

МИНОБРНАУКИ РОССИИ



Федеральное государственное бюджетное образовательное учреждение  
высшего образования  
**«Российский государственный гуманитарный университет»**  
**(ФГБОУ ВО «РГГУ»)**

*ИНСТИТУТ ЛИНГВИСТИКИ*  
*УНЦ компьютерной лингвистики*

***Обработка языковых данных***

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ  
направление **45.03.03** **Фундаментальная и прикладная лингвистика**  
профиль **Фундаментальная и прикладная лингвистика**  
Уровень квалификации выпускника (*бакалавр*)

Форма обучения (*очная*)

РПД адаптирована для лиц  
с ограниченными возможностями  
здоровья и инвалидов

Москва 2019

*Технологии корпусной лингвистики*

Рабочая программа дисциплины

Составитель:

к. филол. н., доцент УНЦ компьютерной лингвистики А.Ч. Пиперски

Ответственный редактор:

д. филол. н., проф. Я.Г. Тестелец

УТВЕРЖДЕНО

Протокол заседания УНЦ компьютерной лингвистики

№ 1 от 28.08.2019

## **ОГЛАВЛЕНИЕ**

### **1. Пояснительная записка**

1.1 Цель и задачи дисциплины

1.2. Формируемые компетенции, соотнесённые с планируемыми результатами обучения по дисциплине

1.3. Место дисциплины в структуре образовательной программы

### **2. Структура дисциплины**

### **3. Содержание дисциплины**

### **4. Образовательные технологии**

### **5. Оценка планируемых результатов обучения**

5.1. Система оценивания

5.2. Критерии выставления оценки по дисциплине

5.3. Типовые задания, используемые для текущего контроля успеваемости и промежуточной аттестации обучающихся по дисциплине

5.3.1. Примеры заданий к семинарам

5.3.2. Примеры статей для докладов

### **6. Учебно-методическое и информационное обеспечение дисциплины**

6.1. Список источников и литературы

6.2. Перечень электронных ресурсов

### **7. Материально-техническое обеспечение дисциплины**

### **8. Обеспечение образовательного процесса для лиц с ограниченными возможностями здоровья и инвалидов**

### **9. Методические материалы**

9.1. Планы семинарских занятий

### **Приложения**

Приложение 1. Аннотация дисциплины

Приложение 2. Лист изменений

## 1. Пояснительная записка

### 1.1. Цель и задачи дисциплины

**Цель дисциплины** – освоение студентами базовых понятий обработки языковых данных и математической статистики в применении к лингвистике.

**Задачи дисциплины:**

Курс нацелен на **формирование** у студентов следующих профессиональных **компетенций:**

- владение основными методами инструментального анализа звучащей речи (ПК-2);
- владение основными способами описания и формальной репрезентации денотативной, концептуальной, коммуникативной и прагматической информации, содержащейся в тексте на естественном языке (ПК-5)
- способность пользоваться лингвистически ориентированными программными продуктами (ПК-10)
- владение принципами создания электронных языковых ресурсов (текстовых, речевых и мультимодальных корпусов; словарей, тезаурусов, онтологий; фонетических, лексических, грамматических и иных баз данных и баз знаний) и умением пользоваться такими ресурсами (ПК-11)
- способность использовать лингвистические технологии для проектирования систем автоматической обработки звучащей речи и письменного текста на естественном языке, лингвистических компонентов интеллектуальных и информационных электронных систем (ПК-12)

### 1.2. Перечень планируемых результатов обучения по дисциплине, соотнесенных с индикаторами достижения компетенций:

В результате освоения дисциплины обучающийся должен:

**знать:**

- основные понятия и методы математической статистики
- устройство лингвистических ресурсов для извлечения необходимых данных

**уметь:**

- анализировать различные уровни языковой структуры с помощью экспериментальных и корпусных методов
- обрабатывать полученные результаты с помощью статистических программ

**владеть:**

- современной терминологией лингвистической документации
- навыками программирования и вычислительной работы.

