

Документ подписан простой электронной подписью

Информация о владельце:

ФИО: Ястребов Олег Александрович

Должность: Ректор

Дата подписания: 25.05.2026 09:50:57

Уникальный программный ключ:

ca953a0120d891083f939673078ef1a989dae18a

**Федеральное государственное автономное образовательное учреждение высшего образования**

**«Российский университет дружбы народов имени Патриса Лумумбы»**

**Институт русского языка**

(наименование основного учебного подразделения (ОУП) – разработчика ОП ВО)

## **РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ**

### **КОРПУСНАЯ ЛИНГВИСТИКА**

(наименование дисциплины/модуля)

**Рекомендована МССН для направления подготовки/специальности:**

#### **45.04.04 ИНТЕЛЛЕКТУАЛЬНЫЕ СИСТЕМЫ В ГУМАНИТАРНОЙ СРЕДЕ**

(код и наименование направления подготовки/специальности)

**Освоение дисциплины ведется в рамках реализации основной профессиональной образовательной программы высшего образования (ОП ВО):**

#### **ИНТЕЛЛЕКТУАЛЬНЫЕ ТЕХНОЛОГИИ И АНАЛИЗ ДАННЫХ В ГУМАНИТАРНОЙ СФЕРЕ**

(наименование (профиль/специализация) ОП ВО)

## 1. ЦЕЛЬ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Дисциплина «Корпусная лингвистика» входит в программу магистратуры «Интеллектуальные технологии и анализ данных в гуманитарной сфере» по направлению 45.04.04 «Интеллектуальные системы в гуманитарной среде» и изучается во 2 семестре 1 курса. Дисциплину реализует Кафедра русского языка 5. Дисциплина состоит из 3 разделов и 7 тем и направлена на изучение лингвистических корпусов.

Целью освоения дисциплины является формирование у обучающихся теоретических знаний в области создания и применения лингвистических корпусов, приобретение практических умений работы с корпусами и связанными с ними компьютерными инструментами в сфере цифровой филологии и цифровой педагогики.

## 2. ТРЕБОВАНИЯ К РЕЗУЛЬТАТАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Освоение дисциплины «Корпусная лингвистика» направлено на формирование у обучающихся следующих компетенций (части компетенций):

Таблица 2.1. Перечень компетенций, формируемых у обучающихся при освоении дисциплины (результаты освоения дисциплины)

Шифр	Компетенция	Индикаторы достижения компетенции (в рамках данной дисциплины)
УК-7	Способен: искать нужные источники информации и данные, воспринимать, анализировать, запоминать и передавать информацию с использованием цифровых средств, а также с помощью алгоритмов при работе с полученными из различных источников данными с целью эффективного использования полученной информации для решения задач; проводить оценку информации, ее достоверность, строить логические умозаключения на основании поступающих информации и данных	УК-7.1 Способен найти источники информации и данные, воспринимать, анализировать, запоминать и передавать информацию с использованием цифровых средств, а также с помощью алгоритмов при работе с полученными из различных источников данными с целью эффективного использования полученной информации для решения профессиональных задач;
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 Применяет средства интеллектуального анализа данных, машинного обучения, компьютерной лингвистики и представления знаний в гуманитарных областях знаний; ОПК-5.2 Использует новые информационные и интеллектуальные технологии в гуманитарных областях знаний;

## 3. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОП ВО

Дисциплина «Корпусная лингвистика» относится к обязательной части блока 1 «Дисциплины (модули)» образовательной программы высшего образования.

В рамках образовательной программы высшего образования обучающиеся также осваивают другие дисциплины и/или практики, способствующие достижению запланированных результатов освоения дисциплины «Корпусная лингвистика».

