

МИНОБРНАУКИ РОССИИ



Федеральное государственное бюджетное образовательное учреждение
высшего образования

**«Российский государственный гуманитарный университет»
(ФГБОУ ВО «РГГУ»)**

ИНСТИТУТ ЛИНГВИСТИКИ

Учебно-научный центр компьютерной лингвистики

АВТОМАТИЧЕСКИЙ ПЕРЕВОД

Рабочая программа дисциплины

45.03.03 – Фундаментальная и прикладная лингвистика

Направленность (профиль): «Фундаментальная и прикладная лингвистика»

Академический бакалавриат

Очная форма обучения

РПД адаптирована для лиц с ограниченными возможностями здоровья и инвалидов

Москва 2020

АВТОМАТИЧЕСКИЙ ПЕРЕВОД

Рабочая программа дисциплины

Составитель:

к. филол. н., профессор УНЦ компьютерной лингвистики *Л.Л.Иомдин*

Ответственный редактор:

д. филол. н., проф. *Я.Г. Тестелец*

УТВЕРЖДЕНО

Протокол заседания кафедры ТиПЛ № 9 от 31.05.20

ОГЛАВЛЕНИЕ

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ
1. Пояснительная записка
1.1. Цели и задачи курса
1.2. Формируемые компетенции и ожидаемые результаты обучения
1.3. Место дисциплины в структуре образовательной программы
2. Структура дисциплины
3. Содержание дисциплины
4. Информационные и образовательные технологии
5. Фонд оценочных средств
5.1. Перечень компетенций и этапы их формирования
5.2. Методика оценивания степени сформированности компетенций
5.3. Описание и обоснование шкал и критериев оценки качества работы студента
5.4. Типовые задания, используемые для оценки степени освоенности учебного материала
5.4.1. Контрольные вопросы по теоретическому материалу
5.4.2. Образцы вопросов для мини-тестов
5.4.3. Образец домашнего задания
5.4.4. Образцы коллективных практических заданий
5.4.5. Примеры заданий из промежуточной письменной контрольной
6. Учебно-методическое и информационное обеспечение дисциплины
6.1. Список источников и литературы
6.2. Перечень электронных ресурсов
7. Материально-техническое обеспечение дисциплины
8. Обеспечение образовательного процесса для лиц с ограниченными возможностями здоровья и инвалидов
9. Методические материалы
9.1 Планы практических (семинарских, лабораторных) занятий
9.2 Иные материалы
Приложения
Приложение 1. Аннотация дисциплины
Приложение 2. Лист изменений

Приложения

Приложение 1. Аннотация дисциплины

Приложение 2. Лист изменений

1. Пояснительная записка

1.1. Цели и задачи курса

Предмет курса – знакомство с современными представлениями об идеях и методах автоматического перевода текстов с одного языка на другой

Цель курса – освоение студентами базовых понятий и методов автоматического перевода, в первую очередь лингвистических понятий и методов, знакомство с историей автоматического перевода, принципов многоуровневого моделирования естественного языка, основных подходов к морфологическому, синтаксическому и семантическому анализу текстов, к соотношению грамматики и словаря

Задачи курса:

Курс нацелен на **формирование** у студентов следующих профессиональных компетенций:

- владение основными понятиями и категориями современной лингвистики (ОПК-1);
- способность использовать лингвистические технологии для проектирования систем автоматической обработки текста на естественном языке, лингвистических компонентов интеллектуальных и информационных электронных систем (ПК-11).

1.2. Формируемые компетенции и ожидаемые результаты обучения

Курс нацелен на **формирование** у студентов следующих профессиональных компетенций:

способностью пользоваться лингвистически ориентированными программными продуктами (ПК-9);

владением принципами создания электронных языковых ресурсов (текстовых, речевых и мультимодальных корпусов; словарей, тезаурусов, онтологии; фонетических, лексических, грамматических и иных баз данных и баз знаний) и умением пользоваться такими ресурсами (ПК-10);

способностью использовать лингвистические технологии для проектирования систем автоматической обработки звучащей речи и письменного текста на естественном языке, лингвистических компонентов интеллектуальных и информационных электронных систем (ПК-11);

способностью проводить квалифицированное тестирование лингвистически ориентированных программных продуктов, электронных ресурсов, лингвистически ориентированных систем и лингвистических компонентов интеллектуальных и информационных электронных систем (ПК-12);

По завершении обучения студент, полностью освоивший программу, должен:

знать:

- основные понятия современной лингвистики, лежащей в основе автоматического анализа и синтеза текстов;
- базовые принципы и методы автоматической обработки естественного текста для целей автоматического перевода и смежных задач (таких, как извлечение знаний из текстов и глубоко семантического анализа);
- основные принципы создания словаря для автоматического перевода.

уметь:

- определять морфологическую структуру предложения, определять основные элементы синтаксической структуры предложения в соответствии с задачей автоматического перевода;
- осуществлять самостоятельный анализ предложения в заданной лингвистической парадигме;
- обнаруживать в реальном предложении ситуации разного типа неоднозначностей, существенных для анализа текста и автоматического перевода;
- строить базовые правила автоматической обработки текста в заданной лингвистической парадигме);
- составлять простые статьи словаря для автоматического перевода (на примере конкретных естественных языков, в первую очередь, русского и английского).

владеть:

- современной терминологией автоматической обработки текста;
- навыками анализа простых предложений для задачи автоматического анализа и синтеза текстов;
- навыками создания простых правил автоматического перевода в рамках выбранной лингвистической парадигмы.

