

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИИ



Федеральное государственное бюджетное образовательное учреждение высшего образования

«РОССИЙСКИЙ ГОСУДАРСТВЕННЫЙ ГУМАНИТАРНЫЙ УНИВЕРСИТЕТ»

(ФГБОУ ВО «РГГУ»)

Институт лингвистики

УНЦ компьютерной лингвистики

Рабочая программа дисциплины

«Статистические модели в лингвистике»

Направление подготовки 45.04.03 Фундаментальная и прикладная лингвистика

Магистерская программа: Фундаментальная и компьютерная лингвистика

Квалификация выпускника: магистр

Форма обучения: очная

РПД адаптирована для лиц
с ограниченными возможностями
здоровья и инвалидов

Москва 2019

Статистические модели в лингвистике

Рабочая программа дисциплины

Составитель:

к.ф-м.н., доцент А.А.Сорокин

Ответственный редактор:

д. филол. н., профессор В.И.Подлеская

УТВЕРЖДЕНО

Протокол заседания УНЦ компьютерной
лингвистики

№ 1 от «28» августа 2019г.

Оглавление

1. Пояснительная записка

- 1.1. Предмет
- 1.2. Цель и задачи дисциплины
- 1.3. Формируемые компетенции и результаты освоения дисциплины
- 1.4. 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 оценок
- 5.3. Оценочные средства для текущего контроля успеваемости и промежуточной аттестации

6. Учебно-методическое и информационное обеспечение дисциплины

- 6.1. Список литературы

7. Материально-техническое обеспечение дисциплины

8. Обеспечение образовательного процесса для лиц с ограниченными возможностями здоровья

9. Приложения

Приложение 1. Аннотация дисциплины

Приложение 2. Лист изменений

1. Пояснительная записка

1.1 Предмет

Предметом дисциплины (модуля) является ознакомление слушателей с современными статистическими моделями и алгоритмическими методами исследования лингвистических данных. На примерах реальных частотных таблиц и таблиц вида «объекты – признаки» изучаются статистические критерии однородности групп текстов (корпусов), рассматривается проблема установления авторства и другие задачи из области компьютерного анализа лингвистической информации.

1.2 Цель и задачи курса

Курс направлен на решение следующих задач:

- освоение магистрантами базовых знаний в области статистического анализа лингвистических данных;
- приобретение навыков исследования данных с помощью программы Excel: фильтрация данных, построение диаграмм, использование генераторов случайных чисел, промежуточных итогов и других встроенных функций;
- обучение методам анализа лингвистической информации с использованием макросов на языке Visual Basic, дополняющих статистические возможности программы Excel;
- выработка у магистрантов умения пользоваться библиотеками функций языка R, реализующих важнейшие методы Data Mining.

1.3 Компетенции обучающегося, формируемые в результате освоения дисциплины

Дисциплина (модуль) направлена на формирование компетенций выпускника:

способностью к абстрактному мышлению, анализу, синтезу (ОК-1);

готовностью действовать в нестандартных ситуациях, нести социальную и этическую ответственность за принятые решения (ОК-2);

готовностью к саморазвитию, самореализации, использованию творческого потенциала (ОК-3);

способностью выбирать оптимальные теоретические подходы и методы решения конкретных научных задач в области лингвистики и новых информационных технологий (ОПК-7);

способностью проводить самостоятельные исследования и получать новые научные результаты в области теории языка, лингвистики конкретных языков, прикладной и компьютерной лингвистики (ПК-1);

способностью изучать и осваивать современные технические средства и информационные технологии, служащие для обеспечения лингвистической деятельности (ПК-2)

и соотнесенных с ними результатов освоения дисциплины (модуля):

Знать:

- фундаментальные понятия и важнейшие теоремы теории вероятностей и математической статистики;
- важнейшие модели и современные методы анализа лингвистических данных.

Уметь:

- пользоваться своими знаниями для решения прикладных задач;
- делать правильные выводы из сопоставления результатов теории и эксперимента;
- проверять статистические гипотезы и производить численные оценки параметров моделей;
- эффективно использовать информационные технологии и компьютерную технику для достижения необходимых теоретических и прикладных результатов.

Владеть:

- навыками освоения большого объема информации;
- практикой исследования и решения теоретических и прикладных задач.

