

УТВЕРЖДАЮ  
 Декан факультета

\_\_\_\_\_  
 (подпись) Шматко А. Д.  
 ФИО  
 «\_\_\_» \_\_\_\_\_ 20\_\_

## РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ ФОРМАЛЬНЫЕ МОДЕЛИ В ЛИНГВИСТИКЕ

Направление/специальность подготовки	45.03.03 Фундаментальная и прикладная лингвистика
Специализация/профиль/программа подготовки	Лингвистика в современной информационно - коммуникационной цифровой среде
Уровень высшего образования	Бакалавриат
Форма обучения	Очная
Факультет	Р Международного промышленного менеджмента и коммуникации
Выпускающая кафедра	Р7 ТЕОРЕТИЧЕСКАЯ И ПРИКЛАДНАЯ ЛИНГВИСТИКА
Кафедра-разработчик рабочей программы	Р7 ТЕОРЕТИЧЕСКАЯ И ПРИКЛАДНАЯ ЛИНГВИСТИКА

КУРС	СЕМЕСТР	ОБЩАЯ ТРУДОЁМКОСТЬ (ЗАЧЕТНЫХ ЕДИНИЦ)	ЧАСЫ (по наличию видов занятий)									ВИД ПРОМЕЖУТОЧНОГО КОНТРОЛЯ
			ОБЩАЯ ТРУДОЁМКОСТЬ	АУДИТОРНЫЕ ЗАНЯТИЯ				САМОСТОЯТЕЛЬНАЯ РАБОТА				
				ВСЕГО	ЛЕКЦИИ	ЛАБОРАТОРНЫЙ ПРАКТИКУМ	ПРАКТИЧЕСКИЕ ЗАНЯТИЯ	ВСЕГО	КУРСОВОЙ ПРОЕКТ	КУРСОВАЯ РАБОТА	ДРУГИЕ ВИДЫ САМОСТ. РАБОТЫ	
4	8	3	108	39	26	0	13	69	0	0	69	диф. зач.

*ЛИСТ СОГЛАСОВАНИЯ*

**РАБОЧАЯ ПРОГРАММА СОСТАВЛЕНА В СООТВЕТСТВИИ С ТРЕБОВАНИЯМИ ФЕДЕРАЛЬНОГО  
ГОСУДАРСТВЕННОГО ОБРАЗОВАТЕЛЬНОГО СТАНДАРТА ВЫСШЕГО ОБРАЗОВАНИЯ (ФГОС ВО)**

**45.03.03 Фундаментальная и прикладная лингвистика**

год набора группы: 2024

Программу составил:

Кафедра Р7 ТЕОРЕТИЧЕСКАЯ И ПРИКЛАДНАЯ ЛИНГВИСТИКА \_\_\_\_\_  
Мамаев Иван Дмитриевич, преподаватель

Программа рассмотрена  
на заседании кафедры-разработчика  
рабочей программы **Р7 ТЕОРЕТИЧЕСКАЯ И ПРИКЛАДНАЯ ЛИНГВИСТИКА**

Заведующий кафедрой Невзорова Г.Д., к.ф.н., доц. \_\_\_\_\_

Программа рассмотрена  
на заседании выпускающей кафедры

**Р7 ТЕОРЕТИЧЕСКАЯ И ПРИКЛАДНАЯ ЛИНГВИСТИКА**

Заведующий кафедрой Невзорова Г.Д., к.ф.н., доц. \_\_\_\_\_

# **РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ ФОРМАЛЬНЫЕ МОДЕЛИ В ЛИНГВИСТИКЕ**

## **Разделы рабочей программы**

1. ЦЕЛИ ОСВОЕНИЯ ДИСЦИПЛИНЫ
2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ООП ВО
3. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ
4. ФОРМЫ КОНТРОЛЯ ОСВОЕНИЯ ДИСЦИПЛИНЫ
5. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ
6. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

## **Приложения к рабочей программе дисциплины**

- Приложение 1. Аннотация рабочей программы
- Приложение 2. Технологии и формы обучения
- Приложение 3. Фонды оценочных средств

## 1. ЦЕЛИ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Целью освоения дисциплины является формирование следующих компетенций:

ОПК-1 — способность применять в профессиональной деятельности основные понятия и категории современной лингвистики
ОПК-2 — способность к ведению профессиональной деятельности с опорой на основы математических дисциплин, необходимых для формализации лингвистических знаний и процедур анализа и синтеза лингвистических структур
ПСК-3.1 — способность анализировать материалы современных исследований в области лингвистики, межкультурной коммуникации и переводоведения для их самостоятельного использования
ПСК-3.3 — способность проводить лингвистический анализ текста и создавать лингвопереводческий и лингвострановедческий комментарий к нему

Формированию компетенций служит достижение следующих результатов образования:

### **ОПК-1**

*знания:*

знать основные формальные лингвистические модели;

*умения:*

уметь оперировать основными формальными лингвистическими моделями;

*навыки:*

обладать навыком использования математических дисциплин, которые необходимы для формального представления лингвистических знаний.