1.3. Место дисциплины в структуре образовательной программы

Курс «Автоматический перевод» относится к вариативной части блока дисциплин учебного плана по направлению подготовки 45.03.03 «Фундаментальная и прикладная лингвистика». Дисциплина адресована бакалаврам, обучающимся по направлению «Фундаментальная и прикладная лингвистика».

Для успешного освоения материала студент должен опираться на знания, умения и навыки, полученные в рамках курсов «Введение в теорию языка», «Общая морфология», «Общий синтаксис», «Общая семантика и лексикология», а также на навыки изучения научной литературы, сформированные при подготовке к другим курсам.

Курс направлен на углубление знаний, усовершенствование умений и навыков студентов в сфере общей и прикладной лингвистики, созданию у студентов представления об основных методах и подходах к автоматическому переводу и в целом к задаче автоматической обработки текстов.

2. СТРУКТУРА ДИСЦИПЛИНЫ

Общая трудоемкость освоения дисциплины составляет 3 зачетных единицы – 90 часов. Программой дисциплины предусмотрены аудиторные занятия (лекции 10 часов, семинары – 32 часа) и самостоятельная работа студента (48 часов), которая предусматривает подготовку к семинарам, выполнение домашних заданий, реализацию индивидуального мини-исследования и подготовку к промежуточной контрольной и экзамену.

Программой предусмотрены следующие **виды контроля**:

- текущий контроль успеваемости в форме проверки качества подготовки студента к семинарам (выполнение домашних заданий, чтение научной литературы, активное участие в обсуждении темы);
- краткие тесты в начале семинаров, требующие ответов на вопросы по понятиям, введенным на лекциях и обсужденным на предыдущих занятиях;
- промежуточная письменная контрольная, предполагающая выполнение практических заданий с опорой на теоретические положения курса;
- презентация индивидуальных исследовательских проектов по выбранным студентами темам (составляющая **экзамена**);
- устные ответы на контрольные вопросы по курсу (составляющая **экзамена**).

Тематический календарный план курса

№№ разд ела	Раздел курса	Семестр: недели	Виды учебной работы и трудоемкость (в часах)			Формы контроля успеваемости
			лекции	семи- нары	самостоя тельная работа студенто в	
1.	Введение. История автоматического перевода. Уроки Джорджаунского эксперимента	1: 1	2			Контроль посещаемости студентов
2.	Лингвистическое моделирование. Действующие модели языка. Теория «Смысл – Текст» и возможности ее использования для построения систем автоматической обработки текста.	1: 2-3	2	2	4	Контроль посещаемости студентов. Краткие тесты по пройденным темам. Обсуждение прочитанной научной литературы.
3.	Грамматика и словарь естественного языка. Представление об интегральном описании языка. Грамматики составляющих и грамматики зависимостей. Гибридные грамматики.	1: 4	1	1	6	Контроль посещаемости студентов. Краткие тесты по пройденным темам. Обсуждение прочитанной научной литературы. Коллективный анализ нетривиальных текстовых фрагментов, требующих выработки неординарных

№№ разд ела	Раздел курса	Семестр: недели	Виды учебной работы и трудоемкость (в часах)			Формы контроля успеваемости
			лекции	семи- нары	самостоя тельная работа студенто в	
						решений
4.	Автоматический анализ и синтез текста. Морфологический анализ текста. Синтаксический анализ текста. Правиловые и статистические подходы к автоматической обработке текста.	1: 5-6	1	3	4	Контроль посещаемости студентов. Краткие тесты по пройденным темам. Обсуждение прочитанной научной литературы. Коллективный анализ нетривиальных текстовых фрагментов, требующих выработки неординарных решений
	Промежуточная контрольная работа и ее обсуждение	1: 7		2		Письменная контрольная работа по разделам 2-4
5.	Языковая неоднозначность и методы ее разрешения при автоматической обработке текста. Применение онтологии и методов машинного обучения и интерактивных методов для разрешения неоднозначности.	1: 8-9		4	6	Контроль посещаемости студентов. Краткие тесты по пройденным темам. Обсуждение прочитанной научной литературы. Коллективный анализ нетривиальных текстовых фрагментов, требующих выработки неординарных решений
6.	Задача машинного перевода в кругу задач автоматической обработки текста на естественном языке. Система автоматического перевода как механизм обратной связи и источник новых лингвистических знаний.	1: 10-		2	6	Контроль посещаемости студентов. Краткие тесты по пройденным темам. Обсуждение прочитанной научной литературы.
7.	Типология систем машинного перевода. Правиловый и статистический подходы. Интерлингва.	1:11		2	2	Контроль посещаемости студентов. Обсуждение прочитанной научной литературы.
8.	Словарь системы автоматического перевода. Структура словарной статьи. Синтаксические и семантические признаки. Теория валентностей. Модель управления.	1:12-15	1	7	4	Контроль посещаемости студентов. Краткие тесты по пройденным темам. Коллективный анализ нетривиальных текстовых фрагментов, требующих выработки неординарных решений
9.	Лексические функции в	1:16-17	1	3	4	Контроль посещаемости