1.4 Место дисциплины в структуре образовательной программы

Дисциплина (модуль) «Статистические модели в лингвистике» является дисциплиной по выбору цикла дисциплин ООП ВПО (магистратуры) по направлению подготовки «Фундаментальная и прикладная лингвистика. Фундаментальная и компьютерная лингвистика» и адресована студентам 1 курса (2 семестр). Дисциплина (модуль) реализуется кафедрой компьютерной лингвистики Института Лингвистики.

Программой дисциплины (модуля) предусмотрены следующие виды контроля: текущий контроль успеваемости в форме: выполнение домашних заданий; автоматическое тестирование по каждой изучаемой теме курса; выполнение практических заданий по ходу изложения лекционного материала; итоговая аттестация в форме экзамена.

Общая трудоемкость освоения дисциплины (модуля) составляет 3 зачетные единицы, 108 часов.

Программой дисциплины (модуля) предусмотрены: лекционные занятия – 4 часа; практические занятия – 26 часов; самостоятельная работа студента – 60 часов, контроль – 18 часов.

2. Структура дисциплины

№ п/ п	Раздел Дисциплины	Семестр	Неделя семестра	Виды учебной работы, включая самостоятельную работу студентов и трудоемкость (в часах)				Формы текущего контроля успеваемости (по неделям семестра) Форма промежуточной аттестации (по семестрам)
				лекц ии	семи- нары	самос тояте льная работ а	ко нтр оль	
1.	Введение. Применение теории вероятности и статистики в	1	1	2		10		

	теоретической и прикладной лингвистике							
2.	Базовые понятия теории вероятностей.	1	2	2	6	10		ДЗ1. Базовые понятия теории вероятностей.
3.	Напоминание: основные понятия математического анализа, операции дифференцирования и интегрирования.	1	3		4	8		ДЗ2. Вычисление производных и интегралов.
4.	Дискретные случайные величины. Математическое ожидание, дисперсия	1	4		4	4		ДЗ3. Характеристики дискретных случайных величин.
5.	Непрерывные случайные величины. Нормальное распределение. Закон больших чисел и центральная предельная теорема	1	5		4	6	4	ДЗ4. Характеристики непрерывных случайных величин.
6.	Основные понятия математической статистики. Оценки параметров случайных величин	3	6	2	2	6		.
7.	Метод максимального правдоподобия и метод моментов.	1	7		2	6		
8.	Критерии значимости для оценок случайных величин. Ошибки первого и второго рода	1	8		4	4		ДЗ5. Оценки и их значимость.
9.	Таблицы сопряженности. Применение критериев значимости к таблицам сопряженности	1	9		2		2	ДЗ6: Оценки значимости для таблиц сопряженности
10.	Применение статистических критериев к реальным лингвистическим данным.	1	10			6		Исследовательский проект (часть 3): Применение статистических критериев к реальным лингвистическим данным
11.	Экзамен							Контрольные вопросы
12.	Итого:			4	26	60	18	

3. Содержание дисциплины

1.1. Введение. Применение теории вероятности и статистики в теоретической и прикладной лингвистике.

Лингвистика как эмпирическая наука. Поиск закономерностей в лингвистических данных. Вероятностная интерпретация задач компьютерной лингвистики.

1.2. Базовые понятия теории вероятностей.

Вероятность, свойства вероятности. Пространство событий. Независимые и зависимые события. Совместная вероятность. Условная вероятность, формула Байеса и формула полной вероятности.

1.3. Напоминание: основные понятия математического анализа, операции дифференцирования и интегрирования.

Функции. Предел функции в точке, производная функции в точке, дифференцируемые функции. Экстремум функции. Первообразная, неопределённый интеграл. Определённый интеграл, формула Ньютона-Лейбница.

1.4. Дискретные случайные величины. Математическое ожидание, дисперсия.

Понятие случайной величины. Основные дискретные распределения: бернуллиево, биномиальное, геометрическое. Математическое ожидание, дисперсия, ковариация, их свойства. Пуассоновское распределение.

1.5. Непрерывные случайные величины. Нормальное распределение. Закон больших чисел и центральная предельная теорема.

