

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

**«Лингвистические основы машинного перевода»**

**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 2021**

# Лингвистические основы машинного перевода

## Рабочая программа дисциплины

**Составитель:**

**к.ф.н Л.Л.Иомдин**

**Ответственный редактор:**

**д. филол. н., профессор В.И.Подлеская**

УТВЕРЖДЕНО

Протокол заседания УНЦ компьютерной  
лингвистики

**№ 6 от «9» апреля 2021г.**

## **ОГЛАВЛЕНИЕ**

### **1. Пояснительная записка**

1.1 Цель и задачи дисциплины

1.2. Перечень планируемых результатов обучения по дисциплине, соотнесенных с индикаторами достижения компетенций

1.3. 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 (материалы) для текущего контроля успеваемости, промежуточной аттестации обучающихся по дисциплине

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

6.1. Список источников и литературы

6.2. Перечень ресурсов информационно-телекоммуникационной сети «Интернет»

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

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

## **Приложения**

Приложение 1. Аннотация дисциплины

Приложение 2. Лист изменений

## 1. Пояснительная записка

### 1.1 Цели и задачи дисциплины

*Предметом дисциплины* является система описания языка с особенностями, присущими такому описанию, когда оно осуществляется в целях автоматической обработки текстов, в первую очередь машинного перевода, основанного на правилах.

Основной внимание уделяется 1) тому, как текст на входном языке анализируется на разных уровнях (морфологическом, синтаксическом, семантическом), как соотносятся грамматическое описание языка для целей автоматической обработки текста и словарь языка, 2) тому, как осуществляется преобразование глубинной структуры входного языка в структуру выходного языка и 3) тому, как синтезируется текст на выходном языке. Курс предполагает интенсивную, практическую работу магистрантов, которые должны освоить основные методы и подходы, применяемые лингвистами, разрабатывающими прикладные системы автоматической обработки текстов, и научиться самостоятельно создавать адекватные правила такой обработки.

Курс направлен на решение следующих задач:

- дать представление о месте теоретической лингвистики в задачах, решаемых компьютерной лингвистикой;
- познакомить магистрантов с важнейшими областями междисциплинарных исследований на стыке лингвистики со смежными дисциплинами, в первую очередь с компьютерной наукой;
- научить магистрантов обнаруживать типологические сходства и различия естественных языков, между которыми осуществляется машинный перевод, и творчески использовать полученные знания;
- научить магистрантов пользоваться методами обратной связи, т.е. применять полученные при разработке автоматических систем результаты для извлечения новых знаний о естественном языке;
- познакомить магистрантов с современными подходами к решению задачи машинного перевода, в том числе с гибридными и статистическими подходами и приемами машинного обучения.

### 1.2 Перечень планируемых результатов обучения по дисциплине, соотнесенных с индикаторами достижения компетенций

Компетенция	Индикаторы компетенций	Результаты обучения
УК-1 Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий	УК-1.2	Умеет определять круг задач и решать поставленные задачи в рамках избранных видов профессиональной деятельности, планировать собственную деятельность исходя из имеющихся ресурсов
УК-2 Способен управлять проектом на всех этапах его жизненного цикла	УК-2.2	Умеет соотносить разнородные явления и систематизировать их в рамках избранных видов профессиональной деятельности
УК-6 Способен определять и реализовывать приоритеты	УК-6.2	Умеет планировать свое рабочее время и время для саморазвития, формулировать

собственной деятельности и способы ее совершенствования на основе самооценки		цели личностного и профессионального развития и условия их достижения, исходя из тенденций развития области профессиональной деятельности, индивидуально-личностных особенностей
ПК-2 Владеет 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-2.3	Имеет практический опыт разработки электронных языковых ресурсов; опыт применения основных методов, способов и средств получения, хранения, переработки информации
ПК-3 Способен использовать лингвистические технологии для проектирования систем автоматической обработки звучащей речи и письменного текста на естественном языке, лингвистических компонентов интеллектуальных и информационных электронных систему	ПК-3.3	Умеет пользоваться существующими системами автоматической обработки текста и звучащей речи, интеллектуальными и информационными электронными системами; проводить их сравнительный анализ; проектировать модули данных систем, составлять технические задания

### 1.3 Место дисциплины в структуре образовательной программы

Дисциплина «Лингвистические основы машинного перевода» относится к части, формируемой участниками образовательных отношений.

