

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Романчук Иван Сергеевич
Должность: Ректор
Дата подписания: 22.04.2025 13:43:14
Уникальный программный ключ:
6319edc2b582fffdacea443f01d5779368d0957ac34f5cd074d81181530452479

МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ ПО ОРГАНИЗАЦИИ САМОСТОЯТЕЛЬНОЙ РАБОТЫ ОБУЧАЮЩИХСЯ

Наименование дисциплины	<i>Корпусная лингвистика / Corpora Linguistics</i>
Направление подготовки	<i>45.04.02 Лингвистика</i>
Направленность (профиль)	<i>Прикладная лингвистика</i>
Форма обучения	<i>очная</i>
Разработчик(и)	<i>Ковязина М.А., доцент кафедры прикладной и теоретической лингвистики</i>

1. Темы дисциплины для самостоятельного освоения обучающимися
Отсутствуют

2. План самостоятельной работы

№ п/п	Учебные встречи	Виды самостоятельной работы	Форма отчетности/ контроля	Количество баллов	Рекоменду емый бюджет времени на выполнен ие (ак.ч.)*
1	2	3	4	5	6
1	УВ №2, Практическое занятие 1, «Введение в корпусную лингвистику»	1. Прочитайте Главу 1 в учебном пособии А. А. Шунейко «Корпусная лингвистика» (2024) на образовательной платформе Юрайт (https://urait.ru/bcode/543746). 1.1. Ознакомьтесь с разделом 1.1. «Поиск имени» и ответьте на вопросы: а) На чём базируется методология корпусной лингвистики? б) С какими науками корпусная лингвистика тесно связана? в) Как на различных этапах своего развития называлась корпусная лингвистика? г) Каковы основные задачи корпусной лингвистики? 1.2. Ознакомьтесь с разделом 1.2. «Два типа учёных и познавательных моделей» и ответьте на вопросы: а) Можно ли корпусную лингвистику считать самостоятельным методом исследования? б) Какие теоретические и прикладные вопросы позволяет решать корпусная лингвистика? в) Какие существуют направления в корпусной лингвистике? 1.3. Ознакомьтесь с разделом 1.3. «Корпус и его окрестности» и ответьте на вопросы: а) Что послужило основой корпусной лингвистики в плане теории? б) Что и как стало основой корпусной лингвистики в плане технологии? в) Когда был дан самый мощный импульс развитию корпусной лингвистики? г) Какие научные направления и поиски сформировали корпусную лингвистику как направление научного знания?	Ключевые слова, ответы на поставленные вопросы	2	3
2	УВ №4,	Посмотрите видео академика РАН	Ключевые	1	4

	Практическое занятие 2, «Национальный корпус русского языка»	В. А. Плунгяна о корпусной лингвистике (https://www.youtube.com/watch?v=ico0zGpiBzc) и ответьте на вопросы: 1) Что такое корпус? 2) Что лингвист делает с корпусом? 3) Корпусы какого объема появились в современной лингвистике? 4) Почему лингвисты полюбили корпусы? 5) Когда появились первые корпусы? 6) Корпусы какого объема появились в 90-е гг.? 7) В каких неанглоязычных странах в 90-е гг. интенсивно развивалась корпусная лингвистика? 8) Какой проект предвосхитил создание Национального корпуса русского языка (НКРЯ)? 9) Когда НКРЯ был открыт для свободного доступа? 10) Кто составляет целевую аудиторию НКРЯ? 11) Каков объем НКРЯ? Какие тексты входят в НКРЯ? 12) Существует ли корпус устного русского языка?	тезисы, ответы на вопросы		
3	УВ №8, Практическое занятие 3, «Англоязычные корпусы»	1. Прочитайте Главу 1 в учебном пособии А. А. Шунейко «Корпусная лингвистика» (2024) на образовательной платформе Юрайт (https://urait.ru/bcode/543746). 1.1. Ознакомьтесь с разделом 1.5. «Типы корпусов» и ответьте на вопросы: а) Дайте развернутое определение лингвистического корпуса. С какой целью было введено это понятие? б) Какие типы лингвистических корпусов вам известны? Для чего они используются? в) Охарактеризуйте диапазон использования корпусов в современном мире. 1.2. Ознакомьтесь с разделами 1.7. «Когда и как родилась корпусная лингвистика» и 1.9. «Основные корпуса второго и третьего поколения». Ответьте на вопросы: а) Как называется корпус, в котором содержится только один тип языковых единиц? б) Как называется корпус, в котором представлено несколько языков? в) Когда и где был разработан первый языковой корпус? г) Какие возможности для лингвистики дало создание первого корпуса?	Ключевые тезисы, ответы на вопросы	1	3