<b>Компетенция</b> (код и наименование)	<b>Индикаторы компетенций</b> (код и наименование)	<b>Результаты обучения</b>
--	---	----------------------------

ПК-2. Владеет основными методами инструментального анализа звучащей речи	ПК-2.1	Знает: основные признаки и параметры вариативности звучания речи; основные достижения современной науки и технологий в рамках автоматического и экспертного анализа речевого сигнала.
	ПК-2.2	Умеет: анализировать вербальные и невербальные компоненты речевой деятельности; пользоваться методами и инструментарием лингвистического анализа звучащей речи; проводить базовую сегментацию и аннотацию звучащей речи, спектральный анализ.
ПК-5. Владеет основными способами описания и формальной репрезентации денотативной, концептуальной, коммуникативной и прагматической информации, содержащейся в тексте на естественном языке	ПК-5.2	Умеет: обрабатывать и анализировать текст на естественном языке; представлять результаты анализа в формализованном виде; производить компонентный анализ языковых данных.
	ПК-5.3	Имеет практический опыт обработки языкового материала с целью установления различных компонентов его значения с последующей формализацией полученных данных
ПК-10. Способен пользоваться лингвистически ориентированными программными продуктами	ПК-10.1	Знает: основные типы систем, использующих модули лингвистического анализа; основные принципы и методы компьютерного моделирования лингвистических задач.
ПК-11. Владеет принципами создания электронных языковых ресурсов (текстовых, речевых и мультимодальных корпусов; словарей, тезаурусов, онтологий; фонетических, лексических, грамматических и иных баз данных и баз знаний) и умением пользоваться такими ресурсами	ПК-11.1	Знает: основные принципы обработки информации; базовые принципы корпусной лингвистики, лексикографии, математической статистики; базовые представления о языковом разнообразии; наиболее полные и значимые лингвистические корпуса, электронные словари и базы данных.
	ПК-11.2	Умеет: пользоваться основными методами, способами и средствами получения, хранения, переработки информации; пользоваться лингвистически ориентированными программными

		продуктами
	ПК-11.3	Имеет практический опыт разработки электронных языковых ресурсов; опыт применения основных методов, способов и средств получения, хранения, переработки информации
ПК-12. Способен использовать лингвистические технологии для проектирования систем автоматической обработки звучащей речи и письменного текста на естественном языке, лингвистических компонентов интеллектуальных и информационных электронных систем	ПК-12.1	Знает: основные системы автоматической обработки звучащей речи и текстов на естественном языке; базовые принципы автоматической обработки языковых данных; основные интеллектуальные и информационные электронные системы и принципы работы с ними.
	ПК-12.2	Умеет: пользоваться существующими системами автоматической обработки текста и звучащей речи, интеллектуальными и информационными электронными системами; проводить их сравнительный анализ; проектировать модули данных систем, составлять технические задания.

### 1.3. Место дисциплины в структуре образовательной программы

Курс «Обработка языковых данных» является факультативом в учебном плане по направлению подготовки 45.03.03 «Фундаментальная и прикладная лингвистика». Дисциплина адресована бакалаврам, обучающимся по направлению «Фундаментальная и прикладная лингвистика» (ФТД.02) без направленности (профиля). Курс читается на 4-м курсе в 1-м семестре УНЦ компьютерной лингвистики ИЛ РГГУ, форма промежуточного контроля – зачет.

Для успешного освоения материала студент должен опираться на знания, умения и навыки, полученные в рамках курсов «Введение в теорию языка», «Общая фонетика», «Общая морфология», «Общий синтаксис», «Общая семантика», «Понятийный аппарат математики», «Введение в компьютерную лингвистику», «Основы программирования» а также на навыки изучения научной литературы, сформированные при подготовке к другим теоретическим курсам.

Курс направлен на углубление знаний усовершенствование умений и навыков студентов в сфере общей лингвистики и созданию у студентов представления об основных методах и задачах математической статистики в лингвистике.

## 2. Структура дисциплины

Общая трудоёмкость дисциплины составляет 3 з.е., 108 ч., в том числе контактная работа обучающихся с преподавателем 42 ч., самостоятельная работа обучающихся 66 ч.