В результате освоения дисциплины формируются знания, умения и владения, необходимые для изучения следующих дисциплин и прохождения практик: Лингвистическая и онтологическая семантика.

## 2. Структура дисциплины

Общая трудоёмкость дисциплины составляет 2 з.е., 76ч., в том числе контактная работа обучающихся с преподавателем 20ч., промежуточная аттестация 0ч., самостоятельная работа обучающихся 56ч.

№ 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	Практически	самостоятельн	контрол	
				лекции	Практически	самостоятельн	контрол	

				и	е занятия	ая работа	ь	
1	Лингвистика как наука о языке. Представление об уровнях представления языка – фонетика, морфология, синтаксис, семантика. Лингвистика и прагматика	1	1	1	1	3		
2	Лингвистическое моделирование. Действующие модели языка. Теория «Смысл – Текст» и возможности ее использования для построения систем автоматической обработки текста.	1	1-2	1	1	3		Реферирование научного сочинения из рекомендованного списка
3	Грамматика и словарь естественного языка. Представление об интегральном описании языка.	1	2	1	1	3		Реферирование научного сочинения из рекомендованного списка
4.	Краткий обзор формальных грамматик. Порождающие грамматики. Грамматики составляющих и грамматики зависимостей. Гибридные грамматики..	1	2-3	1	1	3		Реферирование научного сочинения из рекомендованного списка
5.	Автоматический анализ и синтез текста. Морфологический анализ текста.	1	3-4		1	3		
6.	Синтаксический анализ текста (парсинг). Различные подходы к синтаксическому анализу: анализ «сверху вниз» и «снизу вверх». Правильные и статистические подходы к автоматической обработке текста.	1	3-4		1	3		Выполнение практического задания: работа с конкретной системой машинного перевода ЭТАП-3. Наблюдение за алгоритмом синтаксического анализа.
7.	Языковая неоднозначность	1	4		1	3		Выполнение практического

	как принципиальное свойство языка и методы ее разрешения при автоматической обработке текста. Типы неоднозначности. Интерактивное разрешение лексической и синтаксической неоднозначности. Применение онтологии для разрешения неоднозначности. Методы машинного обучения в применении к разрешению однозначности.							задания: работа с конкретной системой машинного перевода ЭТАП-3. Наблюдение за решением задачи разрешения неоднозначностей разных типов.
8.	Задача машинного перевода в кругу задач автоматической обработки текста на естественном языке. Система машинного перевода как механизм обратной связи и источник новых лингвистических знаний.	1	4-5		1	3		Реферирование научного сочинения из рекомендованного списка
9.	Типы систем машинного перевода. Автоматический и автоматизированный перевод. Память переводов. Интерлингва. UNL (универсальный сетевой язык) как тип интерлингвы и его использование в задаче машинного перевода.	1	5		1	3		Выполнение практического задания: работа с системой UNL
10.	Морфологический компонент системы автоматической обработки текстов. Морфологическая структура слова и предложения.	1	5		1	3		Выполнение практического задания: построение морфологических структур предложений разных типов на примере различных

							естественных языков
11	Синтаксический компонент системы автоматической обработки текстов. Синтаксическая структура предложения. Алгоритм синтаксического анализа. Синтаксические отношения. Синтагмы	1	6		1	3	Выполнение практических заданий: 1)ручное построение синтаксических структур предложений разных типов на примере различных естественных языков, 2) самостоятельное построение синтаксических правил на специальном формальном языке.
12	Словарь системы автоматической обработки текстов. Словарь системы машинного перевода. Структура словарной статьи. Синтаксические признаки. Семантические признаки (дескрипторы). Теория валентностей. Модель управления.	1	6-7		1	3	Выполнение практического задания: создание словарных статей словарей разных языков для машинного перевода.
13.	Правила межъязыкового перевода в узком смысле (трансфер).	1	7		1	3	Выполнение практического задания: написание пробных правил трансфера для перевода с русского языка на английский и обратно
14.	Лексические функции в машинном переводе.	1	8		1	3	Выполнение практического задания: работа с базой данных лексических функций
15.	Краткий обзор действующих систем машинного перевