

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

**«Прикладные пакеты для лингвистических исследований»**

**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**

**Прикладные пакеты для лингвистических исследований**  
**Рабочая программа дисциплины**

**Составитель:**

**к.ф-м.н. С.А.Шаров**

**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 знакомство с основными программными пакетами, предназначенными для решения практических задач компьютерной лингвистики. Курс дополняет более теоретические дисциплины «Основные алгоритмы лингвистического анализа» и «Машинное обучение» и посвящён практическому применению рассматриваемых там алгоритмов и подходов.

### ***1.2 Цель и задачи курса***

Курс направлен на решение следующих задач:

- познакомить обучающихся с основными программными средствами и библиотеками, предназначенными для решения задач компьютерной лингвистики;
- научить магистрантов как предварительно выбирать программный пакет для решения для прикладных лингвистических задач, так и адаптировать его в зависимости от рассмотренной задачи;
- научить магистрантов преобразовывать исходные данные в формат, подходящий для выбранного программного средства, а также подбирать дополнительные данные для повышения эффективности применяемых моделей.
- научить магистрантов анализировать результаты применения компьютерных программ к лингвистическим данным, выявлять проблемы и предлагать способы их решения;
- выработать у магистрантов знания, позволяющие им квалифицированно читать литературу по специальности, включающую в себя прежде всего техническую документацию, а также научные статьи и учебную литературу.

### ***1.3 Компетенции обучающегося, формируемые в результате освоения дисциплины***

Дисциплина (модуль) направлена на формирование компетенций выпускника:  
*владением знаниями об истории лингвистических учений и современного состояния лингвистической науки (ОПК-3);*  
*способностью к осознанию современного состояния в области компьютерной лингвистики и информационных технологий (ОПК-4);*  
*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*  
*способностью изучать и осваивать современные технические средства и*  
*информационные технологии, служащие для обеспечения лингвистической деятельности (ПК-2);*  
*способностью разрабатывать системы автоматической обработки звучащей речи и*  
*письменного текста на естественном языке, лингвистические компоненты*  
*интеллектуальных и информационных электронных систем (ПК-8);*  
*способностью разрабатывать и внедрять в практику компьютерные системы обучения (ПК-9);*  
*способностью разрабатывать и совершенствовать системы автоматизации и*  
*информационной поддержки лингвистических исследований (ПК-10)*

и соотнесенных с ними результатов освоения дисциплины (модуля):

***Знать:***

- структуру научно-практической области исследований «компьютерная лингвистика» и ее место в контексте смежных наук;
- основные программные пакеты, предназначенные для решения задач компьютерной лингвистики, их достоинства и недостатки;
- алгоритмы и методы, лежащие в основе используемых программных пакетов;
- основные типы лингвистических ресурсов, используемых для получения исходных данных, которые впоследствии используются в программных пакетах для решения задач компьютерной лингвистики;

***Уметь:***

- локализовать практическую задачу в контексте организации научно-практической области исследований «компьютерная лингвистика» и находить средства для ее решения;
- самостоятельно подбирать базовый алгоритм для решения той или иной задачи, а также находить программное обеспечение, реализующее данный алгоритм;
- представлять лингвистические данные в виде, который может быть подан на вход выбранному алгоритму, а также обосновывать выбор того или иного представления;
- устанавливать и настраивать выбранное программное обеспечение в зависимости от особенностей поставленной задачи и технических требований;
- анализировать результаты работы выбранной программы и подбирать оптимальные параметры используемых в ней алгоритмов как результат этой оценки;
- при необходимости сравнивать несколько программных пакетов и/или алгоритмов, проводя необходимые эксперименты для выявления их сходства/различия;
- интегрировать выбранную программу в более обширные программные проекты;
- самостоятельно находить и читать техническую документацию, относящуюся к выбранному программному обеспечению, а также анализировать извлечённую отсюда информацию;

***Владеть:***

- основами методологии проведения научного эксперимента;
- основными программными пакетами для решения задач компьютерной лингвистики;
- навыками чтения литературы по специальности, в том числе технической документации.

