

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Наумова Наталия Александровна
Должность: Ректор
Дата подписания: 24.10.2024 14:21:41
Уникальный программный ключ:
6b5279da4e03c1e9171803da5b7b591a9e2

МИНИСТЕРСТВО ПРОСВЕЩЕНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное автономное образовательное учреждение высшего образования
«ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ПРОСВЕЩЕНИЯ»
(ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ ПРОСВЕЩЕНИЯ)

Лингвистический факультет
Кафедра теории языка, англистики и прикладной лингвистики

Согласовано
деканом лингвистического факультета
« 14 » марта 2024 г.
ВЗ
/ Вековищева С.Н./

Согласовано
деканом факультета романо-германских
языков
«20» марта 2024 г.
Шабанова В.П.
/Шабанова В.П./

Рабочая программа дисциплины

Формализация лингвистической информации

Направление подготовки
45.04.02 Лингвистика

Программа подготовки:
Теория обучения иностранным языкам и межкультурная коммуникация
(английский язык)

Квалификация
Магистр

Форма обучения
Заочная

Согласовано учебно-методической комиссией
лингвистического факультета
Протокол «14» марта 2024 г. № 5
Председатель УМКом Горбачева О.А.
/Горбачева О.А./

Рекомендовано кафедрой теории языка,
англистики и прикладной лингвистики
Протокол от «26» февраля 2024 г. № 8
Зав. кафедрой Холстинина Т.В.
/Холстинина Т.В./

Мытищи
2024

Автор-составитель:
Максименко О. И. доктор филологических наук, профессор

Рабочая программа дисциплины «Формализация лингвистической информации» составлена в соответствии с требованиями Федерального государственного образовательного стандарта высшего образования по направлению подготовки 45.04.02 Лингвистика, утвержденного приказом МИНОБРНАУКИ РОССИИ от 12.08.2020 года № 992.

Дисциплина входит в обязательную часть Блока 1 «Дисциплины (модули)» и является обязательной для изучения.

Год начала подготовки (по учебному плану) 2022

СОДЕРЖАНИЕ

1. Планируемые результаты обучения.....	4
2. Место дисциплины в структуре образовательной программы.....	4
3. Объем и содержание дисциплины.....	5
4. Учебно-методическое обеспечение самостоятельной работы обучающихся.....	6
5. Фонд оценочных средств для проведения текущей и промежуточной аттестации по дисциплине.....	7
6. Учебно-методическое и ресурсное обеспечение дисциплины.....	19
7. Методические указания по освоению дисциплины.....	19
8. Информационные технологии для осуществления образовательного процесса по дисциплине.....	19
9. Материально-техническое обеспечение дисциплины.....	20

1. ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ

1.1. Цель и задачи дисциплины

Образовательная цель освоения дисциплины «Формализация лингвистической информации» предполагает повышение уровня образования студентов путем приобщения их к общенаучному знанию, пониманию необходимости формализации лингвистических методов и использования современных информационных технологий для научных исследований.

Практическая цель состоит в формировании у студентов компетенций, необходимых для анализа явлений языка формальными методами, понимания включенности исследований языка в новые информационные технологии, реализация полученных компетенций в научно-исследовательской и научно-практической деятельности.

Задачи дисциплины:

- актуализация и развитие знаний в области математических методов исследования языка и новых информационных технологий;
- формирование навыков анализа количественных методов на разных уровнях языка;
- систематизация знаний о языке с позиций формализованной лингвистики

1.2. Планируемые результаты обучения

В результате освоения данной дисциплины у обучающихся формируются следующие компетенции:

ОПК–6. Способен применять современные технологии при осуществлении сбора, обработки и интерпретации данных эмпирического исследования; составлять и оформлять научную документацию

ОПК–7. Способен работать с основными информационно-поисковыми и экспертными системами, системами представления знаний и обработки вербальной информации

2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ

Данная учебная дисциплина входит в обязательную часть Блока 1 «Дисциплины (модули)» и является обязательной для изучения.

Учебная дисциплина «Формализация лингвистической информации» опирается на знания, полученные при изучении следующих учебных курсов на предыдущих уровнях образования: «Основы языкознания», «Теоретическая грамматика», а также исходит из взаимодействия с дисциплиной программы подготовки магистратуры «Общее языкознание и история лингвистических учений», «Теоретические и методологические основы межкультурной коммуникации», «Сопоставительное языкознание: история и основные направления». Полученные в процессе обучения знания могут быть использованы в научно-исследовательской работе и при написании выпускной квалификационной работы.