Функция распределения и плотность вероятности. Математическое ожидание для непрерывной случайной величины. Основные непрерывные распределения: равномерное, показательное, бета и гамма-распределения, нормальное распределение. Неравенства Чебышева и Маркова. Закон больших чисел. Центральная предельная теорема.

1.6. Основные понятия математической статистики. Оценки параметров случайных величин.

Оценки параметров случайных величин. Эмпирическое среднее и эмпирическая дисперсия. Несмещённые и состоятельные оценки.

1.7. Метод максимального правдоподобия и метод моментов.

Оценки параметров распределения. Метод максимального правдоподобия и метод моментов. Априорное распределение на множестве параметров. Оценки максимального правдоподобия для основных распределений.

1.8. Критерии значимости для оценок случайных величин. Ошибки первого и второго рода.

Проверка статистических гипотез. Нулевая гипотеза, ошибки первого и второго рода. Уровень значимости критерия, критерий Стьюдента

1.9. Таблицы сопряжённости. Применение критериев значимости к таблицам сопряжённости.

Таблицы сопряжённости. Критерий хи-квадрат для таблиц сопряжённости, его применимость.

1.10. Таблицы сопряжённости. Применение критериев значимости к таблицам сопряжённости.

4. Образовательные технологии

В соответствии с требованиями ФГОС по направлению 45.04.03 «Фундаментальная и прикладная лингвистика» и с учетом специфики магистерская программы «Фундаментальная и компьютерная лингвистика» занятия лекционного типа составляют не более 20% аудиторных занятий, а удельный вес занятий, проводимых в интерактивных формах, составляют не менее 40% аудиторных занятий. Интерактивные формы обучения в данном курсе предполагают:

- систематическое использование компьютерных презентаций;
- онлайн демонстрации работы с лингвистическими базами данных и энциклопедическими Интернет-ресурсами;
- использование лингвистических корпусов в режиме онлайн.
- обсуждения курсовых исследовательских проектов;
- работа в группах по выполнению домашних практических заданий.

5. Оценка планируемых результатов обучения

5.1. Система оценивания

При выставлении оценки в ведомость и в зачетную книжку преподаватель должен указать результат в соответствии с традиционной шкалой оценок и со шкалой 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	зачтено	A
83 – 94			B
68 – 82	хорошо		C
56 – 67	удовлетворительно		D
50 – 55			E
20 – 49	неудовлетворительно	не зачтено	FX
0 – 19			F

Распределение баллов по видам учебной деятельности таково:

- посещение семинарских занятий – до 8 баллов,
- уровень активности студента при подготовке к занятиям (конспектирование специальной литературы, готовность отвечать на вопросы по анализу кейсов, активное участие в дискуссиях, коллоквиумах и мозговом штурме и проч.) и во время проведения занятий (участие в обсуждениях и выполнении коллективных заданий) – всего до 32 баллов,
- качество выполнения контрольной работы (текущая аттестация) – до 20 баллов,
- успешность выполнения итогового творческого задания – до 40 баллов.

Оценка «зачтено» выставляется, если студент набрал в сумме не менее 50 баллов. Магистрант, не набравший в сумме 50 баллов, сдаёт зачёт по всему курсу и предъявляет преподавателю собственноручно написанные конспекты специальной литературы и выполненные домашние задания ко всем семинарам.

5.2. Критерии выставления оценок

При выставлении оценки преподаватель ориентируется на следующие содержательные критерии.