#### ***1.4 Место дисциплины в структуре образовательной программы***

Дисциплина (модуль) «Прикладные пакеты для лингвистических исследований» является *дисциплиной по выбору вариативной* части цикла дисциплин ООП ВПО (магистратуры) по направлению подготовки «Фундаментальная и прикладная

лингвистика. Фундаментальная и компьютерная лингвистика» и адресована студентам 2 курса (3 семестр). Дисциплина (модуль) реализуется УНЦ компьютерной лингвистики Института Лингвистики.

Программой дисциплины (модуля) предусмотрены следующие виды контроля: текущий контроль успеваемости в форме: выполнение домашних заданий; лабораторные работы; промежуточные контрольные работы; промежуточная аттестация в форме: зачет.

Общая трудоемкость освоения дисциплины (модуля) составляет 2 зачетные единицы, 72 часа.

Программой дисциплины (модуля) предусмотрены: практические занятия – 20 часов; самостоятельная работа студента – 52 часа.

## 2. Структура дисциплины

### 3. Содержание 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)
				лекции	семинары	самостоятельная работа		
1.	Введение. Краткая история и предмет компьютерной лингвистики (КЛ). Общая структура и основные задачи КЛ. Теоретическая и прикладная КЛ.	1	1			10		ДЗ1. Знакомство с организацией научно-практической области КЛ. Описание термина (по тезаурусу)
2.	Уровни языка в традиционной лингвистике и общие уровневые модели естественного языка (ЕЯ) в КЛ. Задачи анализа и синтеза предложения. Проблема неоднозначности.	1	2		2	10		ДЗ2. Синтаксическая разметка фрагмента текста.
3.	«Понимание» текстов в узкой предметной области. Метод шаблонов. Семантически ориентированный	1	3		2	8		ДЗ3. Тестирование программы Элиза. Выделение видов шаблонов.

	метод анализа.							
4.	Машинный перевод (МП). Схемы МП. Синтаксический анализ (отечественная традиция). Фильтровый метод.	1	4		2	4		ДЗ4. Тестирование систем МП, организованных по разным схемам. Сравнение результатов.
5.	Формализмы синтаксического анализа в англоязычной традиции.	1	5		2	4	4	(Проверка синтаксической разметки)
6.	Корпуса текстов. Распространение эмпирических методов решения задач в КЛ. Метод n-gram – понятие статистической модели языка. Современное состояние МП.	3	6			6		ДЗ5. Исследовательский проект (часть 1): корпусное исследование семантики русского глагола.
7.	Задачи семантики и типы семантических моделей в КЛ, связь с лексикографией. Лексико-семантические базы.	1	7		4	6		ДЗ6. Исследовательский проект (часть 2): Описание значений русского глагола в виде фреймов.
8.	Автоматическая генерация текстов на ЕЯ. Теория риторических структур.	1	8		4	4		ДЗ7. Упражнение на построение модели дискурса конкретного текста в терминах теории риторических отношений.
9.		1	9				2	(Проверка структуры текста. Обсуждение русских фреймов)
10.	Общие модели дискурса. Моделирование стиля и жанра текстов в ресурсах и системах КЛ.	1	10			6		ДЗ8. Исследовательский проект (часть 3): Сопоставление русских фреймов с фреймами базы FrameNet.
11.	Тематический аспект текста и поиск информации. Диалоговые и интерактивные системы.		11		4	4		ДЗ9. Тестирование современных диалоговых систем.
12.	Автоматическая	1	12		2	4		ДЗ10. Оформление

	обработка устной речи и ее приложения.							отчета по исследовательскому проекту.
13.	Обсуждение курсовой работы	1	13			2		
14.	Зачет	1	14			10		Контрольные вопросы
	Итого:				20	52		

#### 4. Образовательные технологии

В соответствии с требованиями ФГОС по направлению 45.04.03 «Фундаментальная и прикладная лингвистика» и с учетом специфики магистерской 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:

1. систематическое использование компьютерных презентаций (как преподавателем в установочной части занятия, так и студентом, выступающим с отчетом по результатам исследования);
2. он-лайн демонстрации работы с лингвистическими интернет-источниками (и др.);
3. он-лайн использование лингвистических ресурсов (Национальный корпус русского языка, Лексико-семантические базы и др.);
4. обсуждения курсовых исследовательских проектов;
5. 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			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, сдает зачет 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>

Количество баллов	Критерии оценки
	<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>
<b>56–67 (D)</b>	<p>Оценка выставляется с учетом текущей и промежуточной аттестации.</p> <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>
<b>50–55 (E)</b>	<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>
<b>21–49 (FX)</b>	<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>

Количество баллов	Критерии оценки
	<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)	<p>Оценка выставляется с учетом текущей и промежуточной 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 (10 заданий – 50 баллов максимум); тестовое задание (максимально 10 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. Знакомство с организацией научно-практической области КЛ. Описание термина (по тезаурусу)
- Д32. Синтаксическая разметка фрагмента текста.
- Д33. Тестирование программы Элиза. Выделение видов шаблонов.
- Д34. Тестирование систем МП, организованных по разным схемам. Сравнение результатов.  
(Проверка синтаксической разметки)
- Д35. Исследовательский проект (часть 1): корпусное исследование семантики русского глагола.
- Д36. Исследовательский проект (часть 2): Описание значений русского глагола в виде фреймов.

- Д37. Упражнение на построение модели дискурса конкретного текста в терминах теории риторических отношений.  
(Проверка структуры текста. Обсуждение русских фреймов)  
Д38. Исследовательский проект (часть 3): Сопоставление русских фреймов с фреймами базы FrameNet.  
Д39. Тестирование современных диалоговых систем.  
Д310. Оформление отчета по исследовательскому проекту.

Зачет ориентирован на следующие контрольные вопросы

КЛ как новая научно-практическая область исследований в контексте смежных наук.  
Задачи теоретической КЛ и приложений.  
Уровни языка и уровневые общие модели языка в КЛ.  
Методы анализа текста.  
Составляющие понятия «понимание». Понятие лингвистического процессора.  
Семантически ориентированный метод анализа текстов.  
Машинный перевод: схемы МП, понятие языка-посредника.  
Синтаксический анализ предложения и фильтровый метод. Различия отечественной и англоязычной традиций в области синтаксического анализа.  
Эволюция формализмов синтаксического анализа в англоязычной традиции.  
Корпусы текстов и распространение эмпирических методов.  
Типы систем МП, основанных на эмпирических методах. Понятие статистической «модели языка».  
Падежные грамматики и семантические примитивы в КЛ. Связь синтаксиса и лексикологии.  
Лексико-семантические базы как лексические ресурсы для систем обработки текстов.  
Моделирование дискурса в ресурсах и системах КЛ.  
Методы автоматической генерации текстов.  
Моделирование параметров текстов (жанр, стиль) в ресурсах и системах КЛ.  
Задачи поиска информации и мера терминологичности слова в тексте.  
Виды поиска информации в КЛ.  
Направления и приложения моделирования устной речи.