### **ОПК-2**

*знания:*

знать основные положения и концепции в области теоретической лингвистики и основного изучаемого языка (языков), теории коммуникации, лингвистического анализа и интерпретации текста;

*умения:*

уметь обсуждать перспективные направления современной лингвистики;

*навыки:*

владеть базовыми навыками анализа языковых фактов с использованием традиционных методов и современных информационных технологий.

### **ПСК-3.1**

*знания:*

знать способы анализа современного материала в области лингвистики, межкультурной коммуникации и переводоведения;

*умения:*

уметь анализировать современный материал в области лингвистики, межкультурной коммуникации и переводоведения;

*навыки:*

обладать навыком сбора, количественного и качественного анализа лингвистического материала.

### **ПСК-3.3**

*знания:*

знать основные способы применения формальных моделей в области переводоведения;

*умения:*

уметь обобщать формальную и содержательную стороны лингвистической информации;

*навыки:*

обладать навыком создания лингвистического комментария по выбранной проблеме.

## 2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ООП ВО

Дисциплина **ФОРМАЛЬНЫЕ МОДЕЛИ В ЛИНГВИСТИКЕ** является дисциплиной **обязательной части блока 1** программы подготовки по направлению *45.03.03 Фундаментальная и прикладная лингвистика*.

Содержание дисциплины является логическим продолжением дисциплин: **ИНФОРМАЦИОННЫЕ ТЕХНОЛОГИИ И ПРОГРАММИРОВАНИЕ, СИНТАКСИС, ОБЩАЯ МОРФОЛОГИЯ, ТЕХНОЛОГИЯ ОБРАБОТКИ ТЕКСТА И ЗВУЧАЩЕЙ РЕЧИ**.

Предварительные компетенции, сформированные у обучающегося до начала изучения дисциплины:

- ОПК-1 — Способен применять в профессиональной деятельности основные понятия и категории современной лингвистики
- ОПК-7 — Способен понимать принципы работы современных информационных технологий и использовать их для решения задач профессиональной деятельности
- ПК-94 — способен к управлению информацией и данными, поиску источников информации и данных, восприятию, анализу, запоминанию и передаче информации с использованием цифровых средств, а также с помощью алгоритмов при работе с полученными из различных источников данными с целью эффективного использования полученной информации для решения задач
- ПК-95 — способен к критическому мышлению в цифровой среде, оценке информации, ее достоверности, построению логических умозаключений на основании поступающих информации и данных

### 3. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

Общая трудоемкость дисциплины составляет 3 з.е., 108 ч.

#### 3.1. Содержание (дидактика) дисциплины

КУРС	СЕМЕСТР	Наименование разделов и дидактических единиц	ВСЕГО	Аудиторные занятия в контактной форме			Самостоятельная работа студентов	Формируемая компетенция, %			
				ВСЕГО	Лекции	Практические занятия		ОПК-1	ОПК-2	ПСК-3.1	ПСК-3.3
4	8	<b>Раздел 1. Моделирование в естественном языке.</b> Лингвистические модели и их типы. Модели синтезирующие и аналитические. Моделирование различных аспектов речевой деятельности. Модели различных уровней языка и речи (фонетические, морфосинтаксические, семантические). Использование лингвистических моделей в системах автоматической обработки естественного языка.	14	4	4	0	10	20	20	20	20
4	8	<b>Раздел 2. Синтаксические структуры составляющих и синтаксические структуры зависимостей.</b> Грамматика зависимостей. Конструкции, представляющие трудность для структуры зависимостей. Понятие структуры составляющих. Структурные отношения между составляющими. Сопоставление структуры зависимостей и структуры составляющих. Синтаксическая омонимия: стрелочная омонимия, конституентная омонимия.	18	8	4	4	10	20	20	20	20
4	8	<b>Раздел 3. Статистико-комбинаторные модели в лингвистике.</b> Направления компьютерной и прикладной лингвистики, которые используют статистические методы: машинный перевод, автоматическое реферирование, извлечение коллокаций из текстов и др. Статистический аппарат в корпусах текстов: обзор используемых метрик и программных средств.	18	8	4	4	10	20	20	20	20
4	8	<b>Раздел 4. Компьютерные модели перевода.</b> История машинного перевода. Три подхода к машинному переводу: rule-based approach, statistical approach, neural approach. Лингвистические проблемы машинного перевода. Понятие постредктирования текстов машинного перевода: требования и стандарты.	14	4	4	0	10	20	20	20	20
4	8	<b>Раздел 5. Компьютерные семантические модели.</b> Машинное обучение в лингвистической науке. Тематическое моделирование текстов на естественном языке. Онтологии как способ формального представления лингвистической предметной области. Классификация и кластеризация текстовых данных.	44	15	10	5	29	20	20	20	20
Всего за 8 семестр			108	39	26	13	69	100	100	100	100
Всего по дисциплине			108	39	26	13	69	100	100	100	100

