (Наука, образование и технологии). – URL: <https://biblioclub.ru/index.php?page=book&id=695077>.
2. Бутусова, А. С. Машинный и автоматизированный перевод : учебное пособие : [16+] / А. С. Бутусова, Ю. В. Бец ; Южный федеральный университет. – Ростов-на-Дону ; Таганрог : Южный федеральный университет, 2021. – 107 с. : ил., табл. – URL: <https://biblioclub.ru/index.php?page=book&id=698672>.
3. Галич, Г. Г. Современные направления лингвистики: учебное пособие для студентов магистратуры, обучающихся по направлению 45.04.02 «Лингвистика», профиль подготовки «Практика и лингводидактика профессионально ориентированного перевода» : [16+] / Г. Г. Галич. – Омск : Омский государственный университет им. Ф.М. Достоевского, 2020. – 184 с. – URL: <https://biblioclub.ru/index.php?page=book&id=614045>.

3.2. Профессиональные базы данных и информационные справочные системы

Наименование ресурса	Адрес	Тип доступа
Электронно-библиотечные системы		
ЭБС «Университетская библиотека онлайн»	https://biblioclub.ru/	Индивидуальный неограниченный доступ после регистрации
КиберЛенинка, российская научная электронная библиотека	https://cyberleninka.ru/	Открытый ресурс
Профессиональные базы данных и информационные справочные системы		
eLIBRARY.RU, российский информационно-аналитический портал	https://elibrary.ru/	Открытый ресурс
Генеральный Интернет-корпус Русского Языка (ГИКРЯ)	http://www.webcorpora.ru/	Открытый ресурс
Корпус русского литературного языка	http://narusco.ru/	Открытый ресурс
Национальный корпус русского языка (НКРЯ)	https://ruscorpora.ru/new/	Открытый ресурс
British National Corpus (BNC)	http://www.natcorp.ox.ac.uk/ ; https://www.english-corpora.org/bnc/	Открытый ресурс
Corpus of Contemporary American English (COCA)	https://www.english-corpora.org/coca/	Открытый ресурс
Corpus of Historical American English (COHA)	https://www.english-corpora.org/coha/	Открытый ресурс
International Computer Archive of Modern and Medieval English	http://icame.uib.no/	Открытый ресурс
iWeb corpus	https://www.english-corpora.org/iweb/	Открытый ресурс
NOW corpus (News on the Web)	https://www.english-corpora.org/now/	Открытый ресурс

3.3. Сетевые ресурсы

Наименование ресурса	Адрес
Не используются	

3.4. Методическое обеспечение дисциплины

3.4.1. Методические указания для преподавателя

Изучение дисциплины проводится в форме лекций, практических занятий, организации самостоятельной работы студентов, консультаций.

Лекции реализуются через изложение учебного материала с возможным мультимедийным сопровождением. Основными целями лекции являются системное освещение ключевых понятий и положений по соответствующей теме, обзор и оценка существующей проблематики, её методологических и социокультурных оснований, возможных вариантов решения, предложение методических рекомендаций для дальнейшего изучения курса, в том числе литературы и источников. Лектор должен стимулировать обучающихся к участию в обсуждении вопросов лекционного занятия, к высказыванию собственной точки зрения по обсуждаемой проблеме. Главное назначение лекции — обеспечить теоретическую основу обучения, развить интерес к учебной деятельности и конкретной учебной дисциплине, сформировать у обучающихся ориентиры для самостоятельной работы над курсом.

Практические занятия направлены на развитие самостоятельности обучающихся в исследовании изучаемых вопросов и приобретение умений и навыков. Практические занятия традиционно проводятся в форме обсуждения проблемных вопросов в группе при активном участии обучающихся. Они способствуют углубленному изучению наиболее фундаментальных и сложных проблем курса, служат важной формой анализа и синтеза исследуемого материала, а также подведения итогов самостоятельной работы обучающихся, стимулируя развитие профессиональной компетентности, навыков и умений. На практических занятиях обучающиеся учатся работать с научной литературой, чётко и доходчиво излагать проблемы и предлагать варианты их решения, аргументировать свою позицию, оценивать и критиковать позиции других, свободно публично высказывать свои мысли и суждения, грамотно вести полемику и представлять результаты собственных исследований. Практические занятия проводятся в форме устных и письменных опросов, диспута, тестирования, выполнения заданий, обсуждения докладов, выполнения контрольных заданий и пр.

Результаты работы на практических занятиях учитываются преподавателем при выставлении итоговой оценки по данной дисциплине. На усмотрение преподавателя обучающиеся, активно отвечающие на занятиях и выполняющие рекомендации преподавателя при подготовке к ним, могут получить повышающий балл к своей оценке в рамках промежуточной аттестации.

3.4.2. Методические рекомендации по самостоятельной работе обучающихся

Самостоятельная работа обучающихся выполняется по заданию и при методическом руководстве преподавателя, но без его непосредственного участия. Самостоятельная работа подразделяется на самостоятельную работу на аудиторных занятиях и на внеаудиторную самостоятельную работу. Самостоятельная работа обучающихся включает как полностью самостоятельное освоение отдельных тем (разделов) дисциплины, так и проработку тем (разделов), осваиваемых во время аудиторной работы. Во время самостоятельной работы обучающиеся читают и конспектируют учебную, научную и справочную литературу, выполняют задания, направленные на закрепление знаний и отработку умений и навыков, готовятся к текущему и промежуточному контролю по дисциплине.

Самостоятельная работа включается в общую трудоёмкость учебной нагрузки обучающегося.

Реализация поставленной цели предполагает решение следующих задач:

- освоение и расширение теоретических знаний по изучаемой дисциплине;
- систематизация и закрепление полученных теоретических знаний и практических навыков;
- формирование умений по поиску и использованию нормативной, справочной и специальной литературы, а также других источников информации;
- развитие познавательных способностей и активности, творческой инициативы, самостоятельности, ответственности и организованности;

- формирование самостоятельности мышления, способностей к саморазвитию, самообразованию, самосовершенствованию и самореализации;
- развитие научно-исследовательских навыков;
- формирование умения решать практические задачи (в профессиональной деятельности), используя приобретённые знания, способности и навыки.

Самостоятельная работа должна быть выполнена индивидуально или являться частью коллективной работы (в случае выполнения группового задания в работе делается соответствующая оговорка).

3.5. Материально-техническое обеспечение дисциплины

3.5.1. Аудитории для проведения занятий

Специальные помещения представляют собой учебные аудитории для проведения занятий лекционного типа, занятий семинарского типа, выполнения курсовых работ, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, а также помещения для самостоятельной работы и помещения для хранения и профилактического обслуживания учебного оборудования.

Специальные помещения укомплектованы учебной мебелью, в том числе мебелью для преподавателя дисциплины, учебной доской.

Обучающиеся из числа лиц с ограниченными возможностями здоровья обеспечиваются электронными образовательными ресурсами в формах, адаптированных к ограничениям их здоровья.

Материально-техническое оснащение учебных аудиторий конкретизировано на официальном сайте Университета в подразделе «Материально-техническое обеспечение и оснащённость образовательного процесса. Доступная среда» раздела «Сведения об образовательной организации».

3.5.2. Оборудование и технические средства обучения

Специальные помещения укомплектованы демонстрационным оборудованием (мультимедийный проектор, экран, компьютер, звуковые колонки).

Помещения для самостоятельной работы обучающихся оснащены компьютерной техникой с возможностью подключения к сети Интернет и обеспечением доступа в электронную информационно-образовательную среду Университета.

Электронно-библиотечная система (электронная библиотека) и электронная информационно-образовательная среда обеспечивают одновременный доступ не менее 25 процентов обучающихся.

3.5.3. Программное обеспечение

Наименование	Сведения о лицензии
Moodle, среда дистанционного обучения	GNU GPL, свободно распространяемое с открытым исходным кодом

4. ОЦЕНОЧНЫЕ СРЕДСТВА

4.1. Методика проведения текущего контроля успеваемости и промежуточной аттестации

Оценивание уровня учебных достижений обучающихся по дисциплине осуществляется в виде текущего контроля успеваемости и промежуточной аттестации. Фонд оценочных средств по дисциплине включает оценочные материалы, направленные на проверку освоения компетенций, в том числе знаний, умений и навыков. Фонд оценочных средств включает оценочные средства текущего контроля и оценочные средства промежуточной аттестации.

4.1.1. Балльно-рейтинговая карта дисциплины

Виды работы	Виды контроля, критерии оценки и количество баллов
Текущий контроль	
Аудиторная работа (0-20 баллов)	Посещение занятий и участие в работе: 10 баллов — посещение не менее 50% занятий 15 баллов — посещение 60-80% занятий, участие в обсуждениях материала 20 баллов — посещение 90-100% занятий, активное участие в обсуждениях материала
Самостоятельная работа (0-40 баллов)	
Промежуточная аттестация	
Контрольное задание (0-40 баллов)	Выполнение письменного контрольного задания в СДО Moodle

4.1.2. Шкала перевода рейтинговых баллов в пятибалльную систему оценивания

Общее количество баллов	Оценка сформированности компетенций	Оценка результатов обучения по дисциплине	Оценка ECTS
0–34	Компетенции не сформированы.	неудовлетворительно (не зачтено)	F
Теоретическое содержание не освоено, практические навыки не сформированы, большинство учебных заданий не выполнено, качество их выполнения минимальное, все задания содержат грубые ошибки. Обучающийся не готов решать типовые профессиональные задачи.			
35–49	Уровень владения компетенциями недостаточный для их формирования в результате обучения по дисциплине.	неудовлетворительно (не зачтено)	FX
Теоретическое содержание освоено частично, практические навыки не сформированы, большинство учебных заданий не выполнено, качество их выполнения минимальное, большинство заданий выполнено с ошибками. Обучающийся не готов решать типовые профессиональные задачи.			
50–59	Уровень владения компетенциями	удовлетворительно	E

Общее количество баллов	Оценка сформированности компетенций	Оценка результатов обучения по дисциплине	Оценка ECTS
	посредственный для их формирования в результате обучения по дисциплине.	(зачтено)	
Теоретическое содержание освоено частично, практические навыки сформированы фрагментарно, многие учебные задания не выполнены, качество их выполнения минимальное, многие задания выполнены с ошибками. Обучающийся готов решать типовые профессиональные задачи.			
60–69	Уровень владения компетенциями удовлетворительный для их формирования в результате обучения по дисциплине.	удовлетворительно (зачтено)	D
Теоретическое содержание освоено частично, пробелы не носят существенного характера, практические навыки в основном сформированы, большинство учебных заданий выполнено, качество их выполнения удовлетворительное, некоторые задания выполнены с ошибками. Обучающийся готов решать типовые профессиональные задачи.			
70–89	Уровень владения компетенциями преимущественно высокий для их формирования в результате обучения по дисциплине.	хорошо (зачтено)	C
Теоретическое содержание освоено полностью, некоторые практические навыки сформированы недостаточно, все учебные задания выполнены, качество их выполнения высокое, некоторые задания выполнены с ошибками. Обучающийся готов решать типовые и ситуативные профессиональные задачи.			
90–94	Уровень владения компетенциями высокий для их формирования в результате обучения по дисциплине.	отлично (зачтено)	B
Теоретическое содержание освоено полностью, практические навыки сформированы, все учебные задания выполнены, качество их выполнения близко к максимальному, однако есть несколько незначительных ошибок. Обучающийся готов эффективно решать типовые и ситуативные профессиональные задачи, в том числе повышенного уровня сложности.			
95–100	Уровень владения компетенциями превосходный для их формирования в результате обучения по дисциплине.	отлично (зачтено)	A
Теоретическое содержание освоено полностью, практические навыки сформированы, все учебные задания выполнены, качество их выполнения близко к максимальному. Обучающийся готов эффективно решать типовые и ситуативные профессиональные задачи, в том числе повышенного уровня сложности, и способен разрабатывать новые решения.			

4.2. Оценочные средства текущего контроля

4.2.1. Вопросы для подготовки к семинарским занятиям

Раздел 1. Теоретические основы и стандартизация в корпусной лингвистике

1. Дайте академические определения базовым понятиям: лингвистический корпус, корпусная лингвистика, репрезентативность, метаданные. Обоснуйте статус корпусной лингвистики как методологии и её противопоставление традиционной интроспекции.

2. Проведите строгое концептуальное разграничение: в чём отличие лингвистического корпуса от электронной библиотеки, полнотекстовой базы данных и случайного текстового массива?
3. Какой корпус текстов в истории мировой науки считается первым? Укажите его ключевые характеристики, объем и роль Брауновского корпуса (Brown Corpus) в стандартизации исследований американского варианта английского языка.
4. Раскройте содержание триады основных характеристик корпуса: репрезентативность, конечность/открытость (динамические/мониторинговые корпуса) и структурированность. Как математическое понятие случайного отбора обеспечивает репрезентативность языковой совокупности?
5. В чём заключается сущность и проблемы лингвостатистического анализа корпусных данных? Сформулируйте его принципиальную цель при изучении распределения языковых единиц.
6. Опишите проблему стандартизации корпусных данных. Почему международный обмен англоязычными и многоязычными текстовыми ресурсами невозможен без единых технических условий?
7. Что подлежит стандартизации в корпусной лингвистике? Раскройте роль и базовые принципы консорциума TEI (Text Encoding Initiative) при разметке документов с помощью XML-тегов.
8. Каковы цели создания и основные направления деятельности международных экспертных групп (в частности, инициативы EAGLES) в области выработки стандартов для компьютерной обработки английского и других европейских языков?
9. Как в корпусных исследованиях определяются статистические понятия доверительного интервала, относительной ошибки и надёжности данных при проверке лингвистических гипотез?
10. Опишите методологию корпусного анализа: перечислите основные методы корпусной лингвистики и опишите, как они применяются в междисциплинарных ИТ-проектах.
11. В чём заключается разница между корпус-ориентированным (corpus-based) и корпус-обусловленным (corpus-driven) подходами к лингвистическому исследованию англоязычного дискурса?
12. Какие лингвистические совокупности и текстовые выборки называют сбалансированными, и по каким параметрам балансируются современные академические корпуса?

Раздел 2. Типология лингвистических корпусов

1. Охарактеризуйте классификацию лингвистических корпусов по критерию прагматической и исследовательской ориентированности. Опишите свойства исследовательских, иллюстративных и динамических корпусов.
2. Классифицируйте корпуса по типу и модальности языкового материала (письменные, устные/разговорные, мультимедийные). В чём заключается специфика и сложность создания корпусов устной английской речи (например, Spoken BNC)?
3. Раскройте типологию корпусов по составу языков. Каково основное назначение параллельных корпусов текстов и в чём их принципиальное отличие от сопоставимых (comparable) корпусов при проведении контрастивных англо-русских исследований?
4. Каковы цели проектирования и сферы применения специализированных корпусов: учебных (learner corpora), диалектных, исторических, поэтических и корпусов терминов?
5. Что такое лонгитюдный корпус текстов? Каковы особенности его проектирования и как он применяется для отслеживания динамики речевого развития (например, при усвоении английского языка)?
6. Дайте подробную характеристику British National Corpus (BNC): история создания, объем, структура, принципы репрезентативности британского национального дискурса и соотношение устной и письменной частей.
7. Опишите архитектуру, жанровый состав и исследовательский потенциал корпуса COCA (Corpus of Contemporary American English). Как COCA позволяет отслеживать актуальные

лексико-грамматические сдвиги в современном американском варианте английского языка?