### Тематический календарный план курса

№ № раз дел а	Раздел курса	Семестр	Виды учебной деятельности и трудоёмкость (в часах)			Форма контроля успеваемости
			лекции	семинары	СРС	
1	Лингвистика как эмпирическая наука. Дизайн эксперимента	1	4		12	Контроль посещаемости студентов.
2	Описательная статистика	1	4	2	16	Контроль посещаемости студентов. Обсуждение прочитанной научной литературы
3	Комбинаторика и теория вероятностей для нужд статистики	1	4		12	Контроль посещаемости студентов. Обсуждение прочитанной научной литературы.
4	Обобщение данных о выборке на генеральную совокупность. Проверка статистических гипотез	1	6	2	12	Контроль посещаемости студентов. Обсуждение прочитанной научной литературы.
5	Обсуждение и анализ данных, собранных учащимися в рамках подготовки ВКР	1		20	10	Контроль посещаемости студентов. Обсуждение докладов.
6	Зачёт	1			4	Итоговая контрольная работа.
	<b>Итого:</b>		<b>18</b>	<b>24</b>	<b>66</b>	

### 3. Содержание дисциплины

№	Наименование раздела дисциплины	Содержание
1	Лингвистика как эмпирическая наука. Дизайн эксперимента	Фальсификационизм. Операционализация лингвистических гипотез. Выделение лингвистических переменных.
2	Описательная статистика	Типы переменных: количественные и категориальные переменные. Способы описания и визуализации распределения значений переменных различных типов. Различные виды средних значений. Построение и понимание диаграмм размаха.
3	Комбинаторика и теория вероятностей для нужд статистики	Понятие статистической значимости и размера эффекта. Вычисление вероятностей наблюдаемых распределений в предположении о независимости переменных.
4	Обобщение данных о выборке на генеральную совокупность. Проверка статистических гипотез	Понятие нулевой и альтернативной гипотезы. Пороги значимости. Основные статистические критерии: критерий хи-квадрат, t-критерий, критерий Уилкоксона — Манна — Уитни, дисперсионный анализ. Линейная и логистическая регрессия.
5	Обсуждение и анализ данных, собранных обучающимися в рамках подготовки ВКР	Практическое применение полученных знаний к обработке реальных лингвистических данных.

#### 4. Образовательные технологии

№ п/п	Наименование раздела	Виды учебных занятий	Образовательные технологии
1	Лингвистика как эмпирическая наука. Дизайн эксперимента	Лекция 1	Вводная лекция с использованием презентации
		Лекция 2	Лекция с использованием презентации
2	Описательная статистика	Лекция 3	Лекция с использованием презентации
		Лекция 4	Лекция с использованием презентации
		Семинар 1	Совместное выполнение заданий по применению методов математической статистики к языковым данным.
3	Комбинаторика и теория вероятностей для нужд статистики	Лекция 5	Лекция с использованием презентации
		Лекция 6	Лекция с использованием презентации
4	Обобщение данных о выборке на генеральную совокупность. Проверка статистических гипотез	Лекция 7	Лекция с использованием презентации
		Лекция 8	Лекция с использованием презентации
		Лекция 9	Лекция с использованием презентации
		Семинар 2	Совместное выполнение заданий по применению методов математической статистики к языковым данным.
5	Обсуждение и анализ данных, собранных обучающимися в рамках подготовки ВКР	Семинар 3	Доклады обучающихся по материалам, собранных в рамках подготовки ВКР. Обсуждение докладов. Совместная обработка данных.
		Семинар 4	Доклады обучающихся по материалам, собранных в рамках подготовки ВКР. Обсуждение докладов. Совместная обработка данных.
		Семинар 5	Доклады обучающихся по материалам, собранных в рамках подготовки ВКР. Обсуждение докладов. Совместная обработка данных.
		Семинар 6	Доклады обучающихся по материалам, собранных в рамках подготовки ВКР. Обсуждение докладов. Совместная обработка данных.
		Семинар 7	Доклады обучающихся по материалам, собранных в рамках подготовки ВКР. Обсуждение докладов. Совместная обработка данных.
		Семинар 8	Доклады обучающихся по материалам, собранных в рамках подготовки ВКР. Обсуждение докладов. Совместная

		обработка данных.
Семинар 9		Доклады обучающихся по материалам, собранных в рамках подготовки ВКР. Обсуждение докладов. Совместная обработка данных.
Семинар 10		Доклады обучающихся по материалам, собранных в рамках подготовки ВКР. Обсуждение докладов. Совместная обработка данных.
Семинар 11		Доклады обучающихся по материалам, собранных в рамках подготовки ВКР. Обсуждение докладов. Совместная обработка данных.
Семинар 12		Доклады обучающихся по материалам, собранных в рамках подготовки ВКР. Обсуждение докладов. Совместная обработка данных.

