

**МИНОБРНАУКИ РОССИИ**

**Федеральное государственное автономное образовательное  
учреждение высшего образования "Пермский  
государственный национальный исследовательский  
университет"**

**Кафедра теоретического и прикладного языкознания**

**Авторы-составители: Белоусов Константин Игоревич**

**Рабочая программа дисциплины**

**МАТЕМАТИЧЕСКАЯ ЛИНГВИСТИКА**

**Код УМК 76454**

**Утверждено  
Протокол №9  
от «30» мая 2023 г.**

**Пермь, 2023**

## **1. Наименование дисциплины**

Математическая лингвистика

## **2. Место дисциплины в структуре образовательной программы**

Дисциплина входит в обязательную часть Блока « Б.1 » образовательной программы по направлениям подготовки (специальностям):

Направление подготовки: **45.03.03** Фундаментальная и прикладная лингвистика  
направленность Компьютерная лингвистика

### **3. Планируемые результаты обучения по дисциплине**

В результате освоения дисциплины **Математическая лингвистика** у обучающегося должны быть сформированы следующие компетенции:

**45.03.03** Фундаментальная и прикладная лингвистика (направленность : Компьютерная лингвистика)

**ОПК.4** Способен к ведению профессиональной деятельности с опорой на основы математических дисциплин, необходимых для формализации лингвистических знаний и процедур анализа и синтеза лингвистических структур

#### **Индикаторы**

**ОПК.4.1** демонстрирует знание основ математических дисциплин

**ОПК.4.2** применяет знание основ математических дисциплин для формализации лингвистических знаний и процедур анализа и синтеза лингвистических структур

#### 4. Объем и содержание дисциплины

|                                                                           |                                                                                                 |
|---------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------|
| <b>Направление подготовки</b>                                             | 45.03.03 Фундаментальная и прикладная лингвистика<br>(направленность: Компьютерная лингвистика) |
| <b>форма обучения</b>                                                     | очная                                                                                           |
| <b>№№ триместров,<br/>выделенных для изучения<br/>дисциплины</b>          | 11                                                                                              |
| <b>Объем дисциплины (з.е.)</b>                                            | 3                                                                                               |
| <b>Объем дисциплины (ак.час.)</b>                                         | 108                                                                                             |
| <b>Контактная работа с<br/>преподавателем (ак.час.),<br/>в том числе:</b> | 42                                                                                              |
| <b>Проведение лекционных<br/>занятий</b>                                  | 28                                                                                              |
| <b>Проведение практических<br/>занятий, семинаров</b>                     | 14                                                                                              |
| <b>Самостоятельная работа<br/>(ак.час.)</b>                               | 66                                                                                              |
| <b>Формы текущего контроля</b>                                            | Письменное контрольное мероприятие (4)                                                          |
| <b>Формы промежуточной<br/>аттестации</b>                                 | Экзамен (11 триместр)                                                                           |

## **5. Аннотированное описание содержания разделов и тем дисциплины**

### **Математическая лингвистика**

В рамках курса "Математическая лингвистика" изучаются формальные модели текстов на естественном языке, основанные на методах математической логики, линейной алгебры, теории вероятностей и математической статистики; рассматриваются формальные грамматики как средство описания синтаксиса языка. Студенты знакомятся со способами представления данных в лингвистических моделях.

### **Множества, функции, отношения**

В разделе рассматриваются вопросы представления данных в лингвистических моделях, категоризации и классификации лингвистических объектов, онтологий и тезаурусов.

### **Представление данных в лингвистических моделях. Классификация и категоризация лингвистических объектов в процессе моделирования.**

Основные группы методов классификации и кластеризации текстов. Классификация текстов, машинное обучение с учителем. Этапы решения задач классификации: предобработка и индексация документов; уменьшение размерности пространства признаков; построение и обучение классификатора с помощью методов машинного обучения; оценка качества классификации). Кластеризация текстов, машинное обучение без учителя. Методы обучения для решения задач кластеризации.