Количество баллов	Критерии оценки
95–100 (А)	<p>Оценка выставляется с учетом текущей и промежуточной аттестации.</p> <p>Теоретическое содержание дисциплины освоено полностью, без пробелов, необходимые практические навыки работы с освоенным материалом сформированы, все предусмотренные рабочей программой дисциплины учебные задания выполнены, качество их выполнения оценено числом баллов, близким к максимальному.</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>Компетенции, закреплённые за дисциплиной, сформированы на уровне «высокий».</p>
83–94 (В)	<p>Оценка выставляется с учетом текущей и промежуточной аттестации.</p> <p>Теоретическое содержание дисциплины освоено полностью, без пробелов, необходимые практические навыки работы с освоенным материалом сформированы, почти все задания, предусмотренные рабочей программой дисциплины, выполнены, качество выполнения большинства из них оценено числом баллов, близким к максимальному.</p> <p>Обучающийся адекватно излагает учебный материал, умеет увязывать теорию с практикой, справляется с решением задач профессиональной направленности высокого уровня сложности, правильно обосновывает принятые решения.</p> <p>Достаточно свободно ориентируется в учебной и профессиональной литературе.</p> <p>Почти все компетенции, закреплённые за дисциплиной, сформированы на уровне «высокий».</p>
68–82 (С)	<p>Оценка выставляется с учетом текущей и промежуточной аттестации.</p> <p>Теоретическое содержание дисциплины освоено полностью, без пробелов, некоторые практические навыки работы с освоенным материалом сформированы недостаточно, все предусмотренные рабочей программой дисциплины учебные задания выполнены, качество выполнения ни одного из них не оценено минимальным числом баллов, некоторые виды заданий выполнены с ошибками.</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>Компетенции, закреплённые за дисциплиной, сформированы на уровне «хороший».</p>
56–67 (D)	<p>Оценка выставляется с учетом текущей и промежуточной аттестации.</p>

Количество баллов	Критерии оценки
	<p>Теоретическое содержание дисциплины освоено частично, но пробелы не носят существенного характера, необходимые практические навыки работы с освоенным материалом в основном сформированы, большинство предусмотренных рабочей программой дисциплины учебных заданий выполнено, некоторые из выполненных заданий, возможно, содержат ошибки.</p> <p>Обучающийся испытывает определённые затруднения 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>Компетенции, закреплённые за дисциплиной, сформированы на уровне – «достаточный».</p>
50–55 (E)	<p>Оценка выставляется с учетом текущей и промежуточной аттестации.</p> <p>Теоретическое содержание дисциплины освоено частично, некоторые практические навыки работы не сформированы, многие предусмотренные рабочей программой дисциплины учебные задания не выполнены, либо качество выполнения некоторых из них оценено числом баллов, близким к минимальному.</p> <p>Обучающийся испытывает определённые затруднения 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>Компетенции, закреплённые за дисциплиной, сформированы на уровне «достаточный».</p>
21–49 (FX)	<p>Оценка выставляется с учетом текущей и промежуточной аттестации.</p> <p>Теоретическое содержание дисциплины освоено частично, необходимые практические навыки работы не сформированы, большинство предусмотренных рабочей программой дисциплины учебных заданий не выполнено, либо качество их выполнения оценено числом баллов, близким к минимальному; при дополнительной самостоятельной работе над материалом курса возможно повышение качества выполнения учебных заданий.</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>Компетенции на уровне «достаточный», закреплённые за дисциплиной, не сформированы.</p>
0–20 (F)	Оценка выставляется с учетом текущей и промежуточной

Количество баллов	Критерии оценки
	<p>аттестации.</p> <p>Теоретическое содержание дисциплины не освоено. Необходимые практические навыки работы не сформированы, все предусмотренные рабочей программой дисциплины учебные задания выполнены с грубыми ошибками. Дополнительная самостоятельная работа над материалом дисциплины не приведет к какому-либо значимому повышению качества выполнения учебных заданий.</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>Компетенции на уровне «достаточный», закреплённые за дисциплиной, не сформированы.</p>

5.3. Оценочные средства для текущего контроля успеваемости и промежуточной аттестации

Текущий контроль успеваемости студентов проводится в следующих формах: выполнение домашних заданий (6 заданий – 30 баллов максимум); тестовое задание (максимально 30 баллов); защита исследовательского проекта (максимально - 40 баллов). Для получения удовлетворительной оценки необходимо набрать минимум 60 баллов.

В качестве домашних заданий предлагаются задания следующих типов

- Д31. Базовые понятия теории вероятностей.
 - Д32. Вычисление производных и интегралов.
 - Д33. Характеристики дискретных случайных величин.
 - Д34. Характеристики непрерывных случайных величин.
 - Д35. Оценки и их значимость.
 - Д36: Оценки значимости для таблиц сопряжённости
- Исследовательский проект (часть 3): Применение статистических критериев к реальным лингвистическим данным

Экзамен ориентирован на следующие контрольные вопросы

- Вычисление вероятности события при условии другого события, формулы Байеса и полной вероятности.
- Нахождение математического ожидания, дисперсии и других характеристик для заданных случайных величин и распределений.
- Проверка несмещённости оценки для параметра распределения.
- Проверка статистической гипотезы на основе выборки с помощью данного критерия.
- Анализ таблицы сопряжённости с помощью критерия хи-квадрат и других критериев.
- Нахождение оптимальных параметров распределения с помощью метода максимального правдоподобия.