8. Проведите сравнительный сопоставительный анализ Национального корпуса русского языка (НКРЯ) и ведущих англоязычных национальных корпусов (BNC, COCA) по параметрам объема, доступности интерфейса и подтверждения критерия репрезентативности.
9. Назовите особенности метаразметки диалектного корпуса текста и специфику фиксации фонетических и лексических девиаций.
10. Какие виды корпусов выделяются по степени доступности и юридическим моделям использования (открытые, коммерческие, закрытые исследовательские)?
11. В чём специфика создания исторических корпусов английского языка (например, Helsinki Corpus) и с какими трудностями орфографической вариативности сталкиваются их разработчики?
12. Каким образом сопоставимые корпуса (comparable corpora) помогают выявлять кросс-культурные и этнолингвистические особенности репрезентации концептов в британском, американском и российском медиадискурсах?

Раздел 3. Технологии аннотирования (разметки) корпусов

1. Дайте определение понятиям «разметка» (аннотация) корпуса, «лемма», «токен» и «парсинг». Обоснуйте, почему неразмеченный текстовый массив не может считаться полноценным лингвистическим корпусом.
2. Подробно опишите содержание и задачи экстралингвистической (метатекстовой/внешней) разметки документов: какие социально-демографические характеристики автора и параметры текста фиксируются в англоязычных корпусах?
3. Что такое графическая и структурная разметка корпуса (токенизация, сегментация предложений). С какими краевыми задачами сталкиваются автоматические токенизаторы при обработке английских сокращений (например, St., Ph.D.) и клитик (например, don't, user's)?
4. Раскройте сущность морфологической разметки (POS-tagging). Какие наборы тегов (Tagsets) являются общепринятыми для аннотирования частей речи в английском языке (например, Penn Treebank, CLAWS)?
5. Опишите проблему многозначности, грамматической омонимии и конверсии в английском языке. Каковы современные методы автоматического снятия омонимии (disambiguation) в морфологических корпусах и какие перспективы они открывают перед исследователем?
6. В чём заключается синтаксическая разметка корпуса? Дайте определение термину Treebank и опишите, какие новые лингвистические перспективы открывают перед пользователями синтаксические корпуса английского языка.
7. Каковы задачи и лингвистические основания семантической разметки корпусов? Как тегятся семантические классы слов и тематические домены в современных корпусных системах?
8. Что такое прагматическая и дискурсивная разметка текста? Какие маркеры (речевые акты, анафорические связи, тональность) подлежат фиксации при расширенном аннотировании англоязычных коммуникативных массивов?
9. Назовите два обязательных условия (непротиворечивость и полнота) при обсуждении лингвистических процедур автоматической обработки текстов корпусов.
10. Назовите основные этапы автоматической обработки текста в конвейере корпусного аннотирования и укажите, на каких этапах возможны наибольшие системные ошибки.
11. Каким образом особенности аналитического строя английского языка облегчают или, наоборот, усложняют этап автоматического POS-теггинга по сравнению с флективным русским языком?
12. Как устроена разметка ошибок в специализированных учебных корпусах английского языка и какие классификаторы девиаций применяются в международной практике?

Раздел 4. Технологии проектирования и создания корпусов

1. Подробно опишите все основные процедуры обработки текстов на естественном языке при самостоятельном проектировании и создании лингвистического корпуса.
2. Как на этапе проектирования корпуса определяются параметры его будущей репрезентативности и формулируются строгие принципы отбора текстового материала?
3. Какие методы и программные сценарии используются для автоматизированного сбора (парсинга) текстовых данных из открытых интернет-источников для научно-исследовательских целей?
4. Раскройте сущность концепции WaC (Web-as-Corpus). Назовите причины, по которым глобальный Веб может рассматриваться как огромный многоязычный корпус, и опишите суть данной технологии.
5. Охарактеризуйте известные международные проекты по созданию интернет-корпусов на базе технологии WaC (семейство корпусов Wacky: UkWac, DeWac и др.) и их значение для лингвистики.
6. Опишите технологический процесс очистки веб-текстов от разметки (web scraping, boilerplate removal). Какие элементы HTML/CSS-кода, навигационные панели и реклама признаются шумом и подлежат обязательному удалению?
7. С какими юридическими и этическими ограничениями (авторское право, лицензии Creative Commons, правила защиты данных) сталкивается цифровой лингвист при парсинге текстов британских и американских СМИ для создания собственного исследовательского корпуса?
8. Каким образом осуществляется автоматическое определение языка (language identification) и кодировки текста при потоковой выгрузке разнородного мультязычного интернет-контента?
9. Опишите методы автоматического или полуавтоматического присвоения метатекстовых признаков (жанр, дата публикации, тип дискурса) веб-документам, извлеченным из сети Интернет.
10. Каковы требования к аппаратному и программному обеспечению для хранения, индексации и поддержки работоспособности локального исследовательского корпуса объемом более 100 миллионов словоупотреблений?
11. Как лингвист-проектировщик решает проблему орфографической нормализации вариативного текстового контента (например, блогов, комментариев, твитов на английском языке) перед этапом лингвистической разметки?
12. Проанализируйте, как концепция Web-as-Corpus изменила подходы к изучению динамических процессов в английском сленге и компьютерно-опосредованной коммуникации.

Раздел 5. Инструментарий корпусного анализа: корпусные менеджеры

1. Дайте определение терминам «конкорданс», «корпусный менеджер», «коллокат», «коллокация» и «информационный запрос». В чём отличие понятия «корпусный менеджер» в широком смысле от понятия в узком смысле?
2. Перечислите названия бесплатных корпусных менеджеров локального типа и раскройте функциональные возможности, интерфейс и архитектуру софта AntConc.
3. Охарактеризуйте возможности веб-ориентированной корпусной платформы Sketch Engine и её бесплатного открытого аналога NoSketch Engine. В чём преимущество серверных систем при анализе сверхбольших текстовых массивов?
4. Что такое язык корпусных запросов CQL (Corpus Query Language) и как он соотносится с общим языком регулярных выражений? Сформулируйте принципы построения CQL-запросов для поиска сложных лексико-грамматических паттернов в английском языке.
5. Как строится и интерпретируется конкорданс (в частности, формат KWIC — Key Word In Context)? Каким образом сортировка левого и правого контекстов помогает лингвисту выявлять скрытые семантические закономерности?
6. Опишите методику извлечения и лингвистической интерпретации частотных списков. В чём разница между абсолютной частотой словоупотребления и нормированной частотой (например, IPM — instances per million)? Зачем нужна нормализация?