### **Онтологии и тезаурусы**

Представление знаний для компьютерной обработки. Тезаурус как разновидность словаря общей или специальной лексики, в котором указаны семантические отношения между терминами. Отличия тезауруса от толкового словаря. Основные семантические отношения, используемые в тезаурусах. Онтология как точная спецификация некоторой области, включающая в себя словарь терминов предметной области и множество логических связей, описывающих соотношения этих терминов между собой. Основные элементы онтологий.

### **Лексические функции (модель "СМЫСЛ - ТЕКСТ")**

Исследование семантики в рамках теории "Смысл - Текст". Лексическая функция как функция, аргументами которой являются некоторые слова или словосочетания данного языка, а значениями - множества слов и словосочетаний этого же языка.

### **Логические методы**

В разделе рассматриваются логические методы, в том числе логика предикатов, интерпретация логических выражений и равносильность формул.

### **Логика предикатов (первого порядка). Интерпретация логических выражений**

Логика предикатов первого порядка как развитие традиционной логики Аристотеля и логики высказываний. Понятие высказывания. Истинность или ложность высказывания. Предикатный символ как выражение, означающее свойство сущности или тип отношения между сущностями. Терм как выражение, означающее значение свойства сущности или связываемые отношением сущности. Предикат как логическая булева функция с аргументами в виде термов. N-местный предикат. Множество истинности предиката. Неделимость предиката. Синтаксис и семантика логики предикатов первого порядка.

### **Равносильность формул**

Предикаты как "переменные" высказывания. Предикатные формулы. Логика высказываний и ее операции. Тожественность логических значений. Равносильность на области. Понятие равносильности формул логики предикатов.

## **Формальные грамматики**

В данном подразделе дается понятие о формальной грамматике как способе описания формального языка

### **Формальные грамматики как средство описания синтаксиса языка**

Способы создания формальных языков. Перечисляющие грамматики. Порождающие грамматики (генеративные грамматики). Распознающие грамматики (аналитические грамматики).

### **Порождающие грамматики. Категориальные грамматики**

Регулярные грамматики (автоматные грамматики). КС-грамматики. КЗ-грамматики. Неограниченные грамматики (грамматики с фразовой структурой). Алфавит. Цепочка нулевой длины. Регулярные множества.

## **Статистические методы**

В разделе рассматривается применение статистических методов в экспериментальных лингвистических исследованиях и в корпусных исследованиях.

### **Статистические методы в экспериментальных исследованиях**

Статистические методы как инструмент сбора и анализа информации. Генеральная совокупность и выборка. Качественная и количественная достоверность выборки. Формирование выборки. Описательные статистики. Меры центральной тенденции и параметры варьирования выборки. Статистические критерии. Статистический анализ. Применение конкретных статистических процедур в экспериментальных лингвистических исследованиях.

### **Статистические методы в корпусных исследованиях**

Корпусно-ориентированные исследования. Статистические параметры корпусов. Вопрос о сбалансированности корпуса. Социолингвистическая, стилистическая и другие виды языковой вариативности в корпусе. Статистические методы измерения вероятности в корпусе. Статистические критерии, применяемые к корпусному материалу. Частотные словари на базе корпусов.

### **Статистические законы и язык**

Вероятностные закономерности в языке. Понятие языковой и текстовой вероятностей. Норма, узус и языковая вероятность. Вероятностные закономерности на фонетическом уровне. Вероятностные закономерности на лексическом уровне. Вероятностные закономерности на грамматическом уровне. Статистические законы на уровне текста. Применение статистических критериев при исследовании генеалогических и типологических отношений языков.

## **6. Методические указания для обучающихся по освоению дисциплины**

Освоение дисциплины требует систематического изучения всех тем в той последовательности, в какой они указаны в рабочей программе.