## 5. Оценка планируемых результатов обучения

### 5.1. Система оценивания

Форма контроля	Макс. количество баллов	
	За одну работу	Всего
Текущий контроль:		
• Практические упражнения и задания, выполняемые на занятиях и в качестве домашней работы.	5 баллов	10 баллов
• Доклады обучающихся	5 баллов	20 баллов
• Активное участие в обсуждении докладов	5 баллов	50
Промежуточная аттестация (зачёт)	20 баллов	20 баллов
<b>Итого за дисциплину</b>		<b>100 баллов</b>

Полученный совокупный результат конвертируется в традиционную шкалу оценок и в шкалу оценок Европейской системы переноса и накопления кредитов (European Credit Transfer System; далее – ECTS) в соответствии с таблицей:

100-балльная шкала	Традиционная шкала	Шкала ECTS
95 – 100	отлично	A

83 – 94		зачтено	B
68 – 82	хорошо		C
56 – 67	удовлетворительно		D
50 – 55			E
20 – 49	неудовлетворительно	не зачтено	FX
0 – 19			F

## 5.2. Критерии выставления оценки по дисциплине

Баллы/ Шкала ECTS	Оценка по дисциплине	Критерии оценки результатов обучения по дисциплине
100-83/ А,В	«отлично»/ «зачтено (отлично)»/ «зачтено»	<p>Выставляется обучающемуся, если он глубоко и прочно усвоил теоретический и практический материал, может продемонстрировать это на занятиях и в ходе промежуточной аттестации.</p> <p>Обучающийся исчерпывающе и логически стройно излагает учебный материал, умеет увязывать теорию с практикой, справляется с решением задач профессиональной направленности высокого уровня сложности, правильно обосновывает принятые решения.</p> <p>Свободно ориентируется в учебной и профессиональной литературе.</p> <p>Оценка по дисциплине выставляется обучающемуся с учётом результатов текущей и промежуточной аттестации.</p> <p>Компетенции, закреплённые за дисциплиной, сформированы на уровне – «высокий».</p>
82-68/ С	«хорошо»/ «зачтено (хорошо)»/ «зачтено»	<p>Выставляется обучающемуся, если он знает теоретический и практический материал, грамотно и по существу излагает его на занятиях и в ходе промежуточной аттестации, не допуская существенных неточностей.</p> <p>Обучающийся правильно применяет теоретические положения при решении практических задач профессиональной направленности разного уровня сложности, владеет необходимыми для этого навыками и приёмами.</p> <p>Достаточно хорошо ориентируется в учебной и профессиональной литературе.</p> <p>Оценка по дисциплине выставляется обучающемуся с учётом результатов текущей и промежуточной аттестации.</p> <p>Компетенции, закреплённые за дисциплиной, сформированы на уровне – «хороший».</p>
67-50/ D,E	«удовлетвори- тельно»/ «зачтено (удовлетвори- тельно)»/	<p>Выставляется обучающемуся, если он знает на базовом уровне теоретический и практический материал, допускает отдельные ошибки при его изложении на занятиях и в ходе промежуточной аттестации.</p> <p>Обучающийся испытывает определённые затруднения</p>

Баллы/ Шкала ECTS	Оценка по дисциплине	Критерии оценки результатов обучения по дисциплине
	«зачтено»	<p>в применении теоретических положений при решении практических задач профессиональной направленности стандартного уровня сложности, владеет необходимыми для этого базовыми навыками и приёмами.</p> <p>Демонстрирует достаточный уровень знания учебной литературы по дисциплине.</p> <p>Оценка по дисциплине выставляются обучающемуся с учётом результатов текущей и промежуточной аттестации.</p> <p>Компетенции, закреплённые за дисциплиной, сформированы на уровне – «достаточный».</p>
49-0/ F,FX	«неудовлетворительно»/ не зачтено	<p>Выставляется обучающемуся, если он не знает на базовом уровне теоретический и практический материал, допускает грубые ошибки при его изложении на занятиях и в ходе промежуточной аттестации.</p> <p>Обучающийся испытывает серьёзные затруднения в применении теоретических положений при решении практических задач профессиональной направленности стандартного уровня сложности, не владеет необходимыми для этого навыками и приёмами.</p> <p>Демонстрирует фрагментарные знания учебной литературы по дисциплине.</p> <p>Оценка по дисциплине выставляются обучающемуся с учётом результатов текущей и промежуточной аттестации.</p> <p>Компетенции на уровне «достаточный», закреплённые за дисциплиной, не сформированы.</p>

### 5.3. Типовые задания, используемые для текущего контроля успеваемости и промежуточной аттестации обучающихся по дисциплине.

#### 5.3.1. Примеры заданий к семинарам

1. Дан набор значений переменной: 11, 13, 18, 11, 9, 4, 5, 1, 25, 29, 42, 32, 32. Определить меры центральной тенденции (моду, медиану, среднее арифметическое) и нарисовать диаграмму размаха.

2. С помощью Национального корпуса русского языка оценить, какая из конструкций «только и знает что» / «только и делает что» чаще встречается в настоящем времени, а какая — в прошедшем; оценить статистическую значимость и размер эффекта.

#### 5.3.2 Примеры статей для докладов

Доклады делаются на материале тем ВКР обучающихся.

## **6. Учебно-методическое и информационное обеспечение дисциплины**

### **6.1. Список источников и литературы**

#### **Основной учебник**

Николаев И. С., Митренина О. В., Ландо Т. М. (ред.). 2016. Прикладная и компьютерная лингвистика. М.: URSS.

### **6.2. Перечень электронных ресурсов**

Национальный корпус русского языка  
ruscorpora.ru

## **7. Материально-техническое обеспечение дисциплины**

Лекционные занятия по курсу «Обработка языковых данных» проводятся с использованием компьютерных презентаций, поэтому в аудитории необходимы компьютер и проектор, а также соответствующее освещение. В ходе семинарских занятий студенты должны быть обеспечены компьютерами с лицензионным программным обеспечением и выходом в Интернет, т.е. занятия должны проходить в компьютерных классах. Для эффективной работы также необходима доска, размер и расположение которой должны позволять всем слушателям видеть написанное.

## **8. Обеспечение образовательного процесса для лиц с ограниченными возможностями здоровья и инвалидов**

В ходе реализации дисциплины используются следующие дополнительные методы обучения, текущего контроля успеваемости и промежуточной аттестации обучающихся в зависимости от их индивидуальных особенностей:

- для слепых и слабовидящих:
  - лекции оформляются в виде электронного документа, доступного с помощью компьютера со специализированным программным обеспечением;
  - письменные задания выполняются на компьютере со специализированным программным обеспечением, или могут быть заменены устным ответом;
  - обеспечивается индивидуальное равномерное освещение не менее 300 люкс;
  - для выполнения задания при необходимости предоставляется увеличивающее устройство; возможно также использование собственных увеличивающих устройств;
  - письменные задания оформляются увеличенным шрифтом;
  - экзамен и зачёт проводятся в устной форме или выполняются в письменной форме на компьютере.
- для глухих и слабослышащих:
  - лекции оформляются в виде электронного документа, либо предоставляется звукоусиливающая аппаратура индивидуального пользования;
  - письменные задания выполняются на компьютере в письменной форме;

- экзамен и зачёт проводятся в письменной форме на компьютере; возможно проведение в форме тестирования.

- для лиц с нарушениями опорно-двигательного аппарата:
  - лекции оформляются в виде электронного документа, доступного с помощью компьютера со специализированным программным обеспечением;
  - письменные задания выполняются на компьютере со специализированным программным обеспечением;
- экзамен и зачёт проводятся в устной форме или выполняются в письменной форме на компьютере.

При необходимости предусматривается увеличение времени для подготовки ответа.

Процедура проведения промежуточной аттестации для обучающихся устанавливается с учётом их индивидуальных психофизических особенностей. Промежуточная аттестация может проводиться в несколько этапов.

При проведении процедуры оценивания результатов обучения предусматривается использование технических средств, необходимых в связи с индивидуальными особенностями обучающихся. Эти средства могут быть предоставлены университетом, или могут использоваться собственные технические средства.

Проведение процедуры оценивания результатов обучения допускается с использованием дистанционных образовательных технологий.