3. ОБЪЕМ И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

3.1. Объем дисциплины

Показатель объема дисциплины	Форма обучения
	Заочная
Объем дисциплины в зачетных единицах	2
Объем дисциплины в часах	72
Контактная работа:	10,2

Лекции	4
Практические занятия	6
Контактные часы на промежуточную аттестацию:	0,2
Зачет	0,2
Самостоятельная работа	54
Контроль	7,8

Формой промежуточной аттестации: зачет во 2 семестре на 1 курсе.

3.2. Содержание дисциплины

Наименование разделов (тем) Дисциплины с кратким содержанием	Виды занятий	
	Лекции	Практические занятия
Тема 1. Введение. Формальные методы в современной лингвистике. Лингвостатистика. Основные статистические категории. Выборка, типы переменных в зависимости от уровня лингвистического анализа, элементарное событие. Частота, распределение частот встречаемости единиц разного уровня. Критерии значимости. Типология распределений.	0.5	1
Тема 2. Применение количественных методов на разных уровнях языка. Использование математических методов в фонетике, морфологии, синтаксисе. Законы Ципфа, новые трактовки. Статистико-комбинаторные, дистрибутивно-статистические методы в грамматике.	1	1
Тема 3. Формальные методы в лексикографии. Частотные словари, словари-индексы, конкордансы. Компьютерная лексикография. Автоматические словари. Национальные корпуса, правила разметки. Зависимость «словарь- текст».	0.5	0.5
Тема 4. Прикладные аспекты формализации лингвистических данных. Задачи атрибуции текста. Математико-статистические методы в стилистике. Семантические сети. Нейронные системы.	0.5	0.5
Тема 5. Системы анализа, представления и обработки лингвистических данных. Data mining. Системы извлечения знаний с машинным обучением. Сентимент-анализ. Количественный контент-анализ.	0.5	1
Тема 6. Новые информационные и коммуникационные технологии. Интернет-лингвистика. Системы обмена информацией. Проблемы обработки данных.	1	1
Итого	4	6

4. УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ САМОСТОЯТЕЛЬНОЙ РАБОТЫ ОБУЧАЮЩИХСЯ

Темы для самостоятельного изучения	Изучаемые вопросы	Количество часов	Формы самостоятельной работы	Методические обеспечения	Формы отчетности
Тема 1. Введение и основные формальные методы в современной лингвистике	Элементы лингвостатистики. Выборка, типы переменных в зависимости от уровня лингвистического анализа. Частота, типы частот встречаемости языковых единиц. Типология распределений	8	Чтение литературы по обсуждаемым вопросам Подготовка доклада Подготовка проекта	Рекомендуемая литература Интернет-ресурсы	Доклад, проект
Тема 2. Применение математических методов на разных уровнях языка	Использование математических методов в фонетике, морфологии, синтаксисе, лексикологии. Стилеметрия	12	Чтение литературы по обсуждаемым вопросам Подготовка доклада Подготовка проекта	Рекомендуемая литература Интернет-ресурсы	Доклад, проект
Тема 3. Квантитативные методы в лексикографии	Частотные словари, словари-индексы, конкордансы. Национальные корпуса, правила разметки.	8	Чтение литературы по обсуждаемым вопросам Подготовка доклада Подготовка проекта	Рекомендуемая литература Интернет-ресурсы	Доклад, проект
Тема 4. Прикладные аспекты квантитативной лингвистики	Контент-анализ (количественный и качественный). Семантические сети. Нейронные системы	10	Чтение литературы по обсуждаемым вопросам Подготовка доклада Подготовка проекта	Рекомендуемая литература Интернет-ресурсы	Доклад, проект
Тема 5. Системы анализа, представления	Извлечение данных. Извлечение мнений,	8	Чтение литературы по обсуждаемым	Рекомендуемая литература Интернет-	Доклад, проект

и обработки лингвистических данных	сентимент-анализ. Системы извлечения знаний с машинным обучением		вопросам Подготовка доклада Подготовка проекта	ресурсы	
Тема 6. Новые информационные и коммуникационные технологии	Интернет-лингвистика. Системы обмена информацией. Проблемы распознавания и обработки данных.	8	Чтение литературы по обсуждаемым вопросам Подготовка доклада Подготовка проекта	Рекомендуемая литература Интернет-ресурсы	Доклад, проект
итого		54			

5. ФОНД ОЦЕНОЧНЫХ СРЕДСТВ ДЛЯ ПРОВЕДЕНИЯ ТЕКУЩЕЙ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ПО ДИСЦИПЛИНЕ

5.1. Перечень компетенций с указанием этапов их формирования в процессе освоения образовательной программы