Основными видами учебной работы являются аудиторные занятия. Их цель - расширить базовые знания обучающихся по осваиваемой дисциплине и систему теоретических ориентиров для последующего более глубокого освоения программного материала в ходе самостоятельной работы. Обучающемуся важно помнить, что контактная работа с преподавателем эффективно помогает ему овладеть программным материалом благодаря расстановке необходимых акцентов и удержанию внимания интонационными модуляциями голоса, а также подключением аудио-визуального механизма восприятия информации.

Самостоятельная работа преследует следующие цели:

- закрепление и совершенствование теоретических знаний, полученных на лекционных занятиях;
- формирование навыков подготовки текстовой составляющей информации учебного и научного назначения для размещения в различных информационных системах;
- совершенствование навыков поиска научных публикаций и образовательных ресурсов, размещенных в сети Интернет;
- самоконтроль освоения программного материала.

Обучающемуся необходимо помнить, что результаты самостоятельной работы контролируются преподавателем во время проведения мероприятий текущего контроля и учитываются при промежуточной аттестации.

Обучающимся с ОВЗ и инвалидов предоставляется возможность выбора форм проведения мероприятий текущего контроля, альтернативных формам, предусмотренным рабочей программой дисциплины. Предусматривается возможность увеличения в пределах 1 академического часа времени, отводимого на выполнение контрольных мероприятий.

Процедура оценивания результатов обучения инвалидов и лиц с ограниченными возможностями здоровья по дисциплине предусматривает предоставление информации в формах, адаптированных к ограничениям их здоровья и восприятия информации.

При проведении текущего контроля применяются оценочные средства, обеспечивающие передачу информации, от обучающегося к преподавателю, с учетом психофизиологических особенностей здоровья обучающихся.

## **7. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине**

При самостоятельной работе обучающимся следует использовать:

- конспекты лекций;
- литературу из перечня основной и дополнительной учебной литературы, необходимой для освоения дисциплины (модуля);
- текст лекций на электронных носителях;
- ресурсы информационно-телекоммуникационной сети "Интернет", необходимые для освоения дисциплины;
- лицензионное и свободно распространяемое программное обеспечение из перечня информационных технологий, используемых при осуществлении образовательного процесса по дисциплине;
- методические указания для обучающихся по освоению дисциплины.

## **8. Перечень основной и дополнительной учебной литературы**

### **Основная:**

1. Семенов, В. А. Математические методы в гуманитарных исследованиях : учебное пособие / В. А. Семенов, В. А. Макаридина. — Москва : Ай Пи Ар Медиа, 2020. — 273 с. — ISBN 978-5-4497-0485-6. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. <http://www.iprbookshop.ru/93993.html>
2. Белоусов К. И.,Блазнова Н. А. Введение в экспериментальную лингвистику:учебное пособие/К. И. Белоусов, Н. А. Блазнова.-Москва:Флинта,2005, ISBN 5-89349-852-6.-136.-Библиогр.: с. 130-136

### **Дополнительная:**

1. Титов, А. Н. Основы математической статистики : учебно-методическое пособие / А. Н. Титов, Р. Ф. Тазиева. — Казань : Издательство КНИТУ, 2022. — 96 с. — ISBN 978-5-7882-3160-0. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. <http://www.iprbookshop.ru/129247.html>
2. Непейвода Н. Н. Прикладная логика:учебное пособие для студентов вузов, обучающихся по специальностям "Математика", "Прикладная математика", "Лингвистика", "Философия" и "Психология"/Н. Н. Непейвода.-Ижевск:Издательство Удмуртского университета,1997, ISBN 5-7029-0074-X.-385.
3. Захаров, В. П. Корпусная лингвистика : учебник для студентов гуманитарных вузов / В. П. Захаров, С. Ю. Богданова. — Иркутск : Иркутский государственный лингвистический университет, 2011. — 161 с. — ISBN 978-5-88267-316-0. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. <http://www.iprbookshop.ru/21088>
4. Зубов А. В.,Зубова И. И. Информационные технологии в лингвистике:учебное пособие для студентов вузов, обучающихся по специальности 021800 - Теоретическая и прикладная лингвистика/А. В. Зубов, И. И. Зубова.-Москва:Академия,2004, ISBN 5-7695-1531-7.-208.-Библиогр.: с. 191-204

## **9. Перечень ресурсов сети Интернет, необходимых для освоения дисциплины**

При освоении дисциплины использование ресурсов сети Интернет не предусмотрено.