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

### **6.1. Список литературы**

Основная литература

7. Николаев И.С., Митренина О.В., Ландо Т. М. Прикладная и компьютерная лингвистика. М.: URSS, 2016.
8. Соколова Е.Г. Синтаксическая разметка в терминах грамматики зависимостей и синтаксических функций [Электронный ресурс]: метод. пособие. М.: РГГУ, 2011. - 33 с. – Библиогр.: с. 33(5 назв.). – Режим доступа: <http://elib.lib.rsuh.ru/elib/000003603.pdf>
9. ACL <https://www.aclweb.org/portal/news-list>
10. NLTK documentation. 2017 <https://www.nltk.org/>
11. Pymystem3 <https://pythonhosted.org/pymystem3/pymystem3.htmlb>
12. R <https://www.rdocumentation.org/packages/base/versions/3.5.1>
13. Sketch engine <https://www.sketchengine.eu/>
14. ГИКРЯ <http://www.webcorpora.ru/>

15. Многоцелевой лингвистический процессор ЭТАП-3  
<http://iitp.ru/ru/science/works/452.htm>
16. НКРЯ <http://www.ruscorpora.ru/corpora-structure.html>

#### Рекомендованная литература

1. Автоматическая обработка текстов на естественном языке и компьютерная лингвистика: учеб. Пособие /Большакова Е.И., Клышинский Э.С., Ландэ Д.В., Носков А.А., Пескова О.В., Ягунова Е.В. – М.: МИЭМ, 2011. -272 с.
2. Болдасов М.В., Соколова Е.Г. Генерация текстов на естественном языке - состояние вопроса и прикладные системы // НТИ, Серия 2, №10, 2005, с.12-22.
3. Виноград Т. Система, понимающая естественный язык. М: Мир, 1976. (Разделы: Образец диалога. С. 21-32; Синтаксис и значение. С. 32-41; Базовый подход к представлению значений. С. 44-47.)
4. Дерюгина О. Программы-собеседники. НТИ серия 1, N 6, стр. 31-35.
5. Жигалов В.А., Соколова Е.Г. InBASE: технология построения ЕЯ интерфейсов к базам данных // Труды Международного семинара Диалог '2001 по компьютерной лингвистике и ее приложениям Том 2, Аксаково, Июнь 2001 с. 123-135. Доступна с сайта: <http://dialog-21.ru/digest/archive/2001/?year=2001&vol=22725&id=6900>
6. Зализняк Анна А. Многозначность и смежные понятия. //Анна А. Зализняк. Многозначность в языке и способы ее представления. М: Языки славянских культур, 2006. Глава 1, 1.1. с. 20-34.
7. Зализняк А.А. Русский грамматический словарь, Изд. 2-е. М.: Рус. словарь, 2003.
8. Искусственный интеллект. Справочник: В 3 кн. Кн. 1-2. М.: Радио и связь, 1990.
9. Кибрик А.А. Анализ дискурса в когнитивной перспективе. [Электронный ресурс] 2003 // [http://www.philol.msu.ru/~otipl/new/main/people/kibrik-aa/files/DA\\_cognitive\\_perspective@Diss\\_2003.pdf](http://www.philol.msu.ru/~otipl/new/main/people/kibrik-aa/files/DA_cognitive_perspective@Diss_2003.pdf)
10. Кобозева И.М. Лингвистическая семантика. Эдиториал УРСС. Москва. 2000. Раздел: Семантические валентности лексемы как семантические отношения, обусловленные ее лексическим значением. С. 140-146.
11. Леонтьева Н.Н. Автоматическое понимание текстов. Системы. Модели. Ресурсы. М.: Academia, 2006.
12. Лингвистический энциклопедический словарь. М., 1990.
13. Тестелец Я.Г. Введение в общий синтаксис. М.: Изд. центр РГГУ, 2001. С. 213-215, 722-747.
14. Шаров С.А. Средства компьютерного представления лингвистической информации //1996. <http://www.ksu.ru/eng/science/ittc/vol000/002/>
15. *Survey of the state of the Art in Human Language Technology* (Ronald Cole, editor in chief) Cambridge University Press. 1997. См. также: (<http://cslu.cse.ogi.edu/HLTsurvey/>).
16. Salton, Gerard. *Automatic text processing. The transformation, analysis, and retrieval of information by computer*. Addison-Wesley Publishing Company, Inc., 1989.
17. Chris Manning and Hinrich Schütze, Chapter 7. Word sense disambiguation. Chapter 8. Lexical acquisition. *Foundations of Statistical Natural Language Processing*, MIT Press. Cambridge, MA: May 1999. Доступна с сайта [http://reslib.com/book/Foundations\\_of\\_Statistical\\_Natural\\_Language\\_Processing](http://reslib.com/book/Foundations_of_Statistical_Natural_Language_Processing)
18. Апресян Ю.Д., Богуславский И.М., Иомдин Л.Л. Лингвистический процессор для сложных информационных систем/ М.: Наука, 1992. 256 с.

