

Министерство науки и высшего образования Российской Федерации  
Федеральное государственное бюджетное образовательное учреждение  
высшего образования  
"Амурский государственный университет"

УТВЕРЖДАЮ

Проректор по учебной и научной  
работе

                    Лейфа                    А.В. Лейфа

« 1 » сентября 2023 г.

РАБОЧАЯ ПРОГРАММА  
«КОРПУСНАЯ ЛИНГВИСТИКА»

Направление подготовки 45.04.03 Фундаментальная и прикладная лингвистика

Направленность (профиль) образовательной программы – Иностранные языки и речевые  
технологии

Квалификация выпускника – Магистр

Год набора – 2023

Форма обучения – Очная

Курс     1     Семестр     2    

Экзамен 2 сем

Общая трудоемкость дисциплины 216.0 (академ. час), 6.00 (з.е)

Составитель О.Н. Морозова, доцент, д-р. филол. наук

Филологический факультет

Кафедра иностранных языков

Рабочая программа составлена на основании Федерального государственного образовательного стандарта ВО для направления подготовки 45.04.03 Фундаментальная и прикладная лингвистика, утвержденного приказом Министерства науки и высшего образования Российской Федерации от 12.08.20 № 993

Рабочая программа обсуждена на заседании кафедры иностранных языков

01.09.2023 г. , протокол № 1

Заведующий кафедрой Морозова О.Н. Морозова

СОГЛАСОВАНО

Учебно-методическое управление

Чалкина Н.А. Чалкина

« 1 » сентября 2023 г.

СОГЛАСОВАНО

Научная библиотека

Петрович О.В. Петрович

« 1 » сентября 2023 г.

СОГЛАСОВАНО

Выпускающая кафедра

Морозова О.Н. Морозова

« 1 » сентября 2023 г.

СОГЛАСОВАНО

Центр цифровой трансформации и  
технического обеспечения

Тодосейчук А.А. Тодосейчук

« 1 » сентября 2023 г.

## 1. ЦЕЛИ И ЗАДАЧИ ОСВОЕНИЯ ДИСЦИПЛИНЫ

### Цель дисциплины:

Ознакомление студентов магистратуры с различными методиками корпусной лингвистики при изучении языков и диалектов для последующих научных изысканий в области прикладной и экспериментальной лингвистики (в том числе: участие в создании электронных языковых ресурсов (корпусов текстов, электронных словарей, фонетических баз данных).

### Задачи дисциплины:

- формирование представления о картине мира в иноязычном обществе, существующих стереотипах и способах их преодоления;
- формирование у студентов глубоких знаний в области проблем, активно обсуждаемых в инокультурном обществе;
- научить студентов научным изысканиям в области прикладной лингвистики.

## 2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ

Дисциплина «Корпусная лингвистика» относится к группе факультативных дисциплин. Дисциплина разработана с учетом содержания цикла дисциплин соответствующего направления и может использоваться в составе комплекса лингвистических дисциплин для обучения магистров гуманитарных специальностей.

## 3. КОМПЕТЕНЦИИ ОБУЧАЮЩЕГОСЯ, ФОРМИРУЕМЫЕ В РЕЗУЛЬТАТЕ ОСВОЕНИЯ ДИСЦИПЛИНЫ И ИНДИКАТОРЫ ИХ ДОСТИЖЕНИЯ

### 3.1 Профессиональные компетенции и индикаторы их достижения

Код и наименование профессиональной компетенции	Код и наименование индикатора достижения профессиональной компетенции
ПК-7. Способен вести мониторинг информационных массивов и готовить на этой основе аналитические материалы	ИД-1 ПК-7 Знает: типы данных и баз данных; историю и становление корпусной лингвистики как научного направления; принципы работы с информационными массивами, способы из обработки и подготовки аналитических материалов. ИД-2 ПК-7 Умеет изучать речевую деятельность носителей языка, описывать новые явления и процессы в современном состоянии языка иноязычного социума; применять современные технологии сбора, обработки и интерпретации полученных экспериментальных данных. ИД-3 ПК-7 Владеет навыками исследовательской работы по анализу языка на базе информационных массивов, готовить на их основе аналитические материалы.

## 4. СТРУКТУРА ДИСЦИПЛИНЫ

Общая трудоемкость дисциплины составляет 6.00 зачетных единицы, 216.0 академических часов.