## **10. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине**

Образовательный процесс по дисциплине **Математическая лингвистика** предполагает использование следующего программного обеспечения и информационных справочных систем:

1. презентационные материалы (слайды по темам лабораторных занятий);
2. доступ в режиме on-line в Электронную библиотечную систему (ЭБС);
3. доступ в электронную информационно-образовательную среду университета;
4. Интернет-сервисы и электронные ресурсы (македонский интернет портал: <http://www.macedonia.eu.org/>).

Перечень необходимого лицензионного и (или) свободно распространяемого программного обеспечения:

1. офисный пакет приложений;
2. приложение позволяющее просматривать и воспроизводить медиаконтент PDF-файлов;
3. программы, демонстрации видео материалов (проигрыватель);
4. программа просмотра интернет контента (браузер).

При освоении материала и выполнения заданий по дисциплине рекомендуется использование материалов, размещенных в Личных кабинетах обучающихся ЕТИС ПГНИУ (**[student.psu.ru](http://student.psu.ru)**).

При организации дистанционной работы и проведении занятий в режиме онлайн могут использоваться:

- система видеоконференцсвязи на основе платформы BigBlueButton (<https://bigbluebutton.org/>).
- система LMS Moodle (<http://e-learn.psu.ru/>), которая поддерживает возможность использования текстовых материалов и презентаций, аудио- и видеоконтент, а так же тесты, проверяемые задания, задания для совместной работы.
- система тестирования Indigo (<https://indigotech.ru/>).

## **11. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине**

Проведение лекционных занятий требует использование учебных аудиторий, оснащенных мультимедийным проектором. Практические занятия и самостоятельная работа требуют использования печатных и электронных изданий художественной и учебной литературы из фондов Научной библиотеки ПГНИУ.

Помещения научной библиотеки ПГНИУ для обеспечения самостоятельной работы обучающихся:

1. Научно-библиографический отдел, корп.1, ауд. 142. Оборудован 3 персональными компьютера с доступом к локальной и глобальной компьютерным сетям.
2. Читальный зал гуманитарной литературы, корп. 2, ауд. 418. Оборудован 7 персональными компьютерами с доступом к локальной и глобальной компьютерным сетям.
3. Читальный зал естественной литературы, корп.6, ауд. 107а. Оборудован 5 персональными компьютерами с доступом к локальной и глобальной компьютерным сетям.
4. Отдел иностранной литературы, корп.2 ауд. 207. Оборудован 1 персональным компьютером с доступом к локальной и глобальной компьютерным сетям.
5. Библиотека юридического факультета, корп.9, ауд. 4. Оборудована 11 персональными

компьютерами с доступом к локальной и глобальной компьютерным сетям.

6. Читальный зал географического факультета, корп.8, ауд. 419. Оборудован 6 персональными компьютерами с доступом к локальной и глобальной компьютерным сетям.

Все компьютеры, установленные в помещениях научной библиотеки, оснащены следующим программным обеспечением:

Операционная система ALT Linux;

Офисный пакет Libreoffice.