19. Баранов А.Н. Введение в прикладную лингвистику. М., 2003. (Разделы: Моделирование общения. С. 20-25, и Моделирование структуры сюжета. С. 25-30.)
20. Болдасов М.В., Соколова Е.Г. Генерация текстов на естественном языке – теории, методы, технологии // НТИ, Серия 2, №7, 2006, с.1-15.
21. Буторов В.Д. Моделирование синтаксиса естественного языка // Прикладное языкознание. Учебник. (ред. А.С.Гердт). СПб., 1996. С. 142-160.
22. Кибрик, А. А. Модус, жанр и другие параметры классификации дискурсов // Вопросы языкознания. - 2009. - N 2. - С. 2-21.
23. Ляшевская О.Н., Кузнецова Ю.Л. Русский фреймнет: к задаче создания корпусного словаря конструкций // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог 2009» (Бекасово, 27-31 мая 2009 г.). Вып. 8 (15). М.: РГГУ, 2009. С. 306-312.
24. Мельчук И.А. Об одной лингвистической модели типа «Смысл-Текст». Уровни представления языковых высказываний // Изв. АН СССР, Серия литературы и языка, 1974. Т. 33, №5-6. <http://feb-web.ru/feb/izvest/1974/05/745-436.htm>
25. Непейвода Н.Н. Квазиискусственный язык // Диалог'2002, Т 1, Москва: Наука. 2002. С.314-318. <http://www.dialog-21.ru/materials/archive.asp?id=7361&y=2002&vol=6077>.
26. Ножов И. Морфогическая и синтаксическая обработка текста (модели и программы)", 2003 год (диссертация). Глава 2. Доступна с сайта <http://aot.ru/technology.html>.
27. Соколова Е.Г. Синтаксическая разметка в терминах грамматики зависимостей и синтаксических функций [Электронный ресурс]: метод. пособие. М.: РГГУ, 2011. - 33 с. – Библиогр.: с. 33(5 назв.). – Режим доступа: <http://elib.lib.rsuh.ru/elib/000003603.pdf>
28. Цибульский В.В., Ежов А.С., Поляков Г.А., Феклистов В.В. Анализ и классификация времени и сроков в российских нормативных и правовых актах. // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог 20012» (в печати).
29. Шенк Р. Обработка концептуальной информации. М.: Энергия, 1980.
30. *The Oxford handbook of computational linguistics* (R. Mitkov ed.) N.Y.: Oxford university press, 2003.
31. Hutchins J. Machine translation: general overview // *The Oxford handbook of computational linguistics* (R. Mitkov ed.) N.Y.: Oxford university press, 2003. P. 501-511.
32. Jurafsky, Daniel, and James H. Martin. 2009. Chapter 10-12 in: *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. 2nd edition. Prentice-Hall. Доступна с сайта <http://lib.mexmat.ru/books/10138>
33. Jurafsky, Daniel, and James H. Martin. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. 2nd edition. Prentice-Hall. 2009.
34. Uszkoreit H. What is computational linguistics? [http://www.coli.uni-saarland.de/~hansu/what\\_is\\_cl.html](http://www.coli.uni-saarland.de/~hansu/what_is_cl.html)