Таблица 3.1. Перечень компонентов ОП ВО, способствующих достижению запланированных результатов освоения дисциплины

Шифр	Наименование компетенции	Предшествующие дисциплины/модули, практики*	Последующие дисциплины/модули, практики*
УК-7	Способен: искать нужные источники информации и данные, воспринимать, анализировать, запоминать и	Введение в специальность. Цифровая гуманитаристика; Спецкурс по программированию на языке Python;	Компьютерная лингвистика**; Технологии и методы программирования;

Шифр	Наименование компетенции	Предшествующие дисциплины/модули, практики*	Последующие дисциплины/модули, практики*
	<p>передавать информацию с использованием цифровых средств, а также с помощью алгоритмов при работе с полученными из различных источников данными с целью эффективного использования полученной информации для решения задач; проводить оценку информации, ее достоверность, строить логические умозаключения на основании поступающих информации и данных</p>	<p>Технологии и методы программирования;</p>	<p>Технологическая (проектно-технологическая) практика;  Научно-исследовательская работа;  Преддипломная практика;</p>
ОПК-5	<p>Способен применять новые информационные технологии в гуманитарных областях знаний с использованием средств интеллектуального анализа данных, машинного обучения, компьютерной лингвистики и представления знаний</p>	<p>Математические методы в гуманитарных исследованиях;</p>	<p>Технологическая (проектно-технологическая) практика;  Научно-исследовательская работа;  Преддипломная практика;  Методы машинного обучения;  Методы распознавания образов;  Математические методы в гуманитарных исследованиях (дополнительные главы);  Базы данных в информационных системах;</p>

\* - заполняется в соответствии с матрицей компетенций и СУП ОП ВО

\*\* - элективные дисциплины /практики

#### 4. ОБЪЕМ ДИСЦИПЛИНЫ И ВИДЫ УЧЕБНОЙ РАБОТЫ

Общая трудоемкость дисциплины «Корпусная лингвистика» составляет «3» зачетные единицы.

Таблица 4.1. Виды учебной работы по периодам освоения образовательной программы высшего образования для очной формы обучения.

Вид учебной работы	ВСЕГО, ак.ч.		Семестр(-ы)
			2
Контактная работа, ак.ч	30		30
Лекции (ЛК)	15		15
Лабораторные работы (ЛР)	0		0
Практические/семинарские занятия (СЗ)	15		15
Самостоятельная работа обучающихся, ак.ч.	60		60
Контроль (экзамен/зачет с оценкой), ак.ч.	18		18
Общая трудоемкость дисциплины ак.ч.	ак.ч.	108	108
	зач.ед.	3	3

## 5. СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

Таблица 5.1. Содержание дисциплины (модуля) по видам учебной работы\*

Номер раздела	Наименование раздела дисциплины	Наименование темы		Содержание темы	Вид учебной работы*
Раздел 1	Корпусы и корпусная лингвистика в гуманитарной науке и практике	1.1	Лингвистический корпус: сущность и основные характеристики. Актуальные сферы применения корпусов.	Понятие лингвистического корпуса как структурированного электронного собрания текстов. Принципы репрезентативности, сбалансированности, разметки и машиночитаемости. История становления корпусной лингвистики как научного направления. Роль корпусов в современных гуманитарных исследованиях. Основные области применения: лексикография, грамматические исследования, дискурс-анализ, переводоведение, обучение языкам, автоматическая обработка текстов.	ЛК, СЗ
		1.2	Классификации лингвистических корпусов. Особые типы корпусов.	Основания классификации корпусов: по языку, объему, структуре, типу текстов, целям использования. Национальные, специализированные, параллельные, сопоставимые, диахронические, синхронические корпуса. Устные, мультимодальные и учебные корпуса. Веб-корпусы и динамически обновляемые корпуса. Особенности и области применения различных типов корпусов.	ЛК, СЗ
Раздел 2	Создание лингвистических корпусов	2.1	Технологии проектирования и разработки корпусов. Отбор источников и процедуры обработки естественного языка: токенизация, лемматизация, стемминг, парсинг.	Этапы создания корпуса: постановка целей и задач, проектирование структуры, отбор текстовых источников. Принципы формирования корпуса: репрезентативность, балансировка, метаданные. Основные процедуры автоматической обработки текста: токенизация, нормализация, лемматизация, стемминг, морфологический и синтаксический анализ (парсинг). Инструменты и программные средства для обработки текстов.	ЛК, СЗ
		2.2	Разметка: виды и средства. Лингвистическая и экстралингвистическая разметка. Стандартизация в корпусной лингвистике. Международные стандарты и проекты.	Понятие разметки и ее роль в корпусной лингвистике. Виды разметки: морфологическая, синтаксическая, семантическая, прагматическая. Экстралингвистическая разметка (метаданные: автор, дата, жанр, стиль и др.). Ручная и автоматическая разметка, методы контроля качества. Форматы представления данных (XML, TEI). Международные стандарты и инициативы (TEI, Universal Dependencies и др.). Интероперабельность и повторное использование корпусов.	ЛК, СЗ
Раздел 3	Использование лингвистических корпусов в цифровой филологии и цифровой педагогике	3.1	Конкордансер и корпусный менеджер как поисковая система. Языки запросов корпусных менеджеров. Корпусные менеджеры на базе веба. Зарубежные национальные корпуса. Современные корпуса русского языка: национальный корпус и устные корпуса русского языка.	Принципы поиска информации в корпусах. Конкорданс как инструмент анализа контекстов. Корпусные менеджеры и их функциональные возможности. Языки запросов (регулярные выражения, CQL и др.). Веб-ориентированные корпусные платформы. Обзор зарубежных национальных корпусов. Современные корпуса русского языка: национальный корпус русского языка, устные и специализированные корпуса.	ЛК, СЗ
		3.2	Компьютерные инструменты	Программные средства для анализа и обработки корпусных данных. Частотный анализ, коллокации, n-граммы, статистические методы. Инструменты визуализации данных.	ЛК, СЗ