7. Что такое закон Ципфа и как он проявляется при анализе частотных словарей и распределения лексики в англоязычных корпусах разного объема? Каковы сферы применения частотных словарей в ИТ?
8. Как вычисляются математические меры связи для автоматического выделения коллокаций в английском языке (MI — Mutual Information, T-score, LogDice)? Как с их помощью можно оценить «идиоматичность» в текстах?
9. Объясните, почему лингвистические корпуса (базы данных коммерческих поисковых систем Google, Яндекс) можно рассматривать как корпуса текстов, и каковы недостатки их использования в качестве корпусных менеджеров.
10. Что представляет собой внешняя и внутренняя статистика в корпусных службах и как результаты корреляции частот словоупотребления могут быть использованы разными специалистами?
11. Какие лингвистические особенности регистра художественной прозы, медиадискурса или конкретного функционального стиля (на материале английского языка) можно выявить с помощью корпусной методологии?
12. Предложите конкретные практические варианты использования корпусных методов, извлечения профилей слов и конкордансов в преподавании иностранного (английского) языка.

Критерии оценки работы на практическом занятии

Критерии	Максимальное количество баллов за занятие
Устный опрос, коллоквиум	
Основные теоретические положения по вопросу раскрыты. Имеются элементы обоснования выводов. Имеются элементы систематизации информации, факты применения профессиональной терминологии. Очевидно использование источников рекомендованной литературы.	5 баллов

4.2.2. Темы докладов

1. Стандарт TEI (Text Encoding Initiative) как технологический фундамент разметки документов в цифровой гуманитаристике.
2. История Брауновского корпуса (Brown Corpus) и его значение для становления количественных исследований американского варианта английского языка.
3. Сравнительный анализ репрезентативности и жанрового баланса British National Corpus (BNC) и Corpus of Contemporary American English (COCA).
4. Корпуса параллельных текстов (на материале англо-русских ресурсов) в теории и практике художественного и специального перевода.
5. Использование специализированных учебных корпусов (Learner Corpora) для оптимизации методик преподавания английского языка как иностранного (EFL).
6. Проблема автоматического снятия лексико-грамматической омонимии и конверсии при POS-теггинге англоязычных медиатекстов.
7. Синтаксические корпуса (Treebanks) английского языка: принципы аннотирования и сфера применения в обучении NLP-моделей.
8. Методология Web-as-Corpus: современные технологии автоматического построения терабайтных интернет-корпусов.
9. Программные инструменты очистки веб-страниц от разметки (Boilerplate Removal) и проблема сохранения метаданных документов.

10. Язык корпусных запросов CQL (Corpus Query Language) как инструмент извлечения сложных лингвистических конструкций.
11. Статистические меры связи (Mutual Information, T-score, LogDice) в задаче автоматического выделения английской терминологической фразеологии.
12. Закон Ципфа и структурно-вероятностные свойства частотных словарей английского и русского языков.
13. Корпусные методы в современной лексикографии: опыт использования Sketch Engine для создания толковых словарей новых поколений.
14. Динамические процессы в лексической системе американского варианта английского языка по данным мониторинговых корпусов.
15. Стилométrический и корпусный анализ text в задачах автоматической атрибуции авторства англоязычных интернет-документов.
16. Особенности корпусного аннотирования и анализа текстов компьютерно-опосредованной коммуникации (социальные сети, мессенджеры).
17. Корпусный анализ профилей слов (Word Sketches) в исследованиях по когнитивной лингвистике и семантическому картированию концептов.
18. Изучение феномена речевой манипуляции и политического дискурса США/Великобритании методами корпусной лингвистики.

Шкала и критерии оценки доклада

Критерии	Показатели	Баллы
1. Степень раскрытия сущности проблемы	– соответствие теме доклада; – полнота и глубина раскрытия основных понятий; – умение работать с литературой, систематизировать и структурировать материал; – умение обобщать, сопоставлять различные точки зрения по рассматриваемому вопросу, аргументировать основные положения и выводы.	70
2. Обоснованность выбора источников	– круг, полнота использования литературных источников по теме; – привлечение новейших работ (журнальные публикации, материалы сборников научных трудов и т.д.).	15
3. Изложение	– литературный стиль.	15

Доклад оценивается по 100 балльной шкале, баллы переводятся в оценки успеваемости следующим образом:

90 – 100 баллов – «отлично»;

70 – 89 баллов – «хорошо»;

50 – 69 баллов – «удовлетворительно»;

менее 50 баллов – «неудовлетворительно».

4.2.3. Примерные темы для индивидуальных проектов

1. Корпусное исследование динамики неологизмов: «Проектирование и анализ микрокорпуса русскоязычных медиатекстов 2024–2026 гг. для фиксации ИТ-метафор».
2. Сравнительный корпусный анализ (Стилметрия): «Особенности дистрибуции дискурсивных маркеров в академическом и научно-популярном лингвистическом дискурсе (на материале параллельных подкорпусов)».

3. Лингвистическая экспертиза и корпусные методы: «Выявление маркеров речевой агрессии в интернет-комментариях с использованием языка корпусных запросов (CQL)».
4. Специализированные учебные корпуса: «Создание и разметка учебного корпуса (Learner Corpus) текстов магистрантов для анализа типичных интерферентных ошибок в академическом английском языке».
5. Терминологическое моделирование: «Автоматическое извлечение двухсловных терминосистем из корпуса текстов по искусственному интеллекту с применением статистических мер ассоциации (*MI*, *T-score*)».

4.3. Оценочные средства промежуточной аттестации

4.3.1. Контрольные задания

ОПК-6.1-1. Прочитайте текст и установите соответствие.

А. электронный словарь — это:

1. введённый в компьютер бумажный словарь, снабженный средствами поиска и отображения информации
2. организованное собрание слов с комментариями, в которых описываются особенности структуры и/или функционирования этих слов
3. организованное собрание слов с описанием их значения, особенностей употребления, структурных свойств, сочетаемости, соотношения с лексическими системами других языков и т.д.
4. словарь в специальном машинном формате, предназначенный для применения на ЭВМ пользователем или компьютерной программой

А	Б	В	Г

ОПК-6.1-2. Прочитайте текст и установите соответствие.

А. Совокупность специально отобранных текстов, размеченных по различным лингвистическим параметрам и обеспеченных системой поиска:

1. база данных
2. корпус
3. информационный массив
4. словарь

А	Б	В	Г

ОПК-6.1-3. Прочитайте текст и запишите развёрнутый обоснованный ответ.

Что такое автоматическое реферирование текста?

--

ОПК-6.1-4. Прочитайте текст и запишите развёрнутый обоснованный ответ.

Что такое анализ тональности текста?

ОПК-6.1-5. Прочитайте текст и запишите развёрнутый обоснованный ответ.

Как искать по определённой морфологической форме в поисковике Яндекс?

ОПК-6.1-6. Прочитайте текст и запишите развёрнутый обоснованный ответ.

Как в поисковике найти результаты, соответствующие точной форме запроса?

ОПК-6.2-1. Прочитайте текст и запишите развёрнутый обоснованный ответ.

С использованием онлайн сервисов текстовой аналитики определите частоту слов в данном тексте (первые десять позиций).

Вот дом,
Который построил Джек.

А это пшеница,
Которая в тёмном чулане хранится
В доме,
Который построил Джек.

А это весёлая птица-синица,
Которая часто ворует пшеницу,
Которая в тёмном чулане хранится
В доме,
Который построил Джек.

Вот кот,
Который пугает и ловит синицу,
Которая часто ворует пшеницу,
Которая в тёмном чулане хранится
В доме,
Который построил Джек.

Вот пёс без хвоста,
Который за шиворот треплет кота,
Который пугает и ловит синицу,
Которая часто ворует пшеницу,
Которая в тёмном чулане хранится
В доме,

Который построил Джек.