№№ разд ела	Раздел курса	Семестр: недели	Виды учебной работы и трудоемкость (в часах)			Формы контроля успеваемости
			лекции	семи- нары	самостоя тельная работа студенто в	
	машинном переводе.					студентов. Обсуждение прочитанной научной литературы. Коллективный анализ нетривиальных словосочетаний, требующих представления в виде лексических функций
10.	Смежные задачи прикладной лингвистики. Информационный поиск. Интеллектуальный анализ данных. Вопросно-ответные системы. Лингвистические онтологии.	1:18		2	4	Контроль посещаемости студентов. Обсуждение прочитанной научной литературы.
11.	Цифровые лингвистические ресурсы и их роль в автоматической обработке текстов. Аннотированные корпуса текстов. Word Net, Frame Net, Treebanks.	1:19-20	2	2	8	Контроль посещаемости студентов. Обсуждение прочитанной научной литературы. Коллективный анализ нетривиальных единиц, представленных в цифровых ресурсах
	Обсуждение индивидуальных решений задач	1: 21		2		Представление индивидуальных решений задач Участие в коллективном обсуждении
	Экзамен				18	Ответы на контрольные вопросы по курсу. Решение практических заданий
	Итого:		10	32	66	

3. СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

Раздел I. История автоматического перевода. (лекция 2 часа)

Джорджтаунский эксперимент 1954 г. Распространение результатов в мировую практику. Автоматический перевод в мире и в СССР. Взлеты и падения. Ограниченность возможностей языкового описания. Кристаллизация подходов к правилочному машинному переводу.

Раздел II. Лингвистическое моделирование (лекция 2 часа, семинар 2 часа, СРС 4 часа)

Действующие модели языка.

Теория «Смысл – Текст» и возможности ее использования для построения систем автоматической обработки текста.

Многоуровневый подход к анализу и синтезу. Поверхностный и глубинный уровни представления языка и текста.

Раздел III. Грамматика и словарь естественного языка (лекция 1 часа, семинар 1 часа, СРС 6 часов)

Представление об интегральном описании языка. Гармонизация словаря и грамматики.

Синтаксис составляющих и синтаксис зависимостей. Гибридные синтаксические представления.

Теоретическая лингвистика и задача автоматической обработки текстов.

Раздел IV. Автоматический анализ и синтез текста (лекция 1 час, семинары 3 часа, СРС 4 часа).

Морфологический анализ текста. Морфологические категории и граммы. Алгоритмы морфологического анализа.

Синтаксический анализ текста. Понятие синтагмы. Алгоритмы синтаксического анализа. Правилочные и статистические подходы к автоматической обработке текста. Ограниченная статистическая поддержка правилочных систем.

Трансфер. Типология правил перевода. Общие и трафаретные правила.

Раздел V. Языковая неоднозначность и методы ее разрешения при автоматической обработке текста (семинары 4 часа, СРС 6 часов).

Языковая неоднозначность как неотъемлемое свойство языка и способы ее разрешения. Типы неоднозначности. Морфологическая, лексическая и синтаксическая неоднозначность.

Методы разрешения неоднозначности при автоматической обработке текста.

Применение онтологии и методов машинного обучения и интерактивных методов для разрешения неоднозначности.

Раздел VI. Задача машинного перевода в кругу задач автоматической обработки текста на естественном языке (семинары 2 часа, СРС 6 часов).

Система автоматического перевода как механизм обратной связи и источник новых лингвистических знаний.

Уроки конкретных систем машинного перевода для теоретической лингвистики и лексикографии.

Раздел VII. Типология систем автоматического перевода (семинары 2 часа, СРС 2 часа).

Типы систем машинного перевода.

Правильный и статистический подходы к машинному переводу.

Поколения машинного перевода.

Интерлингва. Виды интерлингвы (естественные и искусственные семантические языки). UNL как разновидность интерлингвы.

Раздел VIII. Словарь систем автоматического перевода (лекция 1 час, семинары 7 часов, СРС 4 часа).

Словари системы автоматического перевода. Морфологические и комбинаторные словари.

Структура словарной статьи. Одноязычные и переводные зоны.

Синтаксические признаки.

Семантические признаки, дескрипторы и онтологические концепты.

Теория валентностей. Семантические и синтаксические валентности.

Модель управления.

Раздел IX. Лексические функции в машинном переводе (лекция 1 час, семинары 3 часа, СРС 4 часа).

Типология лексических функций. Коллокаты и субституты. Составные лексические функции. Суперпозиция лексических функций.

Применение лексических функций для идиоматического перевода

Лексические функции и перифразирование.

Использование лексических функций для снятия лексической и синтаксической неоднозначности текста.

Раздел X. Смежные задачи прикладной лингвистики (семинары 2 часа, СРС 4 часа).

Методы общие для автоматического перевода и других задач прикладной лингвистики.

Информационный поиск.

Интеллектуальный анализ данных.

Вопросно-ответные системы.

Лингвистические онтологии. Семантический анализ с учетом внеязыковой действительности.

Раздел XI. Цифровые лингвистические ресурсы и их роль в автоматической обработке текстов (лекция 2 часа семинары 2 часа, СРС 2 часа).

Аннотированные корпуса текстов.

Word Net. Frame Net.

Датасеты.

Векторная семантика.