#### Программное обеспечение и Интернет-ресурсы

Учебная дисциплина должна быть обеспечена учебно-методической документацией и материалами. Обязательная литература должна быть представлена в библиотеке ВУЗа, сети Интернет или локальной сети ВУЗа (факультета). Для

обучающихся должен быть обеспечен онлайн доступ к Интернет источникам и системам. В частности, в процессе обучения используются следующие Интернет-ресурсы:

#### Интернет-источники

##### **А. Ресурсы:**

- **Русско-английский тезаурус по компьютерной лингвистике.** Подобласть знаний: <http://uniserv.iis.nsk.su/thes/index.php?ent=74>.
- **Национальный корпус русского языка (НКРЯ):** <http://www.ruscorpora.ru>  
Синтаксический подкорпус НКРЯ: <http://www.ruscorpora.ru/search-syntax.html>  
Параллельный подкорпус НКРЯ:
- **Малый Академический Словарь (МАС):** <http://feb-web.ru/feb/mas/mas-abc/default.asp>.
- **Лексико-семантические базы:**  
**WordNet** (онтология значений полнозначных слов английского языка и лексико-семантические парадигматические отношения : <http://wordnet.princeton.edu/> .  
**Corelex:** (типы регулярной полисемии английских существительных) : <http://www.cs.brandeis.edu/~paulb/CoreLex/overview.html>.  
**FrameNet** (значения и лексико-семантические синтагматические отношения английских предикатных слов в виде фреймов. Онтологические отношения фреймов) : <http://framenet.icsi.berkeley.edu/>.

##### **Б. Системы анализа предложения по уровням:**

- морфологический, синтаксический, поверхностный семантический:  
**Dialing:** <http://aot.ru/>
- семантико-синтаксический:  
**HPSG** (Вершинная грамматика непосредственных составляющих): <http://www2.lingsoft.fi/cgi-bin/engcg?snt=Baby%2C+I+love+you.&h=on>.  
Он-лайн демо: <http://erg.emmtee.net>.  
**LFG** (Лексическая Функциональная Грамматика): <http://decentius.aksis.uib.no/logon/xle.xml>

##### **В. Прикладные системы он-лайн:**

- Программы-собеседники:  
**Eliza** («компьютерный психотерапевт») Дж. Вейценбаума: <http://www.manifestation.com/neurotoys/eliza.php3>.
- Вопросно-ответные системы:  
**START** (в MIT): <http://start.csail.mit.edu/>
- Машинный перевод:  
**ПРОМПТ** (коммерческая, прямой): <http://translate.ru>  
**ЭТАПЗ** (экспериментальная, трансфер): <http://proling.iitp.ru/>  
**Dialing Translator** (экспериментальная): <http://aot.ru/cgi-bin/translate.cgi>  
**GOOGLE. Translate** (коммерческая, статистический): [http://translate.google.com/translate\\_t#](http://translate.google.com/translate_t#)
- Анализ и поиск текстовой информации): <http://demo.rco.ru/>  
**RCO** : <http://demo.rco.ru/>.

##### **Г. Модели и теории (на Интернет сайтах):**

- **RST** – (Rhetorical Structure Theory - Теория Риторических Структур): <http://www.sfu.ca/rst/>.
- **UNL** (Universal Networking Language): <http://www.undl.org/>.

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

Занятия по курсу «Прикладные пакеты для лингвистических исследований» 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 знакомство с основными программными

пакетами, предназначенными для решения практических задач компьютерной лингвистики. Курс дополняет более теоретические дисциплины «Основные алгоритмы лингвистического анализа» и «Машинное обучение» и посвящён практическому применению рассматриваемых там алгоритмов и подходов.