Справочно-правовая система «КонсультантПлюс»

**Фонды оценочных средств для аттестации по дисциплине  
Математическая лингвистика**

**Планируемые результаты обучения по дисциплине для формирования компетенции.  
Индикаторы и критерии их оценивания**

**ОПК.4**

**Способен к ведению профессиональной деятельности с опорой на основы математических дисциплин, необходимых для формализации лингвистических знаний и процедур анализа и синтеза лингвистических структур**

| <b>Компетенция<br/>(индикатор)</b>                                                                                                                              | <b>Планируемые результаты<br/>обучения</b>                                                                                                        | <b>Критерии оценивания результатов<br/>обучения</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                          |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>ОПК.4.1</b><br>демонстрирует знание основ математических дисциплин                                                                                           | Знает основы математических дисциплин; умеет применять их в своей профессиональной деятельности; владеет понятийным аппаратом дисциплины.         | <p><b>Неудовлетворитель</b><br/>Не знает основ математических дисциплин; не умеет применять их в своей профессиональной деятельности; не владеет понятийным аппаратом дисциплины.</p> <p><b>Удовлетворительн</b><br/>Поверхностно знает основы математических дисциплин; умеет под руководством преподавателя применять их в своей профессиональной деятельности, но может допускать грубые ошибки; владеет некоторыми терминами дисциплины.</p> <p><b>Хорошо</b><br/>В целом знает основы математических дисциплин; умеет при консультативной помощи преподавателя применять их в своей профессиональной деятельности, но может допускать ошибки; в целом владеет понятийным аппаратом дисциплины.</p> <p><b>Отлично</b><br/>Знает основы математических дисциплин; умеет самостоятельно применять их в своей профессиональной деятельности, но может допускать отдельные неточности; уверенно владеет понятийным аппаратом дисциплины.</p> |
| <b>ОПК.4.2</b><br>применяет знание основ математических дисциплин для формализации лингвистических знаний и процедур анализа и синтеза лингвистических структур | знает основы математических дисциплин; умеет формализовать лингвистические знания; владеет процедурами анализа и синтеза лингвистических структур | <p><b>Неудовлетворитель</b><br/>не знает основ математических дисциплин; не умеет формализовать лингвистические знания; не владеет процедурами анализа и синтеза лингвистических структур</p> <p><b>Удовлетворительн</b><br/>знает некоторые положения основ математических дисциплин; умеет формализовать некоторые лингвистические знания; владеет некоторыми процедурами</p>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                              |

| Компетенция<br>(индикатор) | Планируемые результаты<br>обучения | Критерии оценивания результатов<br>обучения                                                                                                                                                                                                                                                                                                                                                                                                                                        |
|----------------------------|------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                            |                                    | <p><b>Удовлетворительн</b><br/>анализа и синтеза лингвистических структур</p> <p><b>Хорошо</b><br/>знает основы математических дисциплин;<br/>умеет формализовать базовые<br/>лингвистические знания; владеет основными<br/>процедурами анализа и синтеза<br/>лингвистических структур</p> <p><b>Отлично</b><br/>знает основы математических дисциплин;<br/>умеет формализовать лингвистические<br/>знания; владеет процедурами анализа и<br/>синтеза лингвистических структур</p> |

## Оценочные средства текущего контроля и промежуточной аттестации

Схема доставки : СУОС

**Вид мероприятия промежуточной аттестации :** Экзамен

**Способ проведения мероприятия промежуточной аттестации :** Оценка по дисциплине в рамках промежуточной аттестации определяется на основе баллов, набранных обучающимся на контрольных мероприятиях, проводимых в течение учебного периода.