#### 3.2. Аудиторный практикум

№ п/п	Номер и наименование раздела дисциплины	Тема практического занятия	Объем, ауд. часов
1	Раздел 2. Синтаксические структуры составляющих и синтаксические структуры зависимостей.	Построение грамматик зависимостей и непосредственно составляющих для предложений естественного языка. Анализ систем автоматизированного создания подобных грамматик.	4
2	Раздел 3. Статистико-комбинаторные модели в лингвистике.	Формулы, показывающие меры устойчивости коллокаций. Решение задач.	4
3	Раздел 5. Компьютерные семантические модели.	Сбор корпуса и анализ полученных тематических моделей.	5
<b>Всего за 8 семестр</b>			13

#### 3.3. Самостоятельная работа студента (СРС)

№ п/п	Номер и наименование раздела дисциплины	Содержание учебного задания	Объем, часов
1	Раздел 1. Моделирование в естественном языке.	Самостоятельное изучение формальных моделей в лингвистике по современным научным работам.	10

2	Раздел 2. Синтаксические структуры составляющих и синтаксические структуры зависимостей.	Ручное и автоматическое создание синтаксических структур отобранных для анализа предложений на естественном языке.	10
3	Раздел 3. Статистико-комбинаторные модели в лингвистике.	Знакомство с методами автоматического извлечения коллокаций из текстов.	10
4	Раздел 4. Компьютерные модели перевода.	Работа с современными системами машинного перевода.	10
5	Раздел 5. Компьютерные семантические модели.	Создание собственных тематических моделей для отобранных текстовых коллекций.	29
<b>Всего за 8 семестр</b>			<b>69</b>

#### 4. ФОРМЫ КОНТРОЛЯ ОСВОЕНИЯ ДИСЦИПЛИНЫ

СЕМЕСТР	НЕДЕЛИ СЕМЕСТРА												
	1	2	3	4	5	6	7	8	9	10	11	12	13
<b>8</b>		Дисск.			ДЗ	ДР	ДЗ		Дисск.	ДР		ДЗ	диф. зач.

Условные обозначения:

- ДР – диагностическая работа;
- Дисск. – дискуссия;
- ДЗ – домашнее задание;
- диф. зач. – дифференцированный зачет.

**Текущий контроль успеваемости** студентов проводится в дискретные временные интервалы в следующих формах:

- диагностическая работа;
- дискуссия;
- домашнее задание.

**Промежуточная аттестация** проводится в формах:

- дифференцированный зачет.

## 5. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

### 5.1. Основная литература по дисциплине:

1. А. Н. Гуцин. . Технология обработки текста и звучащей речи. СПб.БГТУ "ВОЕНМЕХ" им. Д. Ф. Устинова, 2018, 66 экз.
2. Э. Алпайдин. . Машинное обучение: новый искусственный интеллект. М.: Альпина Паблишер, 2017, 7 экз.

### 5.2. Дополнительная литература по дисциплине:

1. А. Н. Баранов. . Введение в прикладную лингвистику. М.: Эдиториал УРСС, 2003, 2 экз.
2. А. Н. Баранов. . Введение в прикладную лингвистику. М.: Эдиториал УРСС, 2001, 1 экз.

### 5.3. Периодические издания:

не требуются.

### 5.4. Перечень ресурсов информационно-телекоммуникационной сети "Интернет", необходимых для освоения дисциплины, электронные библиотечные системы:

1. <https://urait.ru/> — Образовательная платформа «Юрайт». Для вузов и ссузов;
2. <https://e.lanbook.com/> — ЭБС Лань;
3. <https://scholar.google.com/> — Google Академия;
4. <http://library.voenmeh.ru/jirbis2> — Сайт фундаментальной библиотеки БГТУ «Военмех» им. Д.Ф. Устинова — Фундаментальная библиотека БГТУ «ВОЕНМЕХ» им. Д.Ф. Устинова;
5. <http://stanza.run/> — Stanza;
6. <https://mimno.infosci.cornell.edu/jsLDA/> — jsLDA: In-browser topic modeling.

### Современные профессиональные базы данных:

1. <https://rusneb.ru> – Национальная электронная библиотека (НЭБ);
2. <https://cyberleninka.ru/> - Научная электронная библиотека «Киберленинка»;
- <http://www.rfbr.ru/rffi/ru/library> - Полнотекстовая электронная библиотека Российского фонда фундаментальных исследований.