4. ИНФОРМАЦИОННЫЕ И ОБРАЗОВАТЕЛЬНЫЕ ТЕХНОЛОГИИ

Дисциплина «Автоматический перевод» преподается достаточно традиционно – в виде лекций и семинаров, между которыми, однако, нет резкой границы: во время лекций допускается обсуждение конкретных тем, во время семинаров, при необходимости, преподаватель обращается к тем или иным теоретическим положениям. Значительное время отводится самостоятельной работе студентов (это чтение литературы и выполнение домашних заданий, а также подготовка к участию в семинарах)

Наименование раздела	Виды учебной работы	Формируемые компетенции	Информационные и образовательные технологии
Раздел I. Введение. История автоматического перевода. Уроки Джорджтаунского эксперимента	Лекция 1	ОПК-1, ПК-11	
Раздел II. Лингвистическое моделирование.	Лекция 2. Постановка проблемы	ОПК-1, ПК-11	
	Семинар 1. Действующие модели языка: общий подход		Коллективный анализ языковых явлений с привлечением демонстрации действующего лингвистического процессора
Раздел III. Грамматика и словарь естественного языка.	Лекция 3, часть 1.	ОПК-1, ПК-11	Анализ языковых явлений с привлечением демонстрации действующего лингвистического процессора
	Семинар 3, часть 1. Введение в интегральное описание языка		
Раздел IV. Анализ и синтез текста	Лекция 3, часть 2.	ОПК-1, ПК-11	
	Семинар 3, часть 2. Морфологический анализ		Анализ языковых явлений с привлечением демонстрации действующего лингвистического процессора

Наименование раздела	Виды учебной работы	Формируемые компетенции	Информационные и образовательные технологии
	Семинар 4. Синтаксический анализ. Статистические подходы		Анализ языковых явлений с привлечением демонстрации действующего лингвистического процессора
Раздел V. Языковая неоднозначность	Семинар 5 Эмпирические Методы разрешения неоднозначности при АОТ	ОПК-1, ПК-11, ПК-5, ПК-6	Анализ языковых явлений с привлечением демонстрации действующего лингвистического процессора
	Семинар 6. Онтологические, машинные и интерактивные методы		
Раздел VI. Автоматический перевод как источник новых лингвистических знаний.	Семинар 7. Уроки машинного перевода	ОПК-1, ПК-11	Анализ языковых явлений с привлечением демонстрации действующего лингвистического процессора
Раздел VII. Типология систем машинного перевода	Семинар 8. Интерлингва и ее разновидности, включая UNL	ОПК-1, ПК-11	Анализ языковых явлений; самостоятельное создание элементов UNL
Раздел VIII. Словарь системы машинного перевода	Лекция 4, часть 1. Введение в компьютерную лексикографию.	ОПК-1, ПК-11	Анализ языковых явлений с привлечением демонстрации действующего лингвистического процессора
	Семинар 9. Структура словарной статьи.		
	Семинар 10. Синтаксические и семантические признаки.		
	Семинар 11. Теория валентностей.		
	Семинар 12, часть 1. Модель управления		
Раздел IX. Лексические функции в машинном переводе	Лекция 4, часть 2.	ОПК-1, ПК-11	Анализ языковых явлений с привлечением демонстрации действующего лингвистического процессора
	Семинар 12, часть 2. Типология лексических функций		

Наименование раздела	Виды учебной работы	Формируемые компетенции	Информационные и образовательные технологии
	Семинар 13. Применение лексических функций для задач машинного перевода и АОТ	ОПК-1, ПК-11	Анализ языковых явлений с привлечением демонстрации действующего лингвистического процессора
Раздел X. Смежные задачи прикладной лингвистики.	Семинар 14. Информационный поиск. Интеллектуальный анализ данных. Вопросно-ответные системы. Лингвистические онтологии.	ОПК-1, ПК-11	Демонстрация и обсуждение излагаемых задач
Раздел XI. Цифровые лингвистические ресурсы	Лекция 5. Общее введение в цифровые ресурсы	ОПК-1, ПК-11	
	Семинар 15. Аннотированные корпуса текстов. Word Net, Frame Net	ОПК-1, ПК-11	Демонстрация и обсуждение ресурсов
Обсуждение индивидуальных решений задач	Семинар 16. Представление и решение индивидуальных задач	ОПК-1, ПК-11	Представление и решение индивидуальных задач. Консультации по курсу
Аттестация	Экзамен		Ответы на контрольные вопросы по курсу. Решение практических заданий

5. Фонд оценочных средств

5.1. Перечень компетенций и этапы их формирования

Курс «Автоматический перевод» направлен на формирование следующих компетенций: ОПК-1, ПК-11.

- **Компетенция ОПК-1** (владение основными понятиями и категориями современной лингвистики) развивается в рамках данного курса на всех занятиях.
- **Компетенция ПК-11** (способность использовать лингвистические технологии для проектирования систем автоматической обработки звучащей речи и письменного текста на естественном языке, лингвистических компонентов интеллектуальных и информационных электронных систем развивается в рамках данного курса на всех занятиях.

5.2. Методика оценивания степени сформированности компетенций

Общая трудоемкость освоения дисциплины составляет 4 зачетных единицы – 144 часа. Программой дисциплины предусмотрены аудиторские занятия (лекции 18 часов, семинары – 38 часов) и самостоятельная работа студента (88 часов), которая предусматривает подготовку к семинарам, выполнение домашних заданий, реализацию индивидуального мини-исследования и подготовку к промежуточной контрольной и экзамену.

Оценка знаний студента производится по 100-балльной шкале и учитывает результаты текущего контроля успеваемости (до 60 баллов) и результаты промежуточной аттестации (до 40 баллов).