А это корова безрогая,
 Лягнувшая старого пса без хвоста,
 Который за шиворот треплет кота,
 Который пугает и ловит синицу,
 Которая часто ворует пшеницу,
 Которая в тёмном чулане хранится
 В доме,
 Который построил Джек.

А это старушка, седая и строгая,
 Которая доит корову безрогую,
 Лягнувшую старого пса без хвоста,
 Который за шиворот треплет кота,
 Который пугает и ловит синицу,
 Которая часто ворует пшеницу,
 Которая в тёмном чулане хранится
 В доме,
 Который построил Джек.

А это ленивый и толстый пастух,
 Который бранится с коровницей строгою,
 Которая доит корову безрогую,
 Лягнувшую старого пса без хвоста,
 Который за шиворот треплет кота,
 Который пугает и ловит синицу,
 Которая часто ворует пшеницу,
 Которая в тёмном чулане хранится
 В доме,
 Который построил Джек.

Вот два петуха,
 Которые будят того пастуха,
 Который бранится с коровницей строгою,
 Которая доит корову безрогую,
 Лягнувшую старого пса без хвоста,
 Который за шиворот треплет кота,
 Который пугает и ловит синицу,
 Которая часто ворует пшеницу,
 Которая в тёмном чулане хранится
 В доме,
 Который построил Джек.

--

ОПК-6.2-2. Прочитайте текст и запишите развёрнутый обоснованный ответ.

С использованием онлайн сервисов текстовой аналитики определите частоту слов в данном тексте (первые десять позиций).

This is the house that Jack built.

This is the malt
That lay in the house that Jack built.

This is the rat,
That ate the malt
That lay in the house that Jack built.

This is the cat,
That kill'd the rat,
That ate the malt
That lay in the house that Jack built.

This is the dog,
That worried the cat,
That kill'd the rat,
That ate the malt
That lay in the house that Jack built.

This is the cow with the crumpled horn,
That toss'd the dog,
That worried the cat,
That kill'd the rat,
That ate the malt
That lay in the house that Jack built.

This is the maiden all forlorn,
That milk'd the cow with the crumpled horn,
That tossed the dog,
That worried the cat,
That kill'd the rat,
That ate the malt
That lay in the house that Jack built.

This is the man all tatter'd and torn,
That kissed the maiden all forlorn,
That milk'd the cow with the crumpled horn,
That tossed the dog,
That worried the cat,
That kill'd the rat,
That ate the malt
That lay in the house that Jack built.

This is the priest all shaven and shorn,
That married the man all tatter'd and torn,
That kissed the maiden all forlorn,
That milked the cow with the crumpled horn,
That tossed the dog,
That worried the cat,
That kill'd the rat,
That ate the malt
That lay in the house that Jack built.

This is the cock that crow'd in the morn,
That waked the priest all shaven and shorn,
That married the man all tatter'd and torn,

That kissed the maiden all forlorn,
 That milk'd the cow with the crumpled horn,
 That tossed the dog,
 That worried the cat,
 That kill'd the rat,
 That ate the malt
 That lay in the house that Jack built.

This is the farmer sowing his corn,
 That kept the cock that crow'd in the morn,
 That waked the priest all shaven and shorn,
 That married the man all tatter'd and torn,
 That kissed the maiden all forlorn,
 That milk'd the cow with the crumpled horn,
 That tossed the dog,
 That worried the cat,
 That killed the rat,
 That ate the malt
 That lay in the house that Jack built.

ОПК-6.2-3. Прочитайте текст и запишите развёрнутый обоснованный ответ.

По Национальному корпусу русского языка (НКРЯ) определите первые пять позиций в конкордансе по слову «хуйня». Определите периоды, когда лексема употреблялась чаще всего (с 1900 по 2022), в каких типах текста (первые три) и жанрах (первые три). Перечислите первые пять 3-граммы.

ОПК-6.2-4. Прочитайте текст и запишите развёрнутый обоснованный ответ.

По Национальному корпусу русского языка (НКРЯ) определите первые пять позиций в конкордансе по слову «диалект». Определите периоды, когда лексема употреблялась чаще всего (с 1900 по 2022), в каких типах текста (первые три) и жанрах (первые три). Перечислите первые пять 3-граммы.

ОПК-6.2-5. Прочитайте текст и запишите развёрнутый обоснованный ответ.

Выберите пару омографов (например, мѡловать и мѡловать) и придумайте предложения с каждым из них (например, «Она его не мѡловала и отправила на верную смерть» и «Они долго мѡловались на скамейке в парке»). Используя переводчики Яндекс, Google и DeepL, переведите эти предложения на английский язык. Оцените степень сохранности смысла. Сравните результаты перевода.

ОПК-6.2-6. Прочитайте текст и запишите развёрнутый обоснованный ответ.

Выберите пару омографов (например, minutes) и придумайте предложения с каждым из них (например, They kept minutes of the meeting и The minutes kept rolling). Используя переводчики Яндекс, Google и DeepL, переведите эти предложения на русский язык. Оцените степень сохранности смысла. Сравните результаты перевода.

--

ПК-2.1-1. Прочитайте текст и установите соответствие.

Б. корпусный менеджер:

1. обеспечивает сортировку результатов поиска, статистические подсчеты, составление списков слов на основе корпуса
2. это специальная программа поиска по корпусу
3. это человек, составляющий корпуса и управляющий ими
4. это специальная программа подготовки текстов к их включению в корпус

А	Б	В	Г

ПК-2.1-2. Прочитайте текст и установите соответствие.

А. требования к корпусам:

1. полнота, адекватность, актуальность, компьютерная поддержка
2. устойчивость, тиражируемость, адаптируемость, оптимальность временных параметров, комфорт пользователя
3. репрезентативность, полнота, экономичность, структуризация, компьютерная поддержка
4. полнота, экономичность, достоверность, структуризация, компьютерная поддержка

А	Б	В	Г

ПК-2.1-3. Прочитайте текст и запишите развёрнутый обоснованный ответ.

В чём особенность поисковика DuckDuckGo?

--

ПК-2.1-4. Прочитайте текст и запишите развёрнутый обоснованный ответ.

Что такое синтаксический анализатор?

--

ПК-2.1-5. Прочитайте текст и запишите развёрнутый обоснованный ответ.

Для чего используется FieldWorks Language Explorer?

ПК-2.1-6. Прочитайте текст и запишите развёрнутый обоснованный ответ.

Для чего используется LingTree?

ПК-2.2-1. Прочитайте текст и запишите развёрнутый обоснованный ответ.

С помощью сервиса реферирования текста (<https://www.yeschat.ai/ru/gpts-2OT09yvbWD-Text-Summarizer>) сделайте автоматический реферат раздела «Лингвистическая основа исследования» из статьи «Речевые жанры электронной коммуникации между преподавателем и студентом» (<https://zhanry-rechi.sgu.ru/ru/articles/rechevye-zhanry-elektronnoy-kommunikacii-mezhdu-prepodavatelem-i-studentom>).

ПК-2.2-2. Прочитайте текст и запишите развёрнутый обоснованный ответ.

С помощью сервиса реферирования текста (<https://www.yeschat.ai/ru/gpts-2OT09yvbWD-Text-Summarizer>) сделайте автоматический реферат раздела «Выводы» из статьи «Речевые жанры электронной коммуникации между преподавателем и студентом» (<https://zhanry-rechi.sgu.ru/ru/articles/rechevye-zhanry-elektronnoy-kommunikacii-mezhdu-prepodavatelem-i-studentom>).

ПК-2.2-3. Прочитайте текст и запишите развёрнутый обоснованный ответ.

С использованием программы AntConc и Блокнот определите первые десять наиболее частотных N-грамм в данном тексте.