Обеспечивается доступ к информационным и библиографическим ресурсам в сети Интернет для каждого обучающегося в формах, адаптированных к ограничениям их здоровья и восприятия информации:

- для слепых и слабовидящих:
  - в печатной форме увеличенным шрифтом;
  - в форме электронного документа;
  - в форме аудиофайла.
- для глухих и слабослышащих:
  - в печатной форме;
  - в форме электронного документа.
- для обучающихся с нарушениями опорно-двигательного аппарата:
  - в печатной форме;
  - в форме электронного документа;
  - в форме аудиофайла.

Учебные аудитории для всех видов контактной и самостоятельной работы, научная библиотека и иные помещения для обучения оснащены специальным оборудованием и учебными местами с техническими средствами обучения:

- для слепых и слабовидящих:
  - устройством для сканирования и чтения с камерой SARA CE;
  - дисплеем Брайля PAC Mate 20;
  - принтером Брайля EmBraille ViewPlus;
- для глухих и слабослышащих:
  - автоматизированным рабочим местом для людей с нарушением слуха и слабослышащих;
  - акустический усилитель и колонки;
- для обучающихся с нарушениями опорно-двигательного аппарата:

- передвижными, регулируемые эргономическими партами СИ-1;
- компьютерной техникой со специальным программным обеспечением.

## **9. Методические материалы**

### **9.1. Планы семинарских занятий**

#### **Семинар 1. Описательная статистика (2 ч.).**

##### Вопросы для обсуждения:

Типы переменных: количественные и категориальные переменные. Способы описания и визуализации распределения значений переменных различных типов. Различные виды средних значений. Построение и понимание диаграмм размаха. Выбор переменных в соответствии с задачами исследования.

##### Список литературы:

- Николаев И. С., Митренина О. В., Ландо Т.М. (ред.). 2016. *Прикладная и компьютерная лингвистика*. М.: URSS.
- Brezina, Vaclav. 2018. *Statistics in corpus linguistics: a practical guide*. Cambridge: Cambridge.

##### Материально-техническое обеспечение занятия:

В ходе лабораторных работ студенты должны быть обеспечены компьютерами с лицензионным программным обеспечением и выходом в Интернет.

#### **Семинар 2. Обобщение данных о выборке на генеральную совокупность. Проверка статистических гипотез (2 ч.).**

##### Вопросы для обсуждения:

Понятие нулевой и альтернативной гипотезы. Пороги значимости. Основные статистические критерии: критерий хи-квадрат, t-критерий, критерий Уилкоксона — Манна — Уитни, дисперсионный анализ. Линейная и логистическая регрессия. Интерпретация результатов, полученных с помощью статистических методов.

##### Список литературы:

- Николаев И. С., Митренина О. В., Ландо Т.М. (ред.). 2016. *Прикладная и компьютерная лингвистика*. М.: URSS.
- Brezina, Vaclav. 2018. *Statistics in corpus linguistics: a practical guide*. Cambridge: Cambridge.

##### Материально-техническое обеспечение занятия:

В ходе лабораторных работ студенты должны быть обеспечены компьютерами с лицензионным программным обеспечением и выходом в Интернет.

**Семинары 3–12. Обсуждение и анализ данных, собранных учащимися в рамках подготовки ВКР (20 ч.).**

Вопросы для обсуждения:

Типы данных, собранных обучающимися. Практическое применение полученных знаний к обработке реальных лингвистических данных.

Список литературы:

- Николаев И. С., Митренина О. В., Ландо Т.М. (ред.). 2016. *Прикладная и компьютерная лингвистика*. М.: URSS.
- McEnery, Tony & Andrew Hardie. 2012. *Corpus linguistics: Method, theory and practice*. Cambridge & New York: Cambridge University Press.

Материально-техническое обеспечение занятия:

В ходе лабораторных работ студенты должны быть обеспечены компьютерами с лицензионным программным обеспечением и выходом в Интернет.

## АННОТАЦИЯ ДИСЦИПЛИНЫ

Курс «Обработка языковых данных» является факультативом в учебном плане по направлению подготовки 45.03.03 «Фундаментальная и прикладная лингвистика». Дисциплина адресована бакалаврам, обучающимся по направлению «Фундаментальная и прикладная лингвистика» (ФТД.02) без направленности (профиля). Курс читается на 4-м курсе в 1-м семестре УНЦ компьютерной лингвистики ИЛ РГГУ, форма промежуточного контроля – зачет.

Дисциплина реализуется в учебно-научном центре компьютерной лингвистики.