Оценка «удовлетворительно» выставляется, если студент набрал в сумме не менее 50 баллов. При выставлении оценки в ведомость и в зачетную книжку преподаватель должен указать результат в соответствии с традиционной шкалой оценок и со шкалой оценок Европейской системы переноса и накопления кредитов (European Credit Transfer System; далее – ECTS) в соответствии с таблицей:

100-балльная шкала	Традиционная шкала		Шкала ECTS
95 – 100	отлично	зачтено	A
83 – 94			B
68 – 82	хорошо		C
56 – 67	удовлетворительно		D
50 – 55			E
20 – 49	неудовлетворительно	не зачтено	FX
0 – 19			F

Распределение баллов по видам учебной деятельности таково:

- 1) уровень активности студента на семинарах (выполнение домашних заданий и участие в их обсуждении; выступления по прочитанной литературе; участие в обсуждении и выполнении коллективных заданий; представление промежуточных результатов индивидуальных исследований и участие в их обсуждении) — всего до 20 баллов;
- 2) написание мини-тестов по темам предшествующих занятий — до 15 баллов;
- 3) написание промежуточной письменной контрольной — до 25 баллов;
- 4) защита индивидуального исследовательского проекта — до 30 баллов;
- 5) ответы на контрольные вопросы по курсу — до 10 баллов

5.3. Описание и обоснование шкал и критериев оценки качества работы студента

При выставлении оценки преподаватель ориентируется на следующие содержательные критерии.

Количество баллов	Критерии оценки
95–100 (A)	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.
83–94 (B)	Теоретическое содержание дисциплины освоено полностью, без пробелов, необходимые практические навыки работы с освоенным материалом сформированы, все предусмотренные рабочей программой дисциплины учебные задания выполнены, качество выполнения большинства из них оценено числом баллов, близким к максимальному.
68–82 (C)	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.
56–67 (D)	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 содержат ошибки.
50–55 (E)	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.
20–54 (FX)	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.
0–19 (F)	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.

Баллы за участие в семинарах. Преподавание дисциплины строится на параллельном обсуждении теоретических вопросов и выполнении индивидуальных и коллективных практических заданий. Все эти задания в обязательном порядке обсуждаются на семинарах. Активное и разумное участие в обсуждении этих заданий, а также доклады о прочитанной литературе и обсуждение индивидуальных исследовательских проектов может принести студенту до 20 баллов.

Баллы за выполнение текущих контрольных работ. В начале большинства семинаров проводятся краткие тесты, включающие в себя базовые вопросы по темам предшествующих занятий. Суммарно за семестр выполнение этих тестов может принести студенту до 15 баллов.

Баллы за промежуточную аттестацию. Экзамен состоит из двух частей. В первой части студенты представляют ответы на практические задания. Оценивается степень овладения тематикой, корректность выбранного метода исследования, наличие содержательных результатов и умение их представить аудитории (в сумме — до 30 баллов). Во второй части студенты устно отвечают на контрольные вопросы по курсу (до 10 баллов).

5.4. Типовые задания, используемые для оценки степени освоенности учебного материала

5.4.1. Контрольные вопросы по теоретическому материалу

Контрольные вопросы преимущественно затрагивают теоретический материал; проверка умений и навыков осуществляется в первую очередь с помощью практических заданий. Контрольные вопросы также используются на экзамене.

Алгоритм автоматического перевода

1. Уровни представления текста. Поверхностные и глубинные уровни.
2. Каковы основные этапы автоматического перевода?
3. Этап трансфера в автоматическом переводе.
4. Что такое интерлингва?
5. Основные принципы статистического машинного перевода.

Морфология

6. Что такое морфологический анализ?
7. Что такое морфологический синтез?
8. Морфологические позиции слова и их роль в морфологическом анализе.
9. Основные средства и этапы морфологического анализа текста.
10. Словоизменительные грамматические категории
11. Классифицирующие грамматические категории: для чего они нужны?
12. Каковы основные грамматические категории английских глаголов.
13. Каковы основные категории русских прилагательных и их значения?

Синтаксис

14. Дерево составляющих и дерево зависимостей – сходства и различия.

15. Что такое синтаксическое отношение?
16. Повторимые и неповторимые синтаксические отношения
17. Что такое синтаксический анализ?
18. Что такое нормализованная синтаксическая структура?
19. Что такое расширенная синтаксическая структура?
20. Что такое проективность синтаксической структуры?
21. Согласование в узком и широком смысле

Словарь

22. Для чего служат синтаксические признаки слов?
23. Семантические признаки слов и концепты,
24. Что такое модель управления слова?
25. Актанты и сирконстанты предикатных слов.
26. Типичная модель управления русского глагола.
27. Типичная модель управления русского предлога.
28. Что такое сильноуправляемые предлоги?
29. Что такое сильноуправляемые союзы?
30. Каковы различия между синтаксическими и семантическими валентностями слова?
31. Активные, пассивные и дистантные валентности слов.

Ресурсы

32. Что такое лексическая функция?
33. Как могут использоваться в автоматическом переводе лексические функции?
34. Что такое размеченный корпус текстов?
35. Как устроен WordNet и для чего он может использоваться?

5.4.2. Образцы вопросов для кратких тестов

- 1) Построить МорфС предложения *Он не спал и трех часов вчера.*
Построить СинтС предложения *В Питере снегу нападало, все деревья в снегу.*
- 2) Построить МорфС предложения *Он взял за правило просто говорить о сложных вещах.*
Построить СинтС этого предложения.