		д) Какой корпус помимо письменной части включает разговорную часть, состоящую из неформальных бесед и разговорного языка?			
4	УВ №9, Практическое занятие 4, «Англоязычные корпусы»	Посмотрите видео разработчика корпусной платформы <i>English-corpora.org</i> Марка Дэвиса об основных инструментах и возможностях, предоставляемых <i>Корпусом современного американского английского языка (Corpus of Contemporary American English)</i> и другими корпусами на данном сайте: “English-Corpora.org: introduction” https://www.youtube.com/watch?v=Xu3Ry6es9Y4&list=PLaJHG9FXj8L6l8sOsYAHpSGzA-5hF-FEt . Подготовьте конспект о следующих аспектах: – элементы интерфейса платформы; – целевая аудитория платформы; – основные виды запросов; – инструменты Корпуса современного американского английского языка; – способы извлечения статистических данных на базе корпуса.	Текст конспекта	1	4
5	УВ №13, Практическое занятие 5, «Аннотирование корпуса»	1. Прочитайте раздел 1.4. «Разметка корпуса в контексте его создания» в учебном пособии А. А. Шунейко «Корпусная лингвистика» (2024) на образовательной платформе Юрайт (https://urait.ru/bcode/543746). Ответьте на вопросы: а) Что такое разметка или аннотирование текста? б) Для чего осуществляется разметка текста? в) Как осуществляется разметка текста? г) Зависит ли качество корпуса от качества и сложности разметки? Приведите примеры. 2. Посмотрите видео профессора Университета Васэда (Токио, Япония) Лоренса Энтони “TagAnt 1.0 - Getting started” (https://www.youtube.com/watch?v=uiWEbLfNkhI) о приложении, позволяющем осуществлять частеречную (или морфологическую) разметку текста на английском языке. Зафиксируйте в письменной форме основные операции, необходимые для получения морфологически размеченного текста.	Ответы на вопросы	1	4
6	УВ №14, Практическое занятие 6, «Корпусный	Посмотрите обучающие видео профессора Университета Васэда (Токио, Япония) Лоренса Энтони о корпусном менеджере AntConc:	Основные тезисы, конспект	1	4

	менеджер AntConc»	1) “AntConc 4 (ver. 4.2) - Getting started” https://www.youtube.com/watch?v=_GSlwIO5QZE 2) “AntConc 4 (ver. 4.2) - Corpus Manager Basics” https://www.youtube.com/watch?v=yDSa1rp8Bqs Подготовьте конспект об основных элементах интерфейса и инструментах данного приложения.			
7	УВ №17, Практическое занятие 7, «Корпусный менеджер AntConc»	Посмотрите обучающие видео профессора Университета Васэда (Токио, Япония) Лоренса Энтони о корпусном менеджере AntConc: 1) “AntConc 4 (ver. 4.0) - Word list tool basics” https://www.youtube.com/watch?v=MSl4WjM9Mw4 2) “AntConc 4 (ver. 4.0) - Keyword list tool basics” https://www.youtube.com/watch?v=SludW4FHatI 3) “AntConc 4 (ver. 4.0) - Collocate tool basics” https://www.youtube.com/watch?v=s0N-89xI23Y Подготовьте конспект об инструментах данного приложения, позволяющих получать данные о наиболее частотных единицах, ключевых словах и сочетаемости.	Основные тезисы, конспект	1	4
8	УВ №18, Практическое занятие 8, «Методология работы с корпус-менеджером»	1. Изучите материалы о корпусном менеджере #LancsBox X на сайте https://lancsbox.lancs.ac.uk/ (Университет Ланкастера, Великобритания). Подготовьте конспект об основных функциях данного приложения. 2. Посмотрите видео разработчика #LancsBox X Вацлава Брезины (Университет Ланкастера, Великобритания). Подготовьте конспект об инструментах и функционале приложения: а) “What is #LancsBox?” https://lancsbox.lancs.ac.uk/#video1 б) “Starting with #LancsBox” https://lancsbox.lancs.ac.uk/#video2 в) “KWIC” https://lancsbox.lancs.ac.uk/#video3	Основные тезисы, конспект	1	4
9	УВ №21-22, Практические занятия 9-10,	Посмотрите видео разработчика #LancsBox X Вацлава Брезины (Университет Ланкастера, Великобритания). Подготовьте	Основные тезисы, конспект	1	4