Код и наименование компетенции	Этапы формирования
ОПК-6. Способен применять современные технологии при осуществлении сбора, обработки и интерпретации данных эмпирического исследования; составлять и оформлять научную документацию	1. Работа на учебных занятиях 2. Самостоятельная работа
ОПК-7. Способен работать с основными информационно-поисковыми и экспертными системами, системами представления знаний и обработки вербальной информации	1. Работа на учебных занятиях 2. Самостоятельная работа

5.2. Описание показателей и критериев оценивания компетенций на различных этапах их формирования, описание шкал оценивания

Оцениваемые компетенции	Уровень сформированности	Этап формирования	Описание показателей	Критерии оценивания	Шкала оценивания
ОПК-6	Пороговый	1. Работа на учебных занятиях 2. Самостоятельная работа	Знает: основные современные технологии сбора информации Умеет: проводить обработку и интерпретацию данных	Тест, проект	Шкала оценивания теста Шкала оценивания проекта

			эмпирического исследования		
	Продви- нутый	1. Работа на учебных занятиях 2. Самостоятельная работа	Знает: основные современные технологии сбора информации Умеет: проводить обработку и интерпретацию данных эмпирического исследования Владеет: навыками оформления научной документации	Тест, проект доклад	Шкала оценивания теста Шкала оценивания проекта Шкала оценивания доклада
ОПК-7	Порогов- ый	1. Работа на учебных занятиях 2. Самостоятельная работа	Знает: основные информационно- поисковые и экспертные системы Умеет: корректно использовать профильные информационно- коммуникационные ресурсы сети интернет	Тест, проект	Шкала оценивания теста Шкала оценивания проекта
	Продви- нутый	1. Работа на учебных занятиях 2. Самостоятельная работа	Знает: основные информационно- поисковые и экспертные системы Умеет: корректно использовать профильные информационно- коммуникационные ресурсы сети интернет Владеет: навыками эффективного использования образовательных ресурсов для повышения квалификации и расширения научной компетенции	Тест, проект, доклад	Шкала оценивания теста Шкала оценивания проекта Шкала оценивания доклада

Описание шкал оценивания

Вид работы	Критерии оценивания
1. Проект	10 баллов , если представленный на защите проект свидетельствует о проведенном самостоятельном исследовании с привлечением различных источников информации, отражает теоретическую и практическую направленность проекта; во время защиты проекта продемонстрированы коммуникативные и рефлексивные умения, а также навыки работы в команде.
	8 баллов , если представленный на защиту проект свидетельствует о проведенном исследовании с привлечением различных источников информации, отражает теоретическую и практическую направленность проекта.
	4 балла , если представленный на защиту проект свидетельствует о проведенном исследовании, отражает теоретическую направленность проекта.
	2 балла , если представленный на защиту проект свидетельствует о проведенном исследовании с привлечением одного источника информации; слабо отражает теоретическую и практическую направленность проекта.
2. Тест	10 баллов , если из всех заданий студент выполнил как минимум 80%
	8 баллов , если из всех заданий студент выполнил как минимум 60%
	4 балла , если из всех заданий студент выполнил 40%
3. Доклад	10 баллов , если представленный доклад свидетельствует о проведенном самостоятельном исследовании с привлечением различных источников информации; логично, связно и полно раскрывается тема; заключение содержит логично вытекающие из содержания выводы; правильно (уместно и достаточно) используются разнообразные средства речи
	8 балла , если представленный доклад свидетельствует о проведенном самостоятельном исследовании с привлечением двух-трех источников информации; логично, связно и полно раскрывается тема; заключение содержит логично вытекающие из содержания выводы; правильно (уместно и достаточно) используются разнообразные средства речи
	4 балла , если представленный доклад свидетельствует о проведенном исследовании с привлечением одного источника информации; тема раскрыта не полностью; логичный вывод не сделан
	0 баллов , если доклад не представлен

5.3. Типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций в процессе освоения образовательной программы

Образец тестовых заданий

1. База данных это

а. Корпус текстов, организованных по внутренним правилам

- б. Именованная совокупность данных, организованных по определенным правилам, предусматривающим общие принципы описания, хранения и манипулирования данными
- в. Именованная совокупность файлов, организованных по определенным правилам
- г. Именованный корпус любых данных с возможностью хранения и оперирования данными

2. Дистрибутивно-статистический анализ это

- а. формальный метод анализа лексики
- б. формальный метод анализа фонетики
- в. формальный метод анализа распределения разных языковых единиц
- г. формальный метод подсчета разных языковых единиц

Образец проекта

1. Провести анализ любого языкового материала (лексической выборки, текстовой выборки и пр.) одним из изученных формальных методов (контент-анализ, дистрибутивно-статистический анализ, сентимент-анализ)

Примерные темы докладов

1. Описание и анализ прикладной математической программы MatLab
2. Описание и анализ автоматической экспертной системы по выбору.
3. Описание и анализ автоматической обучающей системы по выбору.
4. Описание и анализ информационно-поисковой системы по выбору.
5. Описание и анализ любой компьютерной системы контент-анализа по выбору.