6. Учебно-методическое и информационное обеспечение дисциплины

6.1. Список литературы

Основная литература

1. *А.В. Антонов, М.С. Никулин, А.М. Никулин, В.А. Чепурко.* Теория надежности. Статистические модели. Учеб.пособие — М.: ИНФРА-М, 2018. — 576 с. + Доп. Материалы
2. Белько И.В. Теория вероятностей, математическая статистика, математическое программирование: Учебное пособие. - 1. - Москва; Минск: ООО "Научно издательский центр ИНФРА-М»; ООО "Новое знание", 2016. - 299 с.
3. Глинский В.В. Статистика: Учебник. - 4; перераб. и доп. - Москва: ООО "Научно-издательский центр ИНФРА-М", 2017. - 355 с.
4. Кочетков Е.С. Теория вероятностей и математическая статистика: Учебник. - 2; испр. и перераб. - Москва; Москва: Издательство "ФОРУМ»; ООО "Научно-издательский центр ИНФРА-М", 2018. - 240 с.
5. *М.Б. Лагутин.* Наглядная математическая статистика. М.: БИНОМ, Лаборатория знаний, 2007.

Рекомендованная литература

1. *А. А. Боровков* Математическая статистика. М.: Наука, 1984.
2. *Ю. А. Розанов.* Теория вероятностей, случайные процессы и математическая статистика.
3. *Г. В. Ивченко, В. И. Медведев.* Математическая статистика. М.: Высшая школа, 1984.
4. *М.Б. Лагутин.* Наглядная математическая статистика. М.: БИНОМ, Лаборатория знаний, 2007.
5. *В. Н. Тутубалин.* Теория вероятностей и случайных процессов. М.: Издательство МГУ, 1992.

7. Материально-техническое обеспечение дисциплины

Занятия по курсу «Статистические модели в лингвистике» можно проводить с максимальной эффективностью, если проводить их в компьютерном классе с доступом в Интернет, проектором и экраном для презентаций. Необходимо также наличие доски, чтобы преподаватель мог разбирать примеры по ходу объяснения и записывать задания.

8. Обеспечение образовательного процесса для лиц с ограниченными возможностями здоровья

При необходимости рабочая программа дисциплины может быть адаптирована для обеспечения образовательного процесса лицам с ограниченными возможностями здоровья, в том числе для дистанционного обучения. Для этого от студента требуется представить заключение психолого-медико-педагогической комиссии (ПМПК) и личное заявление (заявление законного представителя).

В заключении ПМПК должно быть прописано:

- рекомендуемая учебная нагрузка на обучающегося (количество дней в неделю, часов в день);

- оборудование технических условий (при необходимости);
- сопровождение и (или) присутствие родителей (законных представителей) во время учебного процесса (при необходимости);
- организация психолого-педагогического сопровождение обучающегося с указанием специалистов и допустимой нагрузки (количества часов в неделю).

Для осуществления процедур текущего контроля успеваемости и промежуточной аттестации обучающихся, при необходимости могут быть созданы фонды оценочных средств, адаптированные для лиц с ограниченными возможностями здоровья и позволяющие оценить достижение ими запланированных в основной образовательной программе результатов обучения и уровень сформированности всех компетенций, заявленных в образовательной программе.

Форма проведения текущей и итоговой аттестации для лиц с ограниченными возможностями здоровья устанавливается с учетом индивидуальных психофизических особенностей (устно, письменно (на бумаге, на компьютере), в форме тестирования и т.п.). При необходимости студенту предоставляется дополнительное время для подготовки ответа на зачете или экзамене.

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. Приложения

Приложение 1. Аннотация дисциплины

Предметом дисциплины (модуля) является ознакомление слушателей с современными статистическими моделями и алгоритмическими методами исследования лингвистических данных. На примерах реальных частотных таблиц и таблиц вида «объекты – признаки» изучаются статистические критерии однородности групп текстов (корпусов), рассматривается проблема установления авторства и другие задачи из области компьютерного анализа лингвистической информации.