Вот дом,
Который построил Джек.

А это пшеница,
Которая в тёмном чулане хранится
В доме,
Который построил Джек.

А это весёлая птица-синица,
Которая часто ворует пшеницу,

Которая в тёмном чулане хранится
В доме,
Который построил Джек.

Вот кот,
Который пугает и ловит синицу,
Которая часто ворует пшеницу,
Которая в тёмном чулане хранится
В доме,
Который построил Джек.

Вот пёс без хвоста,
Который за шиворот треплет кота,
Который пугает и ловит синицу,
Которая часто ворует пшеницу,
Которая в тёмном чулане хранится
В доме,
Который построил Джек.

А это корова безрогая,
Лягнувшая старого пса без хвоста,
Который за шиворот треплет кота,
Который пугает и ловит синицу,
Которая часто ворует пшеницу,
Которая в тёмном чулане хранится
В доме,
Который построил Джек.

А это старушка, седая и строгая,
Которая доит корову безрогую,
Лягнувшую старого пса без хвоста,
Который за шиворот треплет кота,
Который пугает и ловит синицу,
Которая часто ворует пшеницу,
Которая в тёмном чулане хранится
В доме,
Который построил Джек.

А это ленивый и толстый пастух,
Который бранится с коровницей строгою,
Которая доит корову безрогую,
Лягнувшую старого пса без хвоста,
Который за шиворот треплет кота,
Который пугает и ловит синицу,
Которая часто ворует пшеницу,
Которая в тёмном чулане хранится
В доме,
Который построил Джек.

Вот два петуха,
Которые будят того пастуха,
Который бранится с коровницей строгою,
Которая доит корову безрогую,
Лягнувшую старого пса без хвоста,
Который за шиворот треплет кота,

Который пугает и ловит синицу,
 Которая часто ворует пшеницу,
 Которая в тёмном чулане хранится
 В доме,
 Который построил Джек.

--

ПК-2.2-4. Прочитайте текст и запишите развернутый обоснованный ответ.

С использованием программы AntConc и Блокнот определите первые десять наиболее частотных N-грамм в данном тексте.

This is the house that Jack built.

This is the malt
 That lay in the house that Jack built.

This is the rat,
 That ate the malt
 That lay in the house that Jack built.

This is the cat,
 That kill'd the rat,
 That ate the malt
 That lay in the house that Jack built.

This is the dog,
 That worried the cat,
 That kill'd the rat,
 That ate the malt
 That lay in the house that Jack built.

This is the cow with the crumpled horn,
 That toss'd the dog,
 That worried the cat,
 That kill'd the rat,
 That ate the malt
 That lay in the house that Jack built.

This is the maiden all forlorn,
 That milk'd the cow with the crumpled horn,
 That tossed the dog,
 That worried the cat,
 That kill'd the rat,
 That ate the malt
 That lay in the house that Jack built.

This is the man all tatter'd and torn,
 That kissed the maiden all forlorn,
 That milk'd the cow with the crumpled horn,
 That tossed the dog,
 That worried the cat,

That kill'd the rat,
That ate the malt
That lay in the house that Jack built.

This is the priest all shaven and shorn,
That married the man all tatter'd and torn,
That kissed the maiden all forlorn,
That milked the cow with the crumpled horn,
That tossed the dog,
That worried the cat,
That kill'd the rat,
That ate the malt
That lay in the house that Jack built.

This is the cock that crow'd in the morn,
That waked the priest all shaven and shorn,
That married the man all tatter'd and torn,
That kissed the maiden all forlorn,
That milk'd the cow with the crumpled horn,
That tossed the dog,
That worried the cat,
That kill'd the rat,
That ate the malt
That lay in the house that Jack built.

This is the farmer sowing his corn,
That kept the cock that crow'd in the morn,
That waked the priest all shaven and shorn,
That married the man all tatter'd and torn,
That kissed the maiden all forlorn,
That milk'd the cow with the crumpled horn,
That tossed the dog,
That worried the cat,
That killed the rat,
That ate the malt
That lay in the house that Jack built.

ПК-2.2-5. Прочитайте текст и запишите развёрнутый обоснованный ответ.

С помощью сервиса Google Books Ngram Viewer определите динамику использования слов *ass*, *cunt*, *dick* в английском языке в период с 1800 по 2022 год, без учёта написания (case-insensitive) и параметром *smoothing* of 2. Сделайте выводы по полученному графику.

ПК-2.2-6. Прочитайте текст и запишите развёрнутый обоснованный ответ.

С помощью сервиса Google Books Ngram Viewer определите динамику использования слов *pupil*, *student* в английском языке в период с 1800 по 2022 год, без учёта написания (case-insensitive) и параметром *smoothing* of 2. Сделайте выводы по полученному графику.

4.3.2. Ключи к контрольным заданиям

ОПК-6.1-1	A4																																				
ОПК-6.1-2	A2																																				
ОПК-6.1-3	Автоматическое реферирование текста (или реферирование документа) — это метод обработки естественного языка (NLP), который сжимает информацию из одного или нескольких входных текстовых документов в исходный выходной текст. / Автоматическое реферирование — извлечение наиболее важных сведений из одного или нескольких документов и составление их краткого описания.																																				
ОПК-6.1-4	Анализ тональности — это процесс анализа цифрового текста для определения того, является ли эмоциональный тон сообщения положительным, отрицательным или нейтральным. / Анализ тональности текста — задача компьютерной лингвистики, заключающаяся в определении эмоциональной окраски (тональности) текста и, в частности, в выявлении эмоциональной оценки авторов по отношению к объектам, описываемым в тексте.																																				
ОПК-6.1-5	Нужно поставить перед словом ! (без пробела). Если слово запроса написано с большой буквы, то при поиске будет учтён регистр.																																				
ОПК-6.1-6	Нужно ввести запрос в кавычках. Тогда будут найдены результаты именно в такой словоформе и таком порядке следования слов.																																				
ОПК-6.2-1	<table><tr><td>Words sorted by its frequency</td><td>Occurrences</td><td>Frequency</td></tr><tr><td>который</td><td>22</td><td>8.91</td></tr><tr><td>которая</td><td>18</td><td>7.29</td></tr><tr><td>в</td><td>16</td><td>6.48</td></tr><tr><td>построил</td><td>9</td><td>3.64</td></tr><tr><td>джек</td><td>9</td><td>3.64</td></tr><tr><td>и</td><td>8</td><td>3.24</td></tr><tr><td>чулане</td><td>8</td><td>3.24</td></tr><tr><td>хранится</td><td>8</td><td>3.24</td></tr><tr><td>тёмном</td><td>8</td><td>3.24</td></tr><tr><td>доме</td><td>8</td><td>3.24</td></tr><tr><td>часто</td><td>7</td><td>2.83</td></tr></table>	Words sorted by its frequency	Occurrences	Frequency	который	22	8.91	которая	18	7.29	в	16	6.48	построил	9	3.64	джек	9	3.64	и	8	3.24	чулане	8	3.24	хранится	8	3.24	тёмном	8	3.24	доме	8	3.24	часто	7	2.83
Words sorted by its frequency	Occurrences	Frequency																																			
который	22	8.91																																			
которая	18	7.29																																			
в	16	6.48																																			
построил	9	3.64																																			
джек	9	3.64																																			
и	8	3.24																																			
чулане	8	3.24																																			
хранится	8	3.24																																			
тёмном	8	3.24																																			
доме	8	3.24																																			
часто	7	2.83																																			
ОПК-6.2-2	<table><tr><td>Words sorted by its frequency</td><td>Occurrences</td><td>Frequency</td></tr><tr><td>the</td><td>74</td><td>18.50</td></tr><tr><td>that</td><td>68</td><td>17.00</td></tr><tr><td>in</td><td>12</td><td>3.00</td></tr><tr><td>all</td><td>12</td><td>3.00</td></tr><tr><td>is</td><td>11</td><td>2.75</td></tr><tr><td>this</td><td>11</td><td>2.75</td></tr><tr><td>jack</td><td>11</td><td>2.75</td></tr><tr><td>house</td><td>11</td><td>2.75</td></tr><tr><td>built</td><td>11</td><td>2.75</td></tr><tr><td>malt</td><td>10</td><td>2.50</td></tr><tr><td>lay</td><td>10</td><td>2.50</td></tr></table>	Words sorted by its frequency	Occurrences	Frequency	the	74	18.50	that	68	17.00	in	12	3.00	all	12	3.00	is	11	2.75	this	11	2.75	jack	11	2.75	house	11	2.75	built	11	2.75	malt	10	2.50	lay	10	2.50
Words sorted by its frequency	Occurrences	Frequency																																			
the	74	18.50																																			
that	68	17.00																																			
in	12	3.00																																			
all	12	3.00																																			
is	11	2.75																																			
this	11	2.75																																			
jack	11	2.75																																			
house	11	2.75																																			
built	11	2.75																																			
malt	10	2.50																																			
lay	10	2.50																																			
ОПК-6.2-3	Конкорданс: скатилась до той самой хуйни настолько хуйня																																				