5.4.3. Образцы домашних заданий

- 1) Домашнее задание № 3 по автоматическому переводу. Бакалавры, 4 курс бакалавриата, 2020/21 уч.г. Валентная структура слова. Вариант 1.

Ответьте на следующие вопросы:

- а) Сколько валентностей глагола *сменять* насыщено в предложении
Одна заря сменить другую спешит, дав ночи полчаса? (А.С.Пушкин):
(A) 5 (B) 4 (C) 3 (D) 2 (E) 1
Какие это валентности?

- б) Сколько валентностей глагола *сменять* насыщено в предложении
Предложения сменить кириллицу на латиницу раздавались и в нашей стране.

(A) 5 (B) 4 (C) 3 (D) 2 (E) 1

Какие это валентности?

в). Сколько валентностей глагола *сменять* насыщено в предложении

Рауль Кастро, сменивший на посту президента Кубы брата Фиделя, разрешил простым кубинцам пользоваться мобильными телефонами?

(A) 5 (B) 4 (C) 3 (D) 2 (E) 1

Какие это валентности?

г). Сколько валентностей глагола *сменять* насыщено в предложении

А уж совсем потом — то есть в третьем классе — нам сменили школьную форму (Д.Драгунский)?

(A) 5 (B) 4 (C) 3 (D) 2 (E) 1

Какие это валентности?

д) Сколько валентностей глагола *сменять* насыщено в предложении

Мы мечтали выиграть звание чемпиона еще в 2001-м, когда высшая лига проходила в Томске (даже сменили название с «Сибири» на «Томск-400») (НКРЯ)

(A) 5 (B) 4 (C) 3 (D) 2 (E) 1

Какие это валентности?

Все ли валентности глагола *сменять* насыщены в предложениях 1-5?

Все ответы прокомментируйте!

2) Домашнее задание № 4 по автоматическому переводу. 4 курс бакалавриата, 2020/21 уч.г. Актантные конструкции. Вариант 2.

Найдите актантные связи (т.е. связи, оформляемые предикативным, агентивным, квазиагентивным, 1-м (2-м, 3-м...) комплетивным и предложным или дательно-субъектным СинтО) в следующем тексте (<https://www.newsru.com/cinema/14oct2020/embrace.html>):

Российский фотограф победил в престижном мировом конкурсе со снимком тигриных "Объятий", на создание которого ушел почти год. «Свет, цвета и текстура здесь как на картине, написанной маслом. Тигрица как бы превратилась в часть леса. Ее хвост сливается с корнями дерева. Они стали одним целым», - описала фото председатель жюри конкурса Wildlife Photographer of the Year Роз Кидманд-Кокс в разговоре с BBC. По словам исполнительного директора Музея естествознания по научной работе Тома Литлвуда, кадр, на котором представить исчезающего вида переживает подобный момент спокойствия, "вселяет надежду".

3) Домашнее задание № 8 по автоматическому переводу. 4 курс бакалавриата, 2020/21 уч.г. Состав компьютерного словаря. Вариант 3.

Постарайтесь определить, какие словарные статьи следовало бы ввести в компьютерный словарь русского языка для многозначных слов, леммы которых таковы:

а) *просыпать*

б) *просыпаться*

в) *единица*

г) *дежурный*

Ответы прокомментируйте!

6. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

6.1. Список источников и литературы

Основная литература

Мельчук И.А. Опыт теории лингвистических моделей класса «Смысл – Текст». – М.: Наука, 1974/М.: Школа « Языки русской культуры», 1999.

Апресян Ю.Д., Богуславский И.М., Иомдин Л.Л. и др. Лингвистический процессор для сложных информационных систем. М.: Наука, 1992.

Апресян Ю.Д. Избранные труды. Том 2: Интегральное описание языка и системная лексикография. М.: Школа « Языки русской культуры», 1995.

Apresjan Ju, Boguslavsky I., Iomdin L et al. ETAP-3 Linguistic Processor: a Full-Fledged NLP Implementation of the MTT // MTT 2003, First International Conference on Meaning – Text Theory (June 16-18 2003). Paris: École Normale Supérieure, 2003. P. 279-288.

Дополнительная литература

Баранов А.Н. Введение в прикладную лингвистику. Серия "Новый лингвистический учебник". М.: Эдиториал УРПС. 2001. Глава 2, раздел 1.3.1. Моделирование общения (с. 20-31); Глава 4, разделы 1.3.1. – 1.3.4. «Естественный» перевод: лингвистические проблемы (с. 143-163); 1.4. Машинный перевод (с. 168-178).

Большакова Е.И., Клышинский Э.С., Ландэ Д.В., Носков А.А., Пескова О.В., Ягунова Е.В. Автоматическая обработка текстов на естественном языке и компьютерная лингвистика : учеб. пособие /. — М.: МИЭМ, 2011. — 272 с.

Кулагина О. С. Исследования по машинному переводу. – М., 1979.

W. John Hutchins and Harold L. Somers. An introduction to machine translation. London: Academic Press, 1992.

Carl Pollard, Ivan A. Sag: Head-Driven Phrase Structure Grammar. Chicago: University of Chicago Press. 1994.

Boguslavsky I, Iomdin L. Nivre J. Parsing the Russian Dependency Treebank. Proceedings of COLING-2008. Manchester, 2008.

Philipp Koehn. Statistical Machine Translation. Cambridge University Press. 2009.