Курс направлен на решение следующих задач:

- освоение магистрантами базовых знаний в области статистического анализа лингвистических данных;
- приобретение навыков исследования данных с помощью программы Excel: фильтрация данных, построение диаграмм, использование генераторов случайных чисел, промежуточных итогов и других встроенных функций;
- обучение методам анализа лингвистической информации с использованием макросов на языке Visual Basic, дополняющих статистические возможности программы Excel;
- выработка у магистрантов умения пользоваться библиотеками функций языка R, реализующих важнейшие методы Data Mining.

Дисциплина (модуль) направлена на формирование компетенций выпускника:
способностью к абстрактному мышлению, анализу, синтезу (ОК-1);

готовностью действовать в нестандартных ситуациях, нести социальную и этическую ответственность за принятые решения (ОК-2);

готовностью к саморазвитию, самореализации, использованию творческого потенциала (ОК-3);

способностью выбирать оптимальные теоретические подходы и методы решения конкретных научных задач в области лингвистики и новых информационных технологий (ОПК-7);

способностью проводить самостоятельные исследования и получать новые научные результаты в области теории языка, лингвистики конкретных языков, прикладной и компьютерной лингвистики (ПК-1);

способностью изучать и осваивать современные технические средства и информационные технологии, служащие для обеспечения лингвистической деятельности (ПК-2)

и соотнесенных с ними результатов освоения дисциплины (модуля):

Знать:

- фундаментальные понятия и важнейшие теоремы теории вероятностей и математической статистики;
- важнейшие модели и современные методы анализа лингвистических данных.

Уметь:

- пользоваться своими знаниями для решения прикладных задач;
- делать правильные выводы из сопоставления результатов теории и эксперимента;
- проверять статистические гипотезы и производить численные оценки параметров моделей;
- эффективно использовать информационные технологии и компьютерную технику для достижения необходимых теоретических и прикладных результатов.

Владеть:

- навыками освоения большого объёма информации;
- практикой исследования и решения теоретических и прикладных задач.

Дисциплина (модуль) *«Статистические модели в лингвистике»* является дисциплиной по выбору частью цикла дисциплин ООП ВПО (магистратуры) по направлению подготовки *«Фундаментальная и прикладная лингвистика. Фундаментальная и компьютерная лингвистика»* и адресована студентам *1 курса (2 семестр)*. Дисциплина (модуль) реализуется кафедрой компьютерной лингвистики Института Лингвистики.

Программой дисциплины (модуля) предусмотрены следующие виды контроля: текущий контроль успеваемости в форме: выполнение домашних заданий; автоматическое тестирование по каждой изучаемой теме курса; выполнение практических заданий по ходу изложения лекционного материала; итоговая аттестация в форме экзамена.

Общая трудоемкость освоения дисциплины (модуля) составляет 3 зачетные единицы, 108 часов.

Программой дисциплины (модуля) предусмотрены: лекционные занятия – 6 часов; практические занятия – 24 часа; самостоятельная работа студента – 60 часов, контроль – 18 часов.

Приложение 2. Лист изменений

ЛИСТ ИЗМЕНЕНИЙ

№	Текст актуализации или прилагаемый к РПД документ, содержащий изменения	Дата	№ протокола
1	Приложение к листу изменений №1	31.08.2020г	1

Приложение к листу изменений №1

1. Структура дисциплины (к п. 2 РПД на 2020)

Общая трудоёмкость дисциплины составляет 3 з.е., 114 ч., в том числе контактная работа обучающихся с преподавателем 30 ч., самостоятельная работа обучающихся 66 ч.

2. 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.

3. Перечень БД и ИСС (к п. 6 на 2020 г.)

№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. Состав программного обеспечения (ПО) (к п. 7 на 2020 г.)

№п	Наименование ПО	Производитель	Способ распространения (лицензионное или свободно распространяемое)

1	Microsoft Share Point 2010	Microsoft	лицензионное
2	Windows 10 Pro	Microsoft	лицензионное
3	Kaspersky Endpoint Security	Kaspersky	лицензионное
4	Microsoft Office 2016	Microsoft	лицензионное
5	Zoom	Zoom	лицензионное