1 – № п/п

2 – Тема (раздел) дисциплины, курсовая работа (проект), промежуточная аттестация

3 – Семестр

4 – Виды контактной работы и трудоемкость (в академических часах)

4.1 – Л (Лекции)

4.2 – Лекции в виде практической подготовки

4.3 – ПЗ (Практические занятия)

4.4 – Практические занятия в виде практической подготовки

4.5 – ЛР (Лабораторные работы)

4.6 – Лабораторные работы в виде практической подготовки

4.7 – ИКР (Иная контактная работа)

4.8 – КТО (Контроль теоретического обучения)

4.9 – КЭ (Контроль на экзамене)

5 – Контроль (в академических часах)

6 – Самостоятельная работа (в академических часах)

7 – Формы текущего контроля успеваемости

1	2	3	4									5	6	7
			4.1	4.2	4.3	4.4	4.5	4.6	4.7	4.8	4.9			
1	Корпусная лингвистика: предмет, терминология. Связь корпусной лингвистики с др. науками.	2	6		4								6	Выполнение тестовых заданий. Устный опрос.
2	Методы корпусной лингвистики.	2	4		16								40	Выполнение тестовых заданий. Устный опрос.
3	Способы использования корпусов.	2	4		16		2						40	Презентация проекта.
4	Типы корпусов.	2	4		12								25.8	Выполнение тестовых заданий. Устный опрос.
5	Экзамен	2									0.3	35.7	0.2	
	Итого			18.0		48.0		2.0	0.0	0.0	0.3	35.7	112.0	

## 5. СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

### 5.1. Лекции

№ п/п	Наименование темы (раздела)	Содержание темы (раздела)
1	Корпусная лингвистика: предмет, терминология. Связь корпусной лингвистики с др. науками.	Корпус как поисковая система. Основные понятия и определения. Разметка. Репрезентативность. Метаданные. Корпусный менеджер. Treebank. Лемматизация. Конкорданс. Параллельный корпус. Лингвистические (языковые) и нелингвистические корпусы. История лингвистических корпусов: от картотеки к корпусу. Корпусная лингвистика: современное

		состояние. Корпусная лингвистика в России. Корпусоподобные интерфейсы между лингвистом и поисковыми системами Интернета. Лингвистические исследования, базирующиеся на корпусах.
2	Методы корпусной лингвистики.	Создание электронной хрестоматии по корпусной лингвистике. Исследование механизмов взаимодействия корпуса текстов и электронной картотеки (корпусы цитат). Создание веб-сайта по корпусной лингвистике. Предварительные работы по созданию корпуса. Проблемы репрезентативности. Отбор источников. Внешние и внутренние критерии отбора. Нормализация файлов. Графематический анализ. Средства создания и разметки корпусов. Понятие разметки. Типы разметки. Автоматический морфологический и синтаксический анализ. Металингвистическая разметка. Параллельные корпуса. Проблема выравнивания. Стандартизация в корпусной лингвистике. Языковые средства представления размеченных текстов. Международные стандарты и проекты (TEI, EAGLES, CDIF, XCES).
3	Способы использования корпусов.	Способы использования корпусов в лингвистических исследованиях. Исследование способов использования корпусов в лексикографии. Изучение средств обработки корпусных данных, представленных на языке XML.
4	Типы корпусов.	Обзор существующих корпусов различных типов. Классификация (типология) корпусов по различным основаниям. Типы корпусов по задачам. Типы корпусов по формальным признакам.

## 5.2. Практические занятия

Наименование темы	Содержание темы
Корпус как поисковая система.	Основные понятия и определения. Разметка. Репрезентативность. Метаданные. Корпусный менеджер. Treebank. Лемматизация. Конкорданс. Параллельный корпус. Лингвистические (языковые) и нелингвистические корпуса. История лингвистических корпусов: от картотеки к корпусу. Корпусная лингвистика: современное состояние. Корпусная лингвистика в России. Корпусоподобные интерфейсы между лингвистом и поисковыми системами Интернета. Лингвистические исследования, базирующиеся на корпусах.
Создание электронной хрестоматии по корпусной лингвистике.	Исследование механизмов взаимодействия корпуса текстов и электронной картотеки (корпусы цитат). Создание веб-сайта по корпусной лингвистике. Предварительные работы по созданию корпуса.

	Проблемы репрезентативности. Отбор источников. Внешние и внутренние критерии отбора. Нормализация файлов. Графематический анализ. Средства создания и разметки корпусов. Понятие разметки. Типы разметки. Автоматический морфологический и синтаксический анализ. Металингвистическая разметка. Параллельные корпусы. Проблема выравнивания. Стандартизация в корпусной лингвистике. Языковые средства представления размеченных текстов. Международные стандарты и проекты (TEI, EAGLES, CDIF, XCES).
Способы использования корпусов в лингвистических исследованиях.	Исследование способов использования корпусов в лексикографии. Изучение средств обработки корпусных данных, представленных на языке XML.
Обзор существующих корпусов различных типов.	Классификация (типология) корпусов по различным основаниям. Типы корпусов по задачам. Типы корпусов по формальным признакам.

### 5.3. Лабораторные занятия

Наименование темы	Содержание темы
Способы использования корпусов в лингвистических исследованиях.	Исследование способов использования корпусов в лексикографии. Изучение средств обработки корпусных данных, представленных на языке XML.

## 6. САМОСТОЯТЕЛЬНАЯ РАБОТА

№ п/п	Наименование темы (раздела)	Содержание темы (раздела)	Трудоемкость в академических часах
1	Корпусная лингвистика: предмет, терминология. Связь корпусной лингвистики с др. науками.	Индивидуальные задания	6
2	Методы корпусной лингвистики.	Индивидуальные задания	40
3	Способы использования корпусов.	Индивидуальные задания	40
4	Типы корпусов.	Индивидуальные задания	25.8
5	Экзамен	Подготовка по экзаменационным вопросам к экзамену	0.2

## 7. ОБРАЗОВАТЕЛЬНЫЕ ТЕХНОЛОГИИ

В ходе преподавания дисциплины лекции с традиционным и проблемным изложением теоретического материала, использование интерактивных обучающих мультимедиа средств; практические занятия с текущим устным опросом, интерактивные методы решения задач, использование наглядных средств, выполнение тестовых заданий, подготовка и презентация проекта.

## **8. ОЦЕНОЧНЫЕ СРЕДСТВА ДЛЯ ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ**

Описание показателей критериев оценивания компетенций на различных этапах их формирования, описание шкал оценивания, типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений и навыков, характеризующих этапы формирования компетенций, а также методические материалы, определяющие процедуры оценивания знаний, умений и навыков отражены в оценочных средствах по дисциплине «Корпусная лингвистика».

Целью всех форм контроля (текущего, промежуточного, итогового) является проверка (устная и письменная) уровня владения студентами изученным материалом. Студент сдает экзамен в II семестре на основании текущей успеваемости и результатов устного ответа и решение практической задачи на экзаменационном мероприятии.

Текущий контроль осуществляется в процессе выполнения индивидуальных заданий, указанных преподавателем к каждому практическому занятию, разбора и повторения основных понятий и методов, изложенных на лекциях. В качестве контроля освоения материала дисциплины предусмотрено выполнение тестовых заданий. Промежуточный контроль проводится в виде экзамена за весь курс обучения дисциплине.

Вопросы к экзамену по дисциплине «Корпусная лингвистика»

1. Предмет корпусной лингвистики
2. Полевая лингвистика и корпусная лингвистика
3. Связь корпусной лингвистики с этнолингвистикой
4. История лингвистических корпусов: от картотеки к корпусу.
5. Классификация (типология) корпусов.
6. Корпусная лингвистика: современное состояние.
7. Корпусная лингвистика в России.
8. Обзор существующих корпусов различных типов.
9. Корпус как поисковая система.
10. Корпусоподобные интерфейсы между лингвистом и поисковыми системами Интернета.
11. Лингвистические исследования, базирующиеся на корпусах.
12. Проблемы репрезентативности корпусов.
13. Проблемы хронологии в общезыковых корпусах.
14. Отбор текстов для корпусов.
15. Графематический анализ.
16. Понятие разметки.
17. Типы разметки.
18. Морфологическая разметка.
19. Синтаксические корпуса (treebanks).
20. Семантическая разметка.
21. Технология создания корпусов. Стадии работы.
22. Понятие корпусоида.
23. Автоматическая морфоразметка.
24. Автоматический синтаксический анализ (parsing).
25. Языковые средства представления размеченных текстов (языки SGML, XML).
26. Международные стандарты (TEI, EAGLES, CDIF, XCES).
27. Способы использования корпусов в лингвистических исследованиях.
28. Исследование способов использования корпусов в лексикографии.
29. Изучение средств обработки корпусных данных, представленных на языке XML.
30. Обзор существующих корпусов различных типов.
31. Классификация (типология) корпусов по различным основаниям.
32. Типы корпусов по задачам. Типы корпусов по формальным признакам.

## **9. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ**

а) литература

1. Захаров В.П. Корпусная лингвистика [Электронный ресурс]: учебник для студентов гуманитарных вузов/ Захаров В.П., Богданова С.Ю.— Электрон. текстовые данные.— Иркутск: Иркутский государственный лингвистический университет, 2011.— 161 с.— Режим доступа: <http://www.iprbookshop.ru/21088>. — Лицензия: весь срок охраны авторского права (дата обращения: 08.06.2023).

2. Федорова О.В. Экспериментальный анализ дискурса [Электронный ресурс]/ Федорова О.В.— Электрон. текстовые данные.— М.: Языки славянской культуры, 2014.— 511 с.— Режим доступа: <http://www.iprbookshop.ru/35742>. Лицензия: весь срок охраны авторского права (дата обращения: 08.06.2023).

3. Корпусная лингвистика [Электронный ресурс] : сб. учеб.-метод. материалов по дисц. для направления подготовки 45.04.03 «Фундаментальная и прикладная лингвистика» / АмГУ, ФФ, сост. О. Н. Морозова. – Благовещенск : Изд-во Амур. гос. ун-та, 2017. – Электронная библиотека АмГУ: [http://irbis.amursu.ru/DigitalLibrary/AmurSU\\_Edition/8303.pdf](http://irbis.amursu.ru/DigitalLibrary/AmurSU_Edition/8303.pdf) (дата обращения: 08.06.2023).

4. Морозова О. Н. Экспериментальные методы в лингвистике (на материале исчезающих языков Дальневосточного региона) [Электронный ресурс]: учеб. пособие / О. Н. Морозова, В. С. Гибалин, С. И. Бобров. - Благовещенск : Изд-во Амур. гос. ун-та, 2017. – 74 с. – Режим доступа: Электронная библиотека АмГУ: [http://irbis.amursu.ru/DigitalLibrary/AmurSU\\_Edition/7515.pdf](http://irbis.amursu.ru/DigitalLibrary/AmurSU_Edition/7515.pdf) (дата обращения: 08.06.2023).

б) программное обеспечение и Интернет-ресурсы

№	Наименование	Описание
1	Электронно-библиотечная система IPRbooks <a href="http://www.iprbookshop.ru/">http://www.iprbookshop.ru/</a>	Электронно-библиотечная система IPRbooks — научно-образовательный ресурс для решения задач обучения в России и за рубежом. Уникальная платформа ЭБС IPRbooks объединяет новейшие информационные технологии и учебную лицензионную литературу. Контент ЭБС IPRbooks отвечает требованиям стандартов высшей школы, СПО, дополнительного и дистанционного образования. ЭБС IPRbooks в полном объеме соответствует требованиям законодательства РФ в сфере образования

в) профессиональные базы данных и информационные справочные системы

№	Наименование	Описание
1	<a href="https://elibrary.ru/">https://elibrary.ru/</a>	Научная электронная библиотека eLIBRARY.RU - российский информационно-аналитический портал в области науки, технологии, медицины и образования
2	<a href="http://www.ruscorpora.ru">http://www.ruscorpora.ru</a>	Национальный корпус русского языка. Информационно-справочная система, основанная на собрании русских текстов в электронной форме
3	<a href="http://neicon.ru">http://neicon.ru</a>	Полнотекстовый архив ведущих западных научных журналов на российской платформе Национального электронно-информационного консорциума (НЭИКОН)
4	<a href="http://diss.rsl.ru/">http://diss.rsl.ru/</a>	Электронная библиотека диссертаций
5	<a href="http://www.gumer.info/bibliotek_Buks/Linguist/Index_Ling.php">http://www.gumer.info/bibliotek_Buks/Linguist/Index_Ling.php</a>	Библиотека Гумер - Языкознание. Лингвистика. Филология. Языкознание.
6	<a href="http://superlinguist.ru/">http://superlinguist.ru/</a>	Superlinguist – электронная научная библиотека, посвященная теоретическим и прикладным вопросам

#### **10. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ**

В учебном процессе используются учебные аудитории, лаборатория фонетики, компьютерный класс, лингафонный кабинет при кафедре ИЯ ФФ АмГУ для проведения лекционных и практических занятий с возможностью использования мультимедийных средств. Самостоятельная работа обучающихся осуществляется в помещениях, оснащенных компьютерной техникой с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду университета.