Jury D. Apresjan, Igor M. Boguslavsky, Leonid L. Iomdin, Leonid L. Tsinman. Lexical Functions in Actual NLP-Applications // Selected Lexical and Grammatical Issues in the Meaning–Text Theory. In honour of Igor Mel'čuk. (Ed. by Leo Wanner). John Benjamins, Studies in Language Companion. Series 84. ISBN 978 90 272 3094 2. 2007. P. 199-230.

New Trends of Research in Ontologies and Lexical Resources: Ideas, Projects, Systems. Eds: Allessandro Oltramari, Piek Vossen, Lu Qin, Eduard Hovy. Springer, Theory and And Applications in Natural Language Processing Series. 2013.

7. Материально-техническое обеспечение дисциплины (модуля)

*В качестве материально-технического обеспечения дисциплины указывается необходимое для обучения **лицензионное программное обеспечение (из рекомендованного списка)**, оборудование, демонстрационные приборы, мультимедийные средства, учебные фильмы, тренажеры, требования к аудиториям – компьютерные классы, академические или специально оборудованные аудитории и лаборатории, наличие доски и т.д.*

8. Обеспечение образовательного процесса для лиц с ограниченными возможностями здоровья и инвалидов

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.

Приложение 1. Аннотация дисциплины

Курс «Автоматический перевод» предназначен для студентов бакалавриата по направлению «Фундаментальная и прикладная лингвистика» Института лингвистики РГГУ. Он преподается силами УНЦ компьютерной лингвистики.

Задачи курса:

Курс нацелен на **формирование** у студентов следующих профессиональных **компетенций**:

способностью пользоваться лингвистически ориентированными программными продуктами (ПК-9);

владением принципами создания электронных языковых ресурсов (текстовых, речевых и мультимодальных корпусов; словарей, тезаурусов, онтологий; фонетических, лексических, грамматических и иных баз данных и баз знаний) и умением пользоваться такими ресурсами (ПК-10);

способностью использовать лингвистические технологии для проектирования систем автоматической обработки звучащей речи и письменного текста на естественном языке, лингвистических компонентов интеллектуальных и информационных электронных систем (ПК-11);

способностью проводить квалифицированное тестирование лингвистически ориентированных программных продуктов, электронных ресурсов, лингвистически ориентированных систем и лингвистических компонентов интеллектуальных и информационных электронных систем (ПК-12).

По завершении обучения студент, полностью освоивший программу, должен:

знать:

- основные понятия современной лингвистики, лежащей в основе автоматического анализа и синтеза текстов;
- базовые принципы и методы автоматической обработки естественного текста для целей автоматического перевода и смежных задач (таких, как извлечение знаний из текстов и глубоко семантического анализа);
- основные принципы создания словаря для автоматического перевода.

уметь:

- определять морфологическую структуру предложения, определять основные элементы синтаксической структуры предложения в соответствии с задачей автоматического перевода;
- осуществлять самостоятельный анализ предложения в заданной лингвистической парадигме;
- обнаруживать в реальном предложении ситуации разного типа неоднозначностей, существенных для анализа текста и автоматического перевода;
- строить базовые правила автоматической обработки текста в заданной лингвистической парадигме);
- составлять простые статьи словаря для автоматического перевода (на примере конкретных естественных языков, в первую очередь, русского и английского).

владеть:

- современной терминологией автоматической обработки текста;
- навыками анализа простых предложений для задачи автоматического анализа и синтеза текстов;
- навыками создания простых правил автоматического перевода в рамках выбранной лингвистической парадигмы.

Курс «Автоматический перевод» относится к вариативной части блока дисциплин учебного плана по направлению подготовки 45.03.03 «Фундаментальная и прикладная лингвистика». Дисциплина адресована бакалаврам, обучающимся по направлению «Фундаментальная и прикладная лингвистика» (Б1.Б.23) без направленности (профиля). Курс читается на 4-м курсе в 1-м семестре УНЦ компьютерной лингвистики Института лингвистики РГГУ, форма промежуточного контроля – экзамен.

Для успешного освоения материала студент должен опираться на знания, умения и навыки, полученные в рамках курсов «Введение в теорию языка», «Общая морфология», «Общий синтаксис», «Общая семантика и лексикология», а также на навыки изучения научной литературы, сформированные при подготовке к другим курсам

№	Текст актуализации или прилагаемый к РПД документ, содержащий изменения	Дата	№ протокола
1	Приложение №1	29.06.2016 г.	3
2	Приложение №2	20.06.2017 г.	5
3	Приложение №3	20.06.2018 г.	1
4	Приложение №4	25.06.2020	4

Состав программного обеспечения (ПО), современных профессиональных баз данных (БД) и информационно-справочные систем (ИСС) (2016 г.)

1. Перечень ПО*

Таблица 1

№п/п	Наименование ПО	Производитель	Способ распространения (лицензионное или свободно распространяемое)
1	Adobe Master Collection CS4	Adobe	лицензионное
2	Microsoft Office 2010	Microsoft	лицензионное
3	Windows 7 Pro	Microsoft	лицензионное
4	AutoCAD 2010 Student	Autodesk	свободно распространяемое
5	Archicad 18 Rus Student	Graphisoft	свободно распространяемое
6	SPSS Statistics 22	IBM	лицензионное
7	Kaspersky Endpoint Security	Kaspersky	лицензионное

** Оставить используемое ПО в рамках учебной дисциплины*

2. Перечень БД и ИСС

Таблица 2

№п/п	Наименование
1	Международные реферативные наукометрические БД, доступные в рамках национальной подписки в 2016 г. Scopus
2	Профессиональные полнотекстовые БД, доступные в рамках национальной подписки в 2016 г. Журналы Oxford University Press SAGE Journals Журналы Taylor and Francis
3	Профессиональные полнотекстовые БД JSTOR Издания по общественным и гуманитарным наукам
4	Компьютерные справочные правовые системы Консультант Плюс, Гарант

Состав программного обеспечения (ПО), современных профессиональных баз данных (БД) и информационно-справочные систем (ИСС) (2017 г.)