### Информационные справочные системы:

1. Техэксперт – Информационный портал технического регулирования: Нормы, правила, стандарты РФ;
2. [http://library.voenmeh.ru/jirbis2/index.php?option=com\\_irbis&view=irbis&Itemid=457](http://library.voenmeh.ru/jirbis2/index.php?option=com_irbis&view=irbis&Itemid=457) - БД ГОСТов собственной генерации БГТУ "ВОЕНМЕХ" им. Д. Ф. Устинова;
3. <http://www.consultant.ru/>- КонсультантПлюс- информационный портал правовой информации.

### 5.5. Программное обеспечение:

не требуется.

### 5.6. Информационные технологии:

взаимодействие с обучающимися посредством ЭИОС Moodle БГТУ «ВОЕНМЕХ» им. Д.Ф. Устинова.

## **6. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ**

### **6.1. Лекционные занятия:**

специализированные требования по оборудованию отсутствуют; аудитория с посадочными местами по количеству студентов; доска.

### **6.2. Практические занятия:**

1. Проектор;
2. Интерактивная доска.

### **6.3. Прочее:**

1. рабочее место преподавателя, оснащенное компьютером с доступом в Интернет;
2. рабочие места студентов, оснащенные компьютерами с доступом в Интернет, предназначенные для работы в электронной образовательной среде.

### **Аннотация рабочей программы**

Дисциплина **ФОРМАЛЬНЫЕ МОДЕЛИ В ЛИНГВИСТИКЕ** является дисциплиной **обязательной части блока 1** программы подготовки по направлению 45.03.03 *Фундаментальная и прикладная лингвистика*. Дисциплина реализуется на факультете Р Международного промышленного менеджмента и коммуникации БГТУ "ВОЕНМЕХ" им. Д.Ф. Устинова кафедрой Р7 ТЕОРЕТИЧЕСКАЯ И ПРИКЛАДНАЯ ЛИНГВИСТИКА.

Дисциплина нацелена на формирование *компетенций*:

ОПК-1 способность применять в профессиональной деятельности основные понятия и категории современной лингвистики;

ОПК-2 способность к ведению профессиональной деятельности с опорой на основы математических дисциплин, необходимых для формализации лингвистических знаний и процедур анализа и синтеза лингвистических структур;

ПСК-3.1 способность анализировать материалы современных исследований в области лингвистики, межкультурной коммуникации и переводоведения для их самостоятельного использования;

ПСК-3.3 способность проводить лингвистический анализ текста и создавать лингвопереводческий и лингвострановедческий комментарий к нему.

Содержание дисциплины охватывает круг вопросов, связанных с современными исследованиями в области прикладной, математической и компьютерной лингвистики. Курс содержит начальные сведения о формальных моделях, формальной семантике, истории структурной лингвистики и компьютерной лингвистики.

Программой дисциплины предусмотрены следующие **виды контроля**:

**Текущий контроль успеваемости** студентов проводится в дискретные временные интервалы в следующих формах:

- диагностическая работа;
- дискуссия;
- домашнее задание.

**Промежуточная аттестация** проводится в формах:

- дифференцированный зачет.

Общая трудоемкость освоения дисциплины составляет 3 з.е., **108 ч**. Программой дисциплины предусмотрены лекционные занятия (**26 ч.**), практические занятия (**13 ч.**), самостоятельная работа студента (**69 ч**).

## ТЕХНОЛОГИИ И ФОРМЫ ОБУЧЕНИЯ

### Рекомендации по освоению дисциплины для студента

Трудоемкость освоения дисциплины составляет 108 ч., из них 39 ч. аудиторных занятий, и 69 ч., отведенных на самостоятельную работу студента.

Рекомендации по распределению учебного времени по видам самостоятельной работы и разделам дисциплины приведены в таблице.

Контроль освоения дисциплины производится в соответствии с Положением о текущем, рубежном контроле успеваемости и промежуточной аттестации обучающихся.

Формы контроля и критерии оценивания приведены в приложении 3 к Рабочей программе.