	хотя хуйню взял и спорол зачем-то нормальные татары такой хуйнёй как ты не страдают Ну, это хуйня-желание. Частотность по периодам: Фиксируется с 1955 года. Пики употребления с 1982 по 1985 и 2008 (наибольшая частотность). Типы текста: роман повесть рассказ Жанры: нежанровая проза документальная проза триллер 3-граммы: что за хуйня это все хуйня хуйня а не все это хуйня за всю хуйню
ОПК-6.2-4	Конкорданс: на каком диалекте он писал ввернуть словечко на местном диалекте Диалекты, на основе которых развились эти языки носители раннеславянских диалектов или их предки основу которого заложили диалекты англов, саксов и ютов Частотность по периодам: Пик использования пришёлся на 1949-1953 годы. Затем менее выраженные пики приходились на 1990-е и начало 2010-х годов. Типы текста: статья монография письмо Жанры: нежанровая проза документальная проза юмор и сатира 3-граммы: на французском диалекте языков и диалектов на местном диалекте на разных диалектах на всех диалектах
ОПК-6.2-5	Она его не миловала и отправила на верную смерть Яндекс: She didn't have mercy on him and sent him to his death. Google: She did not spare him and sent him to certain death. DeepL: She had no mercy on him and sent him to certain death. Они долго миловались на скамейке в парке Яндекс: They made out for a long time on a park bench. Google: They spent a long time making out on a bench in the park. DeepL: They made out for a long time on a park bench. «Миловаться» означает «ласкать(ся)», а <i>make out</i> – «целоваться и ласкать»
ОПК-6.2-6	They kept minutes of the meeting

	Яндекс: Они вели протокол заседания. Google: Они вели протокол встречи. DeepL: Они вели протокол собрания. The minutes kept rolling Яндекс: Минуты шли своим чередом. Google: Минуты продолжали течь. DeepL: Минуты продолжали идти. / Минуты шли своим чередом.
ПК-2.1-1	A1
ПК-2.1-2	A3
ПК-2.1-3	DuckDuckGo позиционирует себя как поисковую систему, которая ставит на первое место конфиденциальность пользовательской информации. Она не хранит IP-адреса, не ведёт лог пользовательской информации и использует куки, только когда это необходимо. / Это поисковая система, ориентированная на приватность: она не хранит данные пользователей и не передаёт их третьим лицам.
ПК-2.1-4	Парсер, или синтаксический анализатор, — часть программы, преобразующей входные данные (как правило, текст) в некий структурированный формат, нужный для задач последующего их (данных) анализа и использования. Синтаксический анализ в лингвистике и информатике — процесс сопоставления линейной последовательности лексем (слов, токенов) естественного или формального языка с его формальной грамматикой. Результатом обычно является дерево разбора (синтаксическое дерево).
ПК-2.1-5	FieldWorks Language Explorer — это комплексный инструмент, позволяющий создать словарь для выбранного языка путём сбора текстов, слов и культурной информации. Он широко используется для документирования и сохранения языков и культур с ограниченными ресурсами. С помощью FieldWorks можно составлять тексты, анализировать грамматику и учитывать культурные особенности. Программа упрощает задачу интерлинеаризации текста и анализа грамматики, а также имеет встроенные функции для публикации словаря или грамматического анализа.
ПК-2.1-6	LingTree — это инструмент, позволяющий быстро создавать лингвистические древовидные диаграммы (синтаксические, морфологические или фонологические) и сохранять их в виде графических файлов.
ПК-2.2-1	– Проблема определения речевого жанра остаётся актуальной из-за сложности и многообразия проявлений этого феномена. – Исследование сосредоточено на речевых жанрах электронной коммуникации между преподавателем и студентами. – К. Ф. Седов определяет речевой жанр как вербальное оформление типичных ситуаций социального взаимодействия. – Т. В. Шмелёва предлагает классификацию речевых жанров по коммуникативной цели: • Информативные (обмен информацией), • Императивные (побуждение к действию), • Этикетные (соответствие социальным нормам), • Оценочные (соотнесение поступков с социальными ценностями). – Эта классификация активно используется в исследованиях педагогического и электронного дискурса. – С. В. Волюнкина и Е. Н. Кузина классифицируют первичные жанры педагогического дискурса по тем же четырём типам. – Л. Ю. Щипицина описывает жанры компьютерно-ориентированной коммуникации как вторичные и выделяет шесть типов: информативные, директивные, фатические, презентационные, эстетические и