1. Перечень ПО

Таблица 1

№п/п	Наименование ПО	Производитель	Способ распространения (лицензионное или свободно распространяемое)
1	Adobe Master Collection CS4	Adobe	лицензионное
2	Microsoft Office 2010	Microsoft	лицензионное
3	Windows 7 Pro	Microsoft	лицензионное
4	AutoCAD 2010 Student	Autodesk	свободно распространяемое
5	Archicad 19 Rus Student	Graphisoft	свободно распространяемое
6	SPSS Statistics 22	IBM	лицензионное
7	Microsoft Share Point 2010	Microsoft	лицензионное
8	SPSS Statistics 25	IBM	лицензионное
9	Microsoft Office 2013	Microsoft	лицензионное
10	ОС «Альт Образование» 8	ООО «Базальт СПО	лицензионное
11	Kaspersky Endpoint Security	Kaspersky	лицензионное

** Оставить используемое ПО в рамках учебной дисциплины*

2. Перечень БД и ИСС

Таблица 2

№п/п	Наименование
	Международные реферативные наукометрические БД, доступные в рамках национальной подписки в 2017 г. Web of Science Scopus
	Профессиональные полнотекстовые БД, доступные в рамках национальной подписки в 2017 г. Журналы Oxford University Press ProQuest Dissertation & Theses Global SAGE Journals Журналы Taylor and Francis
	Профессиональные полнотекстовые БД JSTOR Издания по общественным и гуманитарным наукам
	Компьютерные справочные правовые системы Консультант Плюс,

	Гарант
--	--------

Состав программного обеспечения (ПО), современных профессиональных баз данных (БД) и информационно-справочные систем (ИСС) (2018 г.)

1. Перечень ПО

Таблица 1

№п/п	Наименование ПО	Производитель	Способ распространения (лицензионное или свободно распространяемое)
1	Adobe Master Collection CS4	Adobe	лицензионное
2	Microsoft Office 2010	Microsoft	лицензионное
3	Windows 7 Pro	Microsoft	лицензионное
4	AutoCAD 2010 Student	Autodesk	свободно распространяемое
5	Archicad 21 Rus Student	Graphisoft	свободно распространяемое
6	SPSS Statistics 22	IBM	лицензионное
7	Microsoft Share Point 2010	Microsoft	лицензионное
8	SPSS Statistics 25	IBM	лицензионное
9	Microsoft Office 2013	Microsoft	лицензионное
10	ОС «Альт Образование» 8	ООО «Базальт СПО	лицензионное
11	Microsoft Office 2013	Microsoft	лицензионное
12	Windows 10 Pro	Microsoft	лицензионное
13	Kaspersky Endpoint Security	Kaspersky	лицензионное

** Оставить используемое ПО в рамках учебной дисциплины*

2. Перечень БД и ИСС

Таблица 2

№п/п	Наименование
	Международные реферативные наукометрические БД, доступные в рамках национальной подписки в 2018 г. Web of Science Scopus
	Профессиональные полнотекстовые БД, доступные в рамках национальной подписки в 2018 г. Журналы Cambridge University Press ProQuest Dissertation & Theses Global SAGE Journals Журналы Taylor and Francis Электронные издания издательства Springer
	Профессиональные полнотекстовые БД JSTOR Издания по общественным и гуманитарным наукам

	Компьютерные справочные правовые системы Консультант Плюс, Гарант
--	---

1. 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.

2. Перечень БД и ИСС (к п. 6.2 на 2020 г.)

№п/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	Компьютерные справочные правовые системы Консультант Плюс, Гарант

3. Состав программного обеспечения (ПО) (к п. 7 на 2020 г.)

№п/п	Наименование ПО	Производитель	Способ распространения (лицензионное или свободно распространяемое)
1	Adobe Master Collection CS4	Adobe	лицензионное
2	Microsoft Office 2010	Microsoft	лицензионное

3	Windows 7 Pro	Microsoft	лицензионное
4	AutoCAD 2010 Student	Autodesk	свободно распространяемое
5	Archicad 21 Rus Student	Graphisoft	свободно распространяемое
6	SPSS Statistics 22	IBM	лицензионное
7	Microsoft Share Point 2010	Microsoft	лицензионное
8	SPSS Statistics 25	IBM	лицензионное
9	Microsoft Office 2013	Microsoft	лицензионное
10	ОС «Альт Образование» 8	ООО «Базальт СПО	лицензионное
11	Microsoft Office 2013	Microsoft	лицензионное
12	Windows 10 Pro	Microsoft	лицензионное
13	Kaspersky Endpoint Security	Kaspersky	лицензионное
14	Microsoft Office 2016	Microsoft	лицензионное
15	Visual Studio 2019	Microsoft	лицензионное
16	Adobe Creative Cloud	Adobe	лицензионное
17	Zoom	Zoom	лицензионное