Наименование работы	Рекомендуемая литература	Трудоемкость, час.
<b>Раздел 1. Моделирование в естественном языке.</b>		
Самостоятельное изучение формальных моделей в лингвистике по современным научным работам.	А. Н. Баранов. . Введение в прикладную лингвистику: М.: Эдиториал УРСС, 2003 (1)	10
Итого по разделу 1		10
<b>Раздел 2. Синтаксические структуры составляющих и синтаксические структуры зависимостей.</b>		
Ручное и автоматическое создание синтаксических структур отобранных для анализа предложений на естественном языке.	А. Н. Баранов. . Введение в прикладную лингвистику: М.: Эдиториал УРСС, 2003 (4)	10
Итого по разделу 2		10
<b>Раздел 3. Статистико-комбинаторные модели в лингвистике.</b>		
Знакомство с методами автоматического извлечения коллокаций из текстов.	А. Н. Баранов. . Введение в прикладную лингвистику: М.: Эдиториал УРСС, 2003 (2)	10
Итого по разделу 3		10
<b>Раздел 4. Компьютерные модели перевода.</b>		
Работа с современными системами машинного перевода.	А. Н. Баранов. . Введение в прикладную лингвистику: М.: Эдиториал УРСС, 2001 (4) А. Н. Гущин. . Технология обработки текста и звучащей речи: СПб.БГТУ "ВОЕНМЕХ" им. Д. Ф. Устинова, 2018 (17)	10
Итого по разделу 4		10
<b>Раздел 5. Компьютерные семантические модели.</b>		
Создание собственных тематических моделей для отобранных текстовых коллекций.	Э. Алпайдин. . Машинное обучение: новый искусственный интеллект: М.: Альпина Паблишер, 2017 (1–3)	29
Итого по разделу 5		29

## ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

Фонд оценочных средств, позволяющие оценить результаты обучения по данной дисциплине, включают в себя:

- диагностическая работа
- дискуссия;
- домашнее задание;
- дифференцированный зачет.

### Критерии оценивания

#### Диагностическая работа

Диагностическая работа проводится в форме теста в ЭИОС Moodle:

- при правильном ответе менее чем на 60% вопросов - не аттестация;
- при правильном ответе на 60% вопросов и более - аттестация.

#### Дискуссия

Во время дискуссии студенты и преподаватель обсуждают современные проблемы в области формальных лингвистических моделей. Приблизительный список дискуссионных тем приведён ниже.

1. Синтаксические конструкции зависимостей.
2. Синтаксические конструкции составляющих.
3. Подходы к машинному переводу.
4. Статистические меры анализа устойчивости коллокаций.
5. Семантические модели.

Если во время дискуссии обучающийся всестороннее понимания вопроса, то такое обсуждение может быть зачтено. Если преподаватель не удовлетворён ответом, он может задать студентам дополнительные дискуссионные вопросы.

#### Домашнее задание

Домашнее задание обсуждается с преподавателем перед окончанием обсуждения определённого тематического блока лекционно-практического курса. Домашнее задание может быть оформлено в виде отчёта как в рукописном, так и в печатном формате. При оценке отчёта преподаватель учитывает глубину раскрытия вопроса, а также верность хода решения. Если домашнее задание включало обработку текстовых массивов, то в отчёте также необходимо предоставить листинг программы для обработки или ссылку на используемый для обработки ресурс.

Если обучающийся получил верный ответ, подробно объяснил процесс решения задачи и обосновал выбор того или иного подхода, то домашняя работа может быть зачтена (5 баллов). Наличие не более 2 серьёзных недочётов приводит к снижению общего балла до 4. Если студент допустил от 2 до 4 недочётов, то максимальный общий балл не превышает 3. Если обучающийся неверно решил задачу или же показал поверхностные знания при объяснении процесса решения, то ему рекомендуется решить задачу ещё раз после комментариев преподавателя.

#### Дифференцированный зачет

Дифференцированный зачет выставляется согласно технологической карте, которая составляется с учетом приказа ректора №687-О от 08.09.2023.

Если студент не согласен с предлагаемой оценкой, то он имеет право сдать дифференцированный зачёт в устной форме по вопросам, которые обсуждались в цикле всего курса, а именно: общие положения о формальных моделях в лингвистике, синтаксические формальные модели, модели машинного перевода, статистико-комбинаторные модели в лингвистике, лингвистические онтологии, тематическое моделирование.

Оценка «зачтено-отлично» выставляется обучающемуся, если он показывает всесторонние, систематизированные, глубокие знания вопросов по темам дисциплины, а также умение уверенно применять их на практике.

Оценка «зачтено-хорошо» выставляется обучающемуся, если он твердо знает материал, грамотно и по существу излагает его, умеет применять полученные знания на практике, но допускает в ответе некоторые неточности, которые может устранить с помощью дополнительных вопросов преподавателя.

Оценка «зачтено-удовлетворительно» выставляется обучающемуся, показавшему фрагментарный, разрозненный характер знаний, недостаточно правильные формулировки базовых понятий, нарушения логической последовательности в изложении программного материала, но при этом он владеет основными понятиями выносимых на опрос тем, необходимыми для дальнейшего обучения и может применять полученные знания по образцу в стандартной ситуации.

Оценка «не зачтено» выставляется обучающемуся, который не знает большей части основного содержания вопросов тем дисциплины, допускает грубые ошибки в формулировках основных понятий и не умеет использовать на практике полученные знания.

Паспорт фонда оценочных средств

КУРС	СЕМЕСТР	Наименование разделов и дидактических единиц	ВСЕГО	Аудиторные занятия в контактной форме			Самостоятельная работа студентов	Формируемая компетенция, %				НАИМЕНОВАНИЕ ОЦЕНОЧНОГО СРЕДСТВА
				ВСЕГО	Лекции	Практические занятия		ОПК-1	ОПК-2	ПСК-3.1	ПСК-3.3	
4	8	Раздел 1. Моделирование в естественном языке.	14	4	4	0	10	20	20	20	20	Дискуссия
4	8	Раздел 2. Синтаксические структуры составляющих и синтаксические структуры зависимостей.	18	8	4	4	10	20	20	20	20	Домашнее задание
4	8	Раздел 3. Статистико-комбинаторные модели в лингвистике.	18	8	4	4	10	20	20	20	20	Домашнее задание
4	8	Раздел 4. Компьютерные модели перевода.	14	4	4	0	10	20	20	20	20	Дискуссия
4	8	Раздел 5. Компьютерные семантические модели.	44	15	10	5	29	20	20	20	20	Домашнее задание
Всего за 8 семестр			108	39	26	13	69	100	100	100	100	
Всего по дисциплине			108	39	26	13	69	100	100	100	100	

## Критерии оценивания

### ОПК-1

- Вопросы открытого типа:*
- № 1 Что обозначает свойство экспланаторности в лингвистических моделях?
- № 2 Вставьте одно слово.
- OWL — Ontology Web \_\_\_\_\_.
- № 3 Вставьте одно слово.
- Современная адаптация онтологии WordNet под русский язык, созданная Н.В. Лукашевич, называется \_\_\_\_\_.
- № 4 Вставьте одно слово.
- Онтология задаёт формальную спецификацию, в то время как семантическая \_\_\_\_\_ определяет способ представления знаний.
- № 5 Что такое таксономии?
- № 6 Что такое лингвистическая онтология в общем смысле?
- № 7 Одно из свойств лингвистических моделей можно сформулировать следующим образом: обычно модель оперирует не понятиями о реальных объектах, а конструктами. Что такое конструкты?
- № 8 Что такое дистрибутивная семантика?
- № 9 Закон \_\_\_\_\_ можно сформулировать следующим образом: при упорядочивании всех слов некоторого текста по уменьшению частоты их использования частота случайно выбранного слова будет обратно пропорциональна его рангу.
- № 10 Необходимость в \_\_\_\_\_ возникает во всех тех научных областях, в которых объект науки недоступен непосредственному наблюдению.
- Вопросы закрытого типа:*
- № 1 В порождающих моделях конечной информацией является:
1. текст;
  2. грамматика и словарь;
  3. множество правильных фраз и изображение их структуры;
  4. изображение структуры текста.
- № 2 Лексическая функция Sup обозначает:
1. меронию;
  2. антонимию;
  3. синонимию;
  4. синтаксический дериват.
- № 3 Какова основная цель лингвистических синтетических моделей?
1. Смоделировать деятельность лингвиста;
  2. Смоделировать понимание текста;
  3. Смоделировать умение различать верные и неверные языковые явления;
  4. Смоделировать производство текста.
- № 4 Для «отладки» процесса перевода применяются многочисленные модели сглаживания, одна из которых – \_\_\_\_\_ выравниваний.
1. таблица;
  2. матрица;
  3. кортеж;
  4. столбец.
- № 5 Выберите верные аналоги отношений категоризации в онтологии.
1. IS-A;
  2. PART-OF;

3. A-KIND-OF;  
4. родовидовые отношения.
- № 6 Как называются отдельные представители класса сущностей?
1. аксиомы;
  2. правила;
  3. экземпляры;
  4. нетерминалы.
- № 7 Как называются автоматизированные системы, предназначенные для общения с человеком?
1. диалоговые системы;
  2. дистрибутивные системы;
  3. имитационные системы;
  4. форумные системы.
- № 8 Языковые модели BERT называют:
1. омонимичными;
  2. контекстуализированными;
  3. синонимичными;
  4. прагматическими.
- № 9 К процессам семантической компрессии текстов относится:
- грамматика зависимостей
- выделение именных групп
- тематическое моделирование
- № 10 грамматика непосредственно составляющих
- К формальным моделям синтаксиса относятся:
- грамматика зависимостей
- лемматизация
- тематическое моделирование
- грамматика непосредственно составляющих
- ОПК-2**
- Вопросы открытого типа:*
- № 1 Какое явление продемонстрировано в данном предложении?
- I saw a spider with my binoculars.*
- № 2 Какие модели на выходе имеют множество правильных фраз?
- № 3 Вставьте пропущенное слово.
- Объектом математической лингвистики являются \_\_\_\_\_ языка и речи во всём их многообразии.
- № 4 Вставьте пропущенное слово.
- В задачах автоматической обработки текста (например, при распознавании именованных сущностей) каждая модель оценивается с точки зрения \_\_\_\_\_ и точности.
- № 5 Вставьте пропущенное слово.
- Для вычисления наблюдаемых и ожидаемых частот используются матрицы \_\_\_\_\_.
- № 6 Вставьте пропущенное слово.