Анализ проводится с целью последующего обсуждения на практических занятиях. Особое внимание уделяется проблеме эффективности обработки естественного языка такими программами с позиций пользователя и разработчика подобных систем.

Примерные вопросы к зачету

1. Основные статистические категории. Понятие выборки и совокупности, типы переменных в зависимости от уровня лингвистического анализа. Понятие элементарного события.
2. Понятие частоты и распределения частот встречаемости единиц разного уровня. Случайная выборка. Методы оценки представительности распределения. Критерии значимости.
3. Нормальное распределение. Интерпретация лингвистических распределений.
4. Планирование статистического эксперимента и проверка гипотез.
5. Критерий χ^2 .
6. Системная и прагматическая выборка лингвистических данных. Частоты: индивидуальная, частота класса, абсолютная, относительная, условная и позиционная.
7. Понятие и методы лингвистической дешифровки. Эвристические модели.
8. Применение статистических методов в фонетике: построение системы фонем с использованием критерия частоты встречаемости, количественные модели слогаделения. Фонотактические типы слов. Длина слов.
9. Статистико-комбинаторные, дистрибутивно-статистические методы в грамматике.
10. Метод дистрибутивно-статистического анализа: традиционный и автоматизированный варианты. Использование ДСА для автоматического распознавания жанров текстов в информационном потоке.
11. Частотные словари, словари-индексы, конкордансы. Национальные корпуса.
12. Характеристики частотных словарей и индексы покрытия текста.
13. Ранговое распределение и закон Ципфа-Мандельброта. Разные точки зрения, новые трактовки.
14. Зависимость "словарь-текст". Теория вероятностных систем.

15. Грамматический и семантический аспекты количественного исследования лексики. Словообразовательная структура слов. Математическая теория словообразования. Части речи, лексико-семантические группы. Полисемия, связь с частотой слова.
16. Сентимент-анализ.
17. Контент-анализ, количественный и качественный.
18. Задачи атрибуции текста. Лексико-стилистический анализ текстов. Математико-статистические методы в стилистике.
19. Типология лингвостатистических распределений.
20. Понятие семантических сетей. Нейронные сети.
21. Прикладные аспекты количественной лингвистики. Их использование в реальных пакетах прикладных программ.
22. Нечеткая логика в лингвистике.
23. Новые информационные технологии в лингвистике. Экспертные системы
24. Информационно-поисковые системы. Типология и принципы строения. Поисковые системы Интернета.
25. Компьютерная лингводидактика. Онлайн и оффлайн обучающие системы

5.4. Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций.

В рамках освоения дисциплины предусмотрены следующие формы текущего контроля: тест, проект, доклад.

Промежуточная аттестация по дисциплине «Формализация лингвистической информации» учитывает уровень результатов обучения, общее качество работы магистранта, дисциплинированность, самостоятельность. Освоение дисциплины оценивается по балльной шкале.

Общее количество баллов по дисциплине – 100 баллов.

Максимальное количество баллов, которое может набрать магистрант в течение семестра за текущий контроль, равняется 80 баллам.

Промежуточная аттестация проводится в форме зачета во втором семестре на 1 курсе. Зачет проводится в форме устного собеседования по вопросам.

Максимальная сумма баллов, которые магистрант может набрать на зачете, равняется 20 баллам.