Номер раздела	Наименование раздела дисциплины	Наименование темы		Содержание темы	Вид учебной работы*
			использования корпусов. Программы анализа, лингвистической обработки и преобразования текста. Лексикографические практики на базе корпусов. Грамматические исследования, основанные на корпусах.	Использование корпусов в лексикографии (создание словарей, словарных статей). Корпусно-ориентированные грамматические исследования. Применение корпусов в задачах NLP и анализа текстовых данных.	
		3.3	Лингвистические корпуса в изучении дискурса. Использование корпусов в обучении языкам.	Корпусные методы анализа различных типов дискурса (научного, медийного, разговорного и др.). Использование корпусов в преподавании родного и иностранных языков. Разработка учебных материалов на основе корпусных данных. Формирование языковой компетенции обучающихся с использованием корпусных технологий. Перспективы применения корпусов в цифровой педагогике.	ЛК, СЗ

\* - заполняется только по ОЧНОЙ форме обучения: ЛК – лекции; ЛР – лабораторные работы; СЗ – практические/семинарские занятия.

## 6. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Таблица 6.1. Материально-техническое обеспечение дисциплины

Тип аудитории	Оснащение аудитории	Специализированное учебное/лабораторное оборудование, ПО и материалы для освоения дисциплины (при необходимости)
Лекционная	Аудитория для проведения занятий лекционного типа, оснащенная комплектом специализированной мебели; доской (экраном) и техническими средствами мультимедиа презентаций.	
Семинарская	Аудитория для проведения занятий семинарского типа, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, оснащенная комплектом специализированной мебели и техническими средствами мультимедиа презентаций.	
Для самостоятельной работы	Аудитория для самостоятельной работы обучающихся (может использоваться для проведения семинарских занятий и консультаций), оснащенная комплектом специализированной мебели и компьютерами с доступом в ЭИОС.	

\* - аудитория для самостоятельной работы обучающихся указывается **ОБЯЗАТЕЛЬНО!**

## 7. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

### Основная литература:

1. Захаров В. П., Богданова С. Ю. Корпусная лингвистика: учебник. 3-е изд., перераб. — СПб.:Изд-во С.-Петербур. ун-та, 2024. — 234 с. ISBN 978-5-288- 05997-1
2. Моделирование в корпусной лингвистике: специализированные корпуса русского языка / В.П.Захаров, И.В.Азарова, О.А.Митрофанова, А.М.Попов, М.В.Хохлова; отв. ред. В.П.Захаров. — СПб.: Изд-во С.-Петербур. ун-та, 2023. — 208 с. ISBN 978-5-288-05902-5
3. Шунейко А.А. Корпусная лингвистика: учебник для вузов / А.А. Шунейко — Москва: Юрайт, 2025. — 222 с.