- \_\_\_\_\_ — язык-посредник, который может стать полноценным вспомогательным языком при создании систем машинного перевода.
- № 7 Вставьте пропущенное слово.
- И.А. Мельчук разработал модель «\_\_\_\_\_ ↔ Текст».
- № 8 Вставьте пропущенное слово.
- Синтаксическую структуру предложения наглядно можно изобразить в виде \_\_\_\_\_, узлами которого являются словоформы, а дуги выражают отношения между ними.
- № 9 Вставьте пропущенное слово.
- Грамматика зависимостей – формальная модель, представляющая структуру предложения в виде иерархии компонентов, между которыми установлено отношение \_\_\_\_\_.
- № 10 Вставьте пропущенное слово.
- В вербоцентрической теории вершиной предложения является \_\_\_\_\_.
- Вопросы закрытого типа:*
- № 1 Какая из перечисленных метрик устойчивости коллокаций оперирует наблюдаемыми и ожидаемыми частотами?
1. MI;
  2. t-score;
  3. log-likelihood;
  4. нет правильного ответа.
- № 2 К какому типу лингвистических гипотез, согласно Р. Г. Пиотровскому, относится следующая: баскский язык имеет иберо-кавказское происхождение?
1. статистическая;
  2. статическая;
  3. нестатистическая;
  4. динамическая.
- № 3 Формальный язык L – множество конечных \_\_\_\_\_, построенных путем объединения элементов из алфавита по правилам грамматики G, заданной для данного языка.
1. классов;
  2. цепочек;
  3. экземпляров;
  4. букв.
- № 4 В эпоху Аристотеля онтология определялась как:
1. наука о языке;
  2. наука о бытии;
  3. наука об искусстве;
  4. наука о прекрасном.
- № 5 Элемент C в структуре онтологии означает:
1. аксиомы;
  2. терминалы;
  3. отношения;
  4. классы.
- № 6 Сколько видов распределений должно учитываться в стандартной модели LDA?
1. 1;
  2. 2;
  3. 3;
  4. 4.

- № 7 Какую метрику можно использовать для определения сходства между векторными представлениями текстов некоторого корпуса?
1. косинусная мера;
  2. мера вероятностного правдоподобия;
  3. мера агрегации;
  4. нет правильного ответа.
- № 8 Какие онтологии существуют с точки зрения экземпляров?
1. прикладные
  2. общие
  3. безэкземлярные
  4. репрезентационные
- № 9 Термин «непосредственная составляющая» был введен:
1. Леонардом Блумфилдом;
  2. Викторов Захаровым;
  3. Александром Реформатским;
  4. Виктором Виноградовым.
- № 10 \_\_\_\_\_ Grammar for Finite Lattice – AGFL.
1. Axiom;
  2. Affix;
  3. Axis;
  4. Alphabet.

### **ПСК-3.1**

#### *Вопросы открытого типа:*

- № 1 SUMO — Standard \_\_\_\_\_ Merged Ontology.
- № 2 WebVOWL — Web-based \_\_\_\_\_ of Ontologies
- № 3 Circle в редакторе онтологий WebVOWL описывает такой элемент структуры онтологии, как \_\_\_\_\_.
- № 4 Отношение, описывающее равенство двух сущностей, известно как \_\_\_\_\_.
- № 5 Отношение, описывающее подчинение одной сущности другой, известно как \_\_\_\_\_.
- № 6 Отношение, описывающее двустороннюю направленность между сущностями, известно как \_\_\_\_\_.
- № 7 Наличие посредника Z между X и Y. Если X и Z связаны некоторым отношением, которым связаны Y и Z, то тогда X и Y находятся в этих же отношениях. Такое тип отношений известен как \_\_\_\_\_.
- № 8 Отец и дед — кровные родственники, но сын и отец не являются друг другу родственными. Следствие: отношения между дедом и сыном (внуков) имеют свойство \_\_\_\_\_.
- № 9 Прогоулы приводят к увольнению. С точки зрения рефлексивности данный пример можно отнести к типу, известному как \_\_\_\_\_.
- № 10 Ножка — часть стула. С точки зрения симметричности данный пример можно отнести к типу, известному как \_\_\_\_\_.