**Максимальное количество баллов :** 100

### Конвертация баллов в отметки

«отлично» - от 81 до 100

«хорошо» - от 61 до 80

«удовлетворительно» - от 48 до 60

«неудовлетворительно» / «незачтено» менее 48 балла

| Компетенция<br>(индикатор)                                                                                                                                                                                                                                 | Мероприятие<br>текущего контроля                                       | Контролируемые элементы<br>результатов обучения                                                                                                                     |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>ОПК.4.2</b><br>применяет знание основ<br>математических дисциплин для<br>формализации лингвистических<br>знаний и процедур анализа и<br>синтеза лингвистических<br>структур<br><b>ОПК.4.1</b><br>демонстрирует знание основ<br>математических дисциплин | Онтологии и тезаурусы<br><b>Письменное контрольное<br/>мероприятие</b> | Умение отличить тезаурус от толкового<br>словаря. Знание основных<br>семантических отношений,<br>используемых в тезаурусах. Знание<br>основных элементов онтологий. |
| <b>ОПК.4.2</b><br>применяет знание основ<br>математических дисциплин для<br>формализации лингвистических<br>знаний и процедур анализа и<br>синтеза лингвистических<br>структур<br><b>ОПК.4.1</b><br>демонстрирует знание основ<br>математических дисциплин | Равносильность формул<br><b>Письменное контрольное<br/>мероприятие</b> | Знать и уметь определять логику<br>предикатов, интерпретировать<br>логические выражения и знать<br>равносильность формул.                                           |

| Компетенция<br>(индикатор)                                                                                                                                                                                                               | Мероприятие<br>текущего контроля                                                               | Контролируемые элементы<br>результатов обучения                                                                       |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------|
| <b>ОПК.4.2</b><br>применяет знание основ математических дисциплин для формализации лингвистических знаний и процедур анализа и синтеза лингвистических структур<br><b>ОПК.4.1</b><br>демонстрирует знание основ математических дисциплин | Порождающие грамматики. Категориальные грамматики<br><b>Письменное контрольное мероприятие</b> | Знание способов создания формальных языков.                                                                           |
| <b>ОПК.4.2</b><br>применяет знание основ математических дисциплин для формализации лингвистических знаний и процедур анализа и синтеза лингвистических структур<br><b>ОПК.4.1</b><br>демонстрирует знание основ математических дисциплин | Статистические методы в корпусных исследованиях<br><b>Письменное контрольное мероприятие</b>   | Умение применять статистические методы в экспериментальных лингвистических исследованиях и в корпусных исследованиях. |

### Спецификация мероприятий текущего контроля

#### Онтологии и тезаурусы

Продолжительность проведения мероприятия промежуточной аттестации: **2 часа**

Условия проведения мероприятия: **в часы самостоятельной работы**

Максимальный балл, выставляемый за мероприятие промежуточной аттестации: **25**

Проходной балл: **12**

| Показатели оценивания                                              | Баллы |
|--------------------------------------------------------------------|-------|
| Выполнение письменного задания. За каждую ошибку снимается 1 балл. | 25    |

#### Равносильность формул

Продолжительность проведения мероприятия промежуточной аттестации: **2 часа**

Условия проведения мероприятия: **в часы самостоятельной работы**

Максимальный балл, выставляемый за мероприятие промежуточной аттестации: **25**

Проходной балл: **12**

| Показатели оценивания                                              | Баллы |
|--------------------------------------------------------------------|-------|
| Выполняется письменное задание. За каждую ошибку снимается 1 балл. | 25    |

#### Порождающие грамматики. Категориальные грамматики

Продолжительность проведения мероприятия промежуточной аттестации: **2 часа**

Условия проведения мероприятия: **в часы самостоятельной работы**

Максимальный балл, выставляемый за мероприятие промежуточной аттестации: **25**

Проходной балл: **12**

| Показатели оценивания                                              | Баллы |
|--------------------------------------------------------------------|-------|
| Выполняется письменное задание. За каждую ошибку снимается 1 балл. | 25    |

### **Статистические методы в корпусных исследованиях**

Продолжительность проведения мероприятия промежуточной аттестации: **2 часа**

Условия проведения мероприятия: **в часы самостоятельной работы**

Максимальный балл, выставляемый за мероприятие промежуточной аттестации: **25**

Проходной балл: **12**

| Показатели оценивания                                              | Баллы |
|--------------------------------------------------------------------|-------|
| Выполняется письменное задание. За каждую ошибку снимается 1 балл. | 25    |