	«Методология работы с корпус-менеджером»	конспект об инструментах и функционале приложения: а) “GraphColl” https://lancsbox.lancs.ac.uk/#video4 б) “Words” https://lancsbox.lancs.ac.uk/#video5 в) “Text” https://lancsbox.lancs.ac.uk/#video6 г) “Advanced Queries” https://lancsbox.lancs.ac.uk/#video7			
10	УВ №25, Практическое занятие 11, «Создание собственного корпуса»	Ознакомьтесь с требованиями к корпусному проекту, который разрабатывается студентами индивидуально или в парах и выносится на защиту на завершающем занятии. Студентам предлагается разработать одноязычный или двуязычный корпус не менее 10 тысяч словоупотреблений и на его базе провести лингвистическое исследование.	Проект языкового корпуса	2	10
11	УВ №26, Практическое занятие 12, «Защита корпусных проектов»	Подготовка к защите проекта корпусного исследования: студентам (индивидуально или в парах) предлагается разработать одноязычный или двуязычный корпус не менее 10 тысяч словоупотреблений и на его базе провести лингвистическое исследование. Проект состоит из следующих обязательных модулей: - обоснование корпуса, описание источников материала; - выявление базовых параметров корпуса (соотношение types/tokens, относительная частота и коллокации, составление конкорданса для некоторых слов, ключевые слова); - описание какой-либо лингвистической особенности корпуса, которая вскрылась в процессе выполнения проекта.	Проект языкового корпуса: текст доклада (слайды презентации) и архивный файл корпуса zip или rar, включающий .docx или .pdf файл с названием корпуса и кратким описанием его параметров, а также собственно файл(ы) текста/текстов в формате *.txt.	20	10

3. Требования и рекомендации по выполнению самостоятельных работ обучающихся, критерии оценивания

УВ №25-26 Практическое занятие «Создание собственного корпуса» и практическое занятие 12 «Защита корпусных проектов»:

Студентам предлагается разработать одноязычный или двуязычный корпус не менее 10 тысяч словоупотреблений и на его базе провести лингвистическое исследование.

Проект состоит из следующих обязательных модулей:

- обоснование корпуса, описание источников материала;
- выявление базовых параметров корпуса (соотношение types/tokens, относительная частота и коллокации, составление конкорданса для некоторых слов, ключевые слова);

- описание какой-либо лингвистической особенности корпуса, которая вскрылась в процессе выполнения проекта.

Материалы проекта сдаются преподавателю в электронном виде: студент формирует архивный файл корпуса zip или rar, включающий .docx или .pdf файл с названием корпуса и кратким описанием его параметров, а также собственно файл(ы) текста/текстов в формате .txt, включенных в корпус. В названии архивного файла должна быть указана фамилия (или фамилии) исполнителя проекта; файл отправляется в соответствующее задание в онлайн-классе дисциплины (lхр).

Проект корпуса защищается на семинарском занятии. В ходе устной защиты студентам предлагается ответить на следующие вопросы:

1. Является ли корпус одноязычным / многоязычным (параллельным)?
2. Что послужило источником / источниками для корпуса?
 - укажите жанр(ы) текстов, включённых в корпус;
 - укажите ключевые слова (seed words), использованные при автоматическом формировании корпуса.
3. Цель создания корпуса?
4. Объем корпуса? (количество слов / токенов). Если корпус состоит из нескольких блоков (разные стили, хронологические периоды), можно указать процентное соотношение частей корпуса.
5. Список наиболее частотных единиц корпуса? (Antconc > Word List)
6. Список ключевых слов, выделенных на базе корпуса (Keywords)?
7. Коэффициент лексического разнообразия – types/tokens ratio (для корпуса в целом, либо для его частей).
8. Какая разметка применена в корпусе?
 - частеречная / POS – part of speech tagging, используемый набор меток (tagset)?
 - Free CLAWS web tagger <http://ucrel-api.lancaster.ac.uk/claws/free.html>
 - TagAnt <https://www.laurenceanthony.net/software/tagant/>
 - семантическая (USAS, English online semantic tagger <http://ucrel-api.lancaster.ac.uk/usas/tagger.html>)
9. Результаты лингвистического исследования, проведённого на базе полученного корпуса (исследование сочетаемости, частотности определённых единиц, выделение синонимов, применение CQL при анализе функционирования единиц и пр. (AntConc > Concordance / Clusters & N-Grams / Collocates и пр.).
10. Каковы перспективы применения полученного корпуса?

Студент может получить 20 баллов максимум за корпусный проект, проведённое на его базе исследование и успешную защиту проекта. Баллы распределяются следующим образом:

- формулирование цели проекта (1 балл);
- использование корпусного менеджера для создания корпуса и проведения анализа (2 балла);
- подсчитан объем корпуса, выявлены наиболее частотные единицы (2 балла);
- выделены ключевые слова корпуса (1 балл);
- кратко описана применённая в корпусе разметка (1 балл);

- проведено небольшое лингвистическое исследование на базе корпуса, продемонстрированы его результаты (3 балла);
- защитное слово (4 балла *max*);
- наличие иллюстративных материалов, слайдов (3 балла *max*);
- отчетность в электронном виде (3 балла).

4. Рекомендации по самоподготовке к промежуточной аттестации по дисциплине
Необходимое количество баллов для итоговой аттестации по дисциплине – 61-100.

Студенты, набравшие по итогам изучения дисциплины 61 балл и более, имеют право на аттестацию по итогам текущего контроля согласно балльно-рейтинговой шкале оценивания.

Студенты, не получившие необходимого количества баллов, сдают дифференцированный зачет, который представляет собой защиту корпусного проекта (требования описаны выше).