#### *Вопросы закрытого типа:*

- № 1 Оригинальный WordNet был разработан:
- Джорджем Миллером
- Ноамом Хомским
- Барбарой Парти
- Анной Вежбицкой
- № 2 BalkaNet — онтология ментального лексикона для:

- испанского языка
- бакланских языков
- болгарского языка
- № 3  
русского языка  
Выберите русскоязычную версию WordNet, разработанную на кафедре математической лингвистики СПбГУ.
- YARN
- RuWordNet
- RussNet
- № 4  
RuThes  
Элемент  $T$  в структуре информационно-поискового тезауруса (ИПТ) означает:
- набор терминов некоторой предметной области
- ассоциативные отношения ИПТ
- аксиомы транзитивности иерархических отношений
- иерархические отношения ИПТ
- № 5  
Элемент  $RA$  в структуре информационно-поискового тезауруса (ИПТ) означает:
- набор терминов некоторой предметной области
- ассоциативные отношения ИПТ
- аксиомы транзитивности иерархических отношений
- иерархические отношения ИПТ
- № 6  
Элемент  $RH$  в структуре информационно-поискового тезауруса (ИПТ) означает:
- набор терминов некоторой предметной области
- ассоциативные отношения ИПТ
- аксиомы транзитивности иерархических отношений
- иерархические отношения ИПТ
- № 7  
Элемент  $AT$  в структуре информационно-поискового тезауруса (ИПТ) означает:
- набор терминов некоторой предметной области
- ассоциативные отношения ИПТ
- аксиомы транзитивности иерархических отношений
- иерархические отношения ИПТ
- № 8  
Элемент  $Dth$  в структуре информационно-поискового тезауруса (ИПТ) означает:
- набор терминов некоторой предметной области
- предметную область
- набор дескрипторов предметной области
- иерархические отношения ИПТ
- № 9  
В системе Prolog можно создать онтологию на основе следующего подхода к представлению лингвистических знаний:

логическое представлений знаний

продукционное представление знаний

объектно-ориентированный подход

нейронные сети

№ 10 Для какого языка предназначена онтология AIIRE?

для немецкого языка

для испанского языка

для албанского языка

для русского языка

### ПСК-3.3

*Вопросы открытого типа:*

№ 1 Основным способом выполнения программы на языке Prolog является доказательство некоторой \_\_\_\_\_.

№ 2 Символ запятой при описании сложной цели в среде Prolog — это \_\_\_\_\_.

№ 3 Символ точки с запятой при описании сложной цели в среде Prolog — это \_\_\_\_\_.

№ 4 Символ \+ при описании сложной цели в среде Prolog — это \_\_\_\_\_.

№ 5 Оператор :- в среде Prolog — это \_\_\_\_\_.

№ 6 Prolog — PROgramming \_\_\_\_\_.

№ 7 Одна из основных сфер применения онтологий — Semantic \_\_\_\_\_.

№ 8 До оператора :- в среде Prolog находится \_\_\_\_\_.

№ 9 После оператора :- в среде Prolog находится \_\_\_\_\_.

№ 10 В среде Prolog однократная переменная называется \_\_\_\_\_ variable

*Вопросы закрытого типа:*

№ 1 Декларативные системы отвечают на вопрос:

что мы хотим узнать?

как сделать?

зачем нужно сделать?

почему мы хотим узнать?

№ 2 Императивные системы отвечают на вопрос:

что мы хотим узнать?

как сделать?

зачем нужно сделать?

почему мы хотим узнать?

№ 3 Программа на языке Prolog является совокупностью:

подлежащих

обстоятельств

определений

предикатов

№ 4 Каждый предикат в системе Prolog является совокупностью:

подлежащих

предложений

	высказываний
№ 5	<p>фраза</p> <p>Функтор в системе Prolog — это:</p> <p>имя</p> <p>арность</p> <p>комбинация имени и арности</p>
№ 6	<p>фраза</p> <p>Чему равен функтор высказывания <i>parent(tim, bob)</i>?</p> <p><i>parent/2</i></p> <p><i>parent/1</i></p> <p><i>tim/2</i></p> <p><i>bob/1</i></p>
№ 7	<p>Одноточный комментарий в среде Prolog обозначается символом:</p> <p>#</p> <p>%</p> <p>*</p> <p>&amp;</p>
№ 8	<p>Составной терм в среде Prolog также называется:</p> <p>системой</p> <p>структурой</p> <p>словарём</p> <p>списком</p>
№ 9	<p>Единственный символ нижнего подчеркивания в среде Prolog обозначает:</p> <p>сцепленную переменную</p> <p>конкретизированную переменную</p> <p>числовую переменную</p> <p>анонимную переменную</p>
№ 10	<p>Чем может быть терм в среде Prolog?</p> <p>переменной</p> <p>атомом</p> <p>ссылкой</p> <p>целым числом</p>