	развлекательные. – Для исследования электронной коммуникации преподаватель-студент предпочтительна модель Шмелёвой из-за её соответствия деловому дискурсу и сохранению традиционных форм общения. Вывод: Модель Шмелёвой признана наиболее подходящей для анализа жанров в электронной коммуникации между преподавателем и студентом. Её классификация учитывает цели общения, а специфичность ситуации диктует сохранение традиционных норм делового дискурса, несмотря на цифровой формат.																				
ПК-2.2-2	– Электронная коммуникация между преподавателем и студентом изучается как особая форма взаимодействия. – Основной фокус исследования – жанровое разнообразие такой коммуникации. – Выявлены характеристики речевых жанров: вторичность, письменность, официальность, но также элементы устной и неофициальной речи. – Это говорит о формирующемся характере данного типа общения. – Студенты часто испытывают трудности в коммуникативной и жанровой компетентности в официальном виртуальном общении. – Выделены четыре группы жанров: • Информативные – самые многочисленные, выполняют задачу устранения информационного дефицита. • Этикетные – почти столь же распространены, показывают сохранение норм вежливости. • Императивные и оценочные – преимущественно используются преподавателями. – Общение характеризуется нестабильностью жанров и может измениться в будущем. Вывод: Электронное взаимодействие в образовании пока формируется, сочетает черты делового и повседневного общения, а его жанровая структура может трансформироваться с накоплением опыта у участников.																				
ПК-2.2-3	<table> <tr><td>который построил</td><td>9</td></tr> <tr><td>построил джек</td><td>9</td></tr> <tr><td>в доме</td><td>8</td></tr> <tr><td>в тёмном</td><td>8</td></tr> <tr><td>доме который</td><td>8</td></tr> <tr><td>которая в</td><td>8</td></tr> <tr><td>тёмном чулане</td><td>8</td></tr> <tr><td>хранится в</td><td>8</td></tr> <tr><td>чулане хранится</td><td>8</td></tr> <tr><td>ворует пшеницу</td><td>7</td></tr> </table>	который построил	9	построил джек	9	в доме	8	в тёмном	8	доме который	8	которая в	8	тёмном чулане	8	хранится в	8	чулане хранится	8	ворует пшеницу	7
который построил	9																				
построил джек	9																				
в доме	8																				
в тёмном	8																				
доме который	8																				
которая в	8																				
тёмном чулане	8																				
хранится в	8																				
чулане хранится	8																				
ворует пшеницу	7																				
ПК-2.2-4	<table> <tr><td>in the</td><td>12</td></tr> <tr><td>house that</td><td>11</td></tr> <tr><td>is the</td><td>11</td></tr> <tr><td>jack built</td><td>11</td></tr> <tr><td>that jack</td><td>11</td></tr> <tr><td>the house</td><td>11</td></tr> <tr><td>this is</td><td>11</td></tr> <tr><td>built this</td><td>10</td></tr> <tr><td>lay in</td><td>10</td></tr> <tr><td>malt that</td><td>10</td></tr> </table>	in the	12	house that	11	is the	11	jack built	11	that jack	11	the house	11	this is	11	built this	10	lay in	10	malt that	10
in the	12																				
house that	11																				
is the	11																				
jack built	11																				
that jack	11																				
the house	11																				
this is	11																				
built this	10																				
lay in	10																				
malt that	10																				
ПК-2.2-5	Слово <i>cunt</i> начинает использоваться после 1960 года. Слова <i>dick</i> и <i>ass</i> отмечены																				

	на всём протяжении. С 2007 года <i>ass</i> употребляется чаще, чем <i>dick</i> . <i>Ass</i> было частотным и до 1865 года, а также в короткий период 1966-1969 годов.
ПК-2.2-6	Практически во весь наблюдаемый период <i>student</i> употреблялось чаще, чем <i>pupil</i> , достигнув пика в 2006 году.

Шкала и критерии оценки текущего тестирования

Число правильных ответов	Оценка
90-100% правильных ответов	Оценка «отлично»
70-89% правильных ответов	Оценка «хорошо»
50-69% правильных ответов	Оценка «удовлетворительно»
Менее 50% правильных ответов	Оценка «неудовлетворительно»

5. ОСОБЕННОСТИ РЕАЛИЗАЦИИ ДИСЦИПЛИНЫ В ОТНОШЕНИИ ЛИЦ ИЗ ЧИСЛА ИНВАЛИДОВ И ЛИЦ С ОГРАНИЧЕННЫМИ ВОЗМОЖНОСТЯМИ ЗДОРОВЬЯ

Профессорско-преподавательский состав знакомится с психолого-физиологическими особенностями обучающихся инвалидов и лиц с ограниченными возможностями здоровья, индивидуальными программами реабилитации инвалидов (при наличии). При необходимости осуществляется дополнительная поддержка преподавания психологами, социальными работниками, прошедшими подготовку ассистентами.

В соответствии с методическими рекомендациями Министерства науки и высшего образования Российской Федерации, в курсе предполагается использовать социально-активные и рефлексивные методы обучения, технологии социокультурной реабилитации с целью оказания помощи в установлении полноценных межличностных отношений с другими студентами, создания комфортного психологического климата в студенческой группе. Подбор и разработка учебных материалов производятся с учётом предоставления материала в различных формах: аудиальной, визуальной, с использованием специальных технических средств и информационных систем.

Согласно требованиям к порядку реализации образовательной деятельности в отношении инвалидов и лиц с ОВЗ, установленным Министерством науки и высшего образования Российской Федерации, необходимо иметь в виду, что:

1) инвалиды и лица с ОВЗ по зрению имеют право присутствовать на занятиях вместе с ассистентом, оказывающим обучающемуся необходимую помощь;

2) инвалиды и лица с ОВЗ по слуху имеют право на использование звукоусиливающей аппаратуры.

При проведении промежуточной аттестации по дисциплине обеспечивается соблюдение следующих общих требований:

- проведение аттестации для инвалидов в одной аудитории совместно с обучающимися, не являющимися инвалидами, если это не создает трудностей для инвалидов и иных обучающихся при промежуточной аттестации;

- присутствие в аудитории ассистента (ассистентов), оказывающего обучающимся инвалидам необходимую техническую помощь с учётом их индивидуальных особенностей (занять рабочее место, передвигаться, прочитать и оформить задание, общаться с экзаменатором);

- пользование необходимыми обучающимся инвалидам техническими средствами при прохождении промежуточной аттестации с учётом их индивидуальных особенностей;

- обеспечение возможности беспрепятственного доступа обучающихся инвалидов в аудитории, туалетные и другие помещения, а также их пребывания в указанных помещениях.

По письменному заявлению обучающегося инвалида продолжительность прохождения испытания промежуточной аттестации (зачёта, экзамена, и др.) обучающимся инвалидом может быть увеличена по отношению к установленной продолжительности его сдачи:

- продолжительность сдачи испытания, проводимого в письменной форме, — не более чем на 90 минут;

- продолжительность подготовки обучающегося к ответу, проводимому в устной форме, — не более чем на 20 минут.

В зависимости от индивидуальных особенностей обучающихся с ОВЗ Университет обеспечивает выполнение следующих требований при проведении аттестации:

а) для слепых:

- задания и иные материалы для прохождения промежуточной аттестации оформляются в виде электронного документа, доступного с помощью компьютера со специализированным программным обеспечением для слепых, либо зачитываются ассистентом;

- письменные задания выполняются обучающимися на бумаге рельефно-точечным шрифтом Брайля или на компьютере со специализированным программным обеспечением для слепых, либо надиктовываются ассистенту;

- при необходимости обучающимся предоставляется комплект письменных принадлежностей и бумага для письма рельефно-точечным шрифтом Брайля, компьютер со специализированным программным обеспечением для слепых;

б) для слабовидящих:

- задания и иные материалы для сдачи экзамена оформляются увеличенным шрифтом;

- обеспечивается индивидуальное равномерное освещение не менее 300 люкс;

- при необходимости обучающимся предоставляется увеличивающее устройство, допускается использование увеличивающих устройств, имеющихся у обучающихся;

в) для глухих и слабослышащих, с тяжелыми нарушениями речи:

- обеспечивается наличие звукоусиливающей аппаратуры коллективного пользования, при необходимости обучающимся предоставляется звукоусиливающая аппаратура индивидуального пользования;

- по их желанию испытания проводятся в письменной форме;

г) для лиц с нарушениями опорно-двигательного аппарата (тяжелыми нарушениями двигательных функций верхних конечностей или отсутствием верхних конечностей):

- письменные задания выполняются обучающимися на компьютере со специализированным программным обеспечением или надиктовываются ассистенту;

- по их желанию испытания проводятся в устной форме.

О необходимости обеспечения специальных условий для проведения аттестации обучающийся должен сообщить письменно не позднее, чем за 10 дней до начала аттестации. К заявлению прилагаются документы, подтверждающие наличие у обучающегося индивидуальных особенностей (при отсутствии указанных документов в организации).

При необходимости для обучающихся с инвалидностью процедура оценивания результатов обучения может проводиться в несколько этапов.