**Цель дисциплины** – освоение студентами базовых понятий обработки языковых данных и математической статистики в применении к лингвистике.

### **Задачи дисциплины:**

Курс нацелен на **формирование** у студентов следующих профессиональных **компетенций**:

- владение основными методами инструментального анализа звучащей речи (ПК-2);
- владение основными способами описания и формальной репрезентации денотативной, концептуальной, коммуникативной и прагматической информации, содержащейся в тексте на естественном языке (ПК-5)
- способность пользоваться лингвистически ориентированными программными продуктами (ПК-10)
- владение принципами создания электронных языковых ресурсов (текстовых, речевых и мультимодальных корпусов; словарей, тезаурусов, онтологий; фонетических, лексических, грамматических и иных баз данных и баз знаний) и умением пользоваться такими ресурсами (ПК-11)
- способность использовать лингвистические технологии для проектирования систем автоматической обработки звучащей речи и письменного текста на естественном языке, лингвистических компонентов интеллектуальных и информационных электронных систем (ПК-12)

В результате освоения дисциплины обучающийся должен:

### **знать:**

- основные понятия и методы математической статистики
- устройство лингвистических ресурсов для извлечения необходимых данных

### **уметь:**

- анализировать различные уровни языковой структуры с помощью экспериментальных и корпусных методов
- обрабатывать полученные результаты с помощью статистических программ

### **владеть:**

- современной терминологией лингвистической документации

- навыками программирования и вычислительной работы.

По дисциплине предусмотрена промежуточная аттестация в форме зачета.

Общая трудоемкость освоения дисциплины составляет 2 зачетные единицы.

**ЛИСТ ИЗМЕНЕНИЙ**

№	Текст актуализации или прилагаемый к РПД документ, содержащий изменения	Дата	№ протокола
1	Приложение №1	25.06.2020	4

### 1. Образовательные технологии (к п.4 на 2020 г.)

В период временного приостановления посещения обучающимися помещений и территории РГГУ. для организации учебного процесса с применением электронного обучения и дистанционных образовательных технологий могут быть использованы следующие образовательные технологии:

- видео-лекции;
- онлайн-лекции в режиме реального времени;
- электронные учебники, учебные пособия, научные издания в электронном виде и доступ к иным электронным образовательным ресурсам;
- системы для электронного тестирования;
- консультации с использованием телекоммуникационных средств.

### 2. Перечень БД и ИСС (к п. 6.2 на 2020 г.)

№п/п	Наименование
1	Международные реферативные наукометрические БД, доступные в рамках национальной подписки в 2020 г. Web of Science Scopus
2	Профессиональные полнотекстовые БД, доступные в рамках национальной подписки в 2020 г. Журналы Cambridge University Press ProQuest Dissertation & Theses Global SAGE Journals Журналы Taylor and Francis
3	Профессиональные полнотекстовые БД JSTOR Издания по общественным и гуманитарным наукам Электронная библиотека Grebennikon.ru
4	Компьютерные справочные правовые системы Консультант Плюс, Гарант

### 3. Состав программного обеспечения (ПО) (к п. 7 на 2020 г.)

№п/п	Наименование ПО	Производитель	Способ распространения ( <i>лицензионное или свободно распространяемое</i> )
1	Adobe Master Collection CS4	Adobe	лицензионное
2	Microsoft Office 2010	Microsoft	лицензионное
3	Windows 7 Pro	Microsoft	лицензионное
4	AutoCAD 2010 Student	Autodesk	свободно распространяемое
5	Archicad 21 Rus Student	Graphisoft	свободно распространяемое
6	SPSS Statistics 22	IBM	лицензионное
7	Microsoft Share	Microsoft	лицензионное

	Point 2010		
8	SPSS Statistics 25	IBM	лицензионное
9	Microsoft Office 2013	Microsoft	лицензионное
10	ОС «Алът Образование» 8	ООО «Базальт СПО	лицензионное
11	Microsoft Office 2013	Microsoft	лицензионное
12	Windows 10 Pro	Microsoft	лицензионное
13	Kaspersky Endpoint Security	Kaspersky	лицензионное
14	Microsoft Office 2016	Microsoft	лицензионное
15	Visual Studio 2019	Microsoft	лицензионное
16	Adobe Creative Cloud	Adobe	лицензионное
17	Zoom	Zoom	лицензионное