Курс направлен на решение следующих задач:

- познакомить обучающихся с основными программными средствами и библиотеками, предназначенными для решения задач компьютерной лингвистики;
- научить магистрантов как предварительно выбирать программный пакет для решения для прикладных лингвистических задач, так и адаптировать его в зависимости от рассмотренной задачи;
- научить магистрантов преобразовывать исходные данные в формат, подходящий для выбранного программного средства, а также подбирать дополнительные данные для повышения эффективности применяемых моделей.
- научить магистрантов анализировать результаты применения компьютерных программ к лингвистическим данным, выявлять проблемы и предлагать способы их решения;
- выработать у магистрантов знания, позволяющие им квалифицированно читать литературу по специальности, включающую в себя прежде всего техническую документацию, а также научные статьи и учебную литературу.

Дисциплина (модуль) направлена на формирование компетенций выпускника: владением знаниями об истории лингвистических учений и современного состояния лингвистической науки (ОПК-3);

способностью к осознанию современного состояния в области компьютерной лингвистики и информационных технологий (ОПК-4);

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

способностью разрабатывать системы автоматической обработки звучащей речи и письменного текста на естественном языке, лингвистические компоненты интеллектуальных и информационных электронных систем (ПК-8);

способностью разрабатывать и внедрять в практику компьютерные системы обучения (ПК-9);

способностью разрабатывать и совершенствовать системы автоматизации и информационной поддержки лингвистических исследований (ПК-10)

и соотношенных с ними результатов освоения дисциплины (модуля):

**Знать:**

- структуру научно-практической области исследований «компьютерная лингвистика» и ее место в контексте смежных наук;
- основные программные пакеты, предназначенные для решения задач компьютерной лингвистики, их достоинства и недостатки;
- алгоритмы и методы, лежащие в основе используемых программных пакетов;

- основные типы лингвистических ресурсов, используемых для получения исходных данных, которые впоследствии используются в программных пакетах для решения задач компьютерной лингвистики;

#### **Уметь:**

- локализовать практическую задачу в контексте организации научно-практической области исследований «компьютерная лингвистика» и находить средства для ее решения;
- самостоятельно подбирать базовый алгоритм для решения той или иной задачи, а также находить программное обеспечение, реализующее данный алгоритм;
- представлять лингвистические данные в виде, который может быть подан на вход выбранному алгоритму, а также обосновывать выбор того или иного представления;
- устанавливать и настраивать выбранное программное обеспечение в зависимости от особенностей поставленной задачи и технических требований;
- анализировать результаты работы выбранной программы и подбирать оптимальные параметры используемых в ней алгоритмов как результат этой оценки;
- при необходимости сравнивать несколько программных пакетов и/или алгоритмов, проводя необходимые эксперименты для выявления их сходства/различия;
- интегрировать выбранную программу в более обширные программные проекты;
- самостоятельно находить и читать техническую документацию, относящуюся к выбранному программному обеспечению, а также анализировать извлечённую отсюда информацию;

#### **Владеть:**

- основами методологии проведения научного эксперимента;
- основными программными пакетами для решения задач компьютерной лингвистики;
- навыками чтения литературы по специальности, в том числе технической документации.

Дисциплина (модуль) «Прикладные пакеты для лингвистических исследований» является *дисциплиной по выбору вариативной* части цикла дисциплин ООП ВПО (магистратуры) по направлению подготовки «Фундаментальная и прикладная лингвистика. Фундаментальная и компьютерная лингвистика» и адресована студентам 2 курса (3 семестр). Дисциплина (модуль) реализуется УНЦ компьютерной лингвистики Института Лингвистики.

Программой дисциплины (модуля) предусмотрены следующие виды контроля: текущий контроль успеваемости в форме: *выполнение домашних заданий; лабораторные работы; промежуточные контрольные работы;* промежуточная аттестация в форме: *зачет.*

Общая трудоемкость освоения дисциплины (модуля) составляет 2 зачетные единицы, 72 часа.

Программой дисциплины (модуля) предусмотрены: практические занятия – 20 часов; самостоятельная работа студента – 52 часа.



***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 2 з.е., 76 ч., в том числе контактная работа обучающихся с преподавателем 20 ч., самостоятельная работа обучающихся 56 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