### Дополнительная литература:

1. Копотев М.В. Введение в корпусную лингвистику: Учебное пособие для студентов филологических и лингвистических специальностей университетов / М. В. Копотев — «Animedia», 2014.
2. McEnery T., Hardie A., Corpus Linguistics: Method, Theory and Practice. Cambridge University Press, 2012.
3. Корпусы и корпусная лингвистика. Основные понятия. — Текст : электронный // Myfilology.ru – информационный филологический ресурс : [сайт]. – URL: <https://myfilology.ru//177/korpusy-i-korpusnaya-lingvistika-osnovnye-ponyatiya/> (дата обращения: 12.12.2021)
4. Плунгян В. Корпусная лингвистика // ПостНаука [Электронный ресурс]. 2013. URL: <https://postnauka.ru/video/7783>
5. Перспективы: Корпусная лингвистика и корпус русского языка // ПостНаука. Режим доступа: <http://www.youtube.com/watch?v=OBTLuLx962U>
6. Публикации по Национальному корпусу русского языка. <http://www.ruscorpora.ru/corpora-biblio.html>
7. Biber D., Conrad S., Reppen R. Corpus Linguistics: Investigating Language Structure and Use. Cambridge University Press, 1998. <http://books.google.com/books?id=2h5F7TXa6psC>
8. Developing linguistic corpora: a guide to good practice. Edited by Martin Wynne. <http://www.ahds.ac.uk/creating/guides/linguistic-corpora/index.htm>

9. Kutuzov A. A lecture on corpus linguistics “Corpus Turn in Translation Studies”.  
<http://www.youtube.com/watch?v=VhNGfkP6JJ0>

10. McEnery T., Wilson A. Corpus Linguistics. Edinburgh University Press, 2001.  
[http://books.google.com/books?id=nwmgdvN\\_akAC](http://books.google.com/books?id=nwmgdvN_akAC)

11. Tymoczko M. Computerized Corpora and the Future of Translation Studies // Meta: journal des traducteurs / Meta: Translators' Journal, Vol. 43, No. 4, 1998, p. 652-660.

[http://tc.utmn.ru/files/Computerized%20corpora%20and%20the%20future%20of%20translation%20studies\\_%D0%A2%D0%B8%D0%BC%D0%BE%D1%88%D0%BA%D0%BE.pdf](http://tc.utmn.ru/files/Computerized%20corpora%20and%20the%20future%20of%20translation%20studies_%D0%A2%D0%B8%D0%BC%D0%BE%D1%88%D0%BA%D0%BE.pdf)

0translation%20studies\_%D0%A2%D0%B8%D0%BC%D0%BE%D1%88%D0%BA%D0%BE.pdf

*Ресурсы информационно-телекоммуникационной сети «Интернет»:*

1. ЭБС РУДН и сторонние ЭБС, к которым студенты университета имеют доступ на основании заключенных договоров

- Электронно-библиотечная система РУДН – ЭБС РУДН <http://lib.rudn.ru/MegaPro/Web>
- ЭБС «Университетская библиотека онлайн» <http://www.biblioclub.ru>
- ЭБС Юрайт <http://www.biblio-online.ru>
- ЭБС «Консультант студента» [www.studentlibrary.ru](http://www.studentlibrary.ru)
- ЭБС «Троицкий мост»

2. Базы данных и поисковые системы

- электронный фонд правовой и нормативно-технической документации <http://docs.cntd.ru/>
- поисковая система Яндекс <https://www.yandex.ru/>
- поисковая система Google <https://www.google.ru/>
- реферативная база данных SCOPUS <http://www.elsevier.com/locate/scopus/>

*Учебно-методические материалы для самостоятельной работы обучающихся при освоении дисциплины/модуля\*:*

1. Курс лекций по дисциплине «Корпусная лингвистика».

2. Презентационные материалы по дисциплине «Корпусная лингвистика»

\* - все учебно-методические материалы для самостоятельной работы обучающихся размещаются в соответствии с действующим порядком на странице дисциплины **в ТУИС!**

**РАЗРАБОТЧИКИ**

Доцент кафедры русского языка №5

---

Должность

**РУКОВОДИТЕЛЬ ОП ВО**

Заведующий кафедрой

---

Должность

**РУКОВОДИТЕЛЬ БУП**

Заведующий кафедрой

---

Должность

Козеренко Е.Б.

---

Фамилия И.О

Софронова Е.А.

---

Фамилия И.О

Куриленко В.Б.

---

Фамилия И.О