Шкала оценивания ответа на зачете

Критерии оценивания	Количество баллов			
Полнота ответа на теоретический вопрос	4 Ответ полный, с привлечением знаний из разных разделов курса	2 Ответ полный	1 Ответ неполный	0 Ответ, не соответствующий теоретическому вопросу
Знание терминологии, умение давать определения понятиям	4 Четкие определения, умение объяснить их и дополнить	2 Определения даются без собственных объяснений и дополнений	1 Определения даются с некоторыми неточностями	0 отсутствует

Знание персоналий, сопряженных с теоретическим вопросом	4 5 и более примеров	2 3-4 примера	1 1-2 примера	0 отсутствует
Умение проиллюстрировать явление практическими примерами	4 5 и более примеров примеров	2 3-4 примера	1 1-2 примера	0 отсутствие
Ответы на вопросы экзаменатора	4 Ответы на вопросы полные с приведением примеров и/или пояснений	2 Ответы на вопросы полные или частично полные	1 Только ответы на элементарные вопросы	0 Нет ответов на вопросы
Итоговый балл (максимальный)	20	10	5	0

Итоговая шкала оценивания результатов освоения дисциплины

Итоговая оценка по дисциплине формируется из суммы баллов по результатам текущего контроля и промежуточной аттестации и выставляется в соответствии с приведенной ниже таблицей.

Оценка по 100-балльной системе	Оценка по дисциплине
41-100	«зачтено»
0-40	«не зачтено»

6. УЧЕБНО-МЕТОДИЧЕСКОЕ И РЕСУРСНОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

6.1. Основная литература

1. Баранов, А.Н. Введение в прикладную лингвистику / А.Н. Баранов. 6-е изд., стереотип. М.: URSS, 2021. - 368с. – Текст: непосредственный.
2. Волосатова, Т. М. Информатика и лингвистика : учебное пособие / Т.М. Волосатова, Н.В. Чичварин. — Москва : ИНФРА-М, 2021. — 196 с. + Доп. материалы [Электронный ресурс]. — (Высшее образование: Бакалавриат). — DOI 10.12737/16175. - ISBN 978-5-16-010977-0. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1212524> (дата обращения: 30.11.2021). – Режим доступа: по подписке.
3. Гребенщикова, А.В. Основы количественной лингвистики и новых информационных технологий: учебное пособие/А.В. Гребенщикова 3-е изд. стереотип. – М.: Флинта, 2018. – 152 с. – Текст: непосредственный.

6.2. Дополнительная литература

4. Костюк А.В., Бобонец С.А., Флагонтов А.В., Черных А.К. Информационные етехнологии. Базовый курс: учеб.пособие / А.В. Костюк. - М.: Лань, 2019. – 604 с.
5. Лапшин, В.А. Лекции по математической лингвистике / В. А. Лапшин. - М. : Науч.мир, 2010. - 248с. – Текст: непосредственный.
6. Максименко О.И. Формализованная лингвистика. М., 2013.
7. Потапова Р.Г. Новые информационные технологии и лингвистика. [Текст] М.: УРСС, 2021. – 368 с.
8. Корнеев В.В. и др. Базы данных. Интеллектуальная обработка информации. М.: Нолидж, 2000, 351 с.
9. Шайкевич А.Я.. Введение в лингвистику. Учебное пособие. 2-е изд., испр. М.: «Академия», 2010. 394 с.
10. Шемакин Ю.И. Семантика самоорганизующихся систем. – М.: Академический Проект, 2003.

6.3. Ресурсы информационно-телекоммуникационной сети «Интернет»

www.philology.ru
www.postnauka.ru
www.elibrary.ru
www.krugosvet.ru
www.knigafund.ru

7. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ

1. Методические рекомендации по организации самостоятельной работы обучающихся
2. Методические рекомендации по проведению практических занятий по теоретическим дисциплинам

8. ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ ДЛЯ ОСУЩЕСТВЛЕНИЯ ОБРАЗОВАТЕЛЬНОГО ПРОЦЕССА ПО ДИСЦИПЛИНЕ

Лицензионное программное обеспечение:

Microsoft Windows

Microsoft Office

Kaspersky Endpoint Security

Информационные справочные системы:

Система ГАРАНТ

Система «КонсультантПлюс»

Профессиональные базы данных:

fgosvo.ru

pravo.gov.ru

www.edu.ru

Свободно распространяемое программное обеспечение, в том числе отечественного производства

ОМС Плеер (для воспроизведения Электронных Учебных Модулей)

7-zip

Google Chrome

9. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Материально-техническое обеспечение дисциплины включает в себя:

- учебные аудитории для проведения занятий лекционного и семинарского типа, курсового проектирования (выполнения курсовых работ), групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, укомплектованные учебной мебелью, доской, демонстрационным оборудованием.
- помещения для самостоятельной работы, укомплектованные учебной мебелью, персональными компьютерами с подключением к сети Интернет и обеспечением доступа к электронным библиотекам и в электронную информационно-образовательную среду МГОУ;
- помещения для хранения и профилактического обслуживания учебного оборудования, укомплектованные мебелью (шкафы/стеллажи), наборами демонстрационного оборудования и учебно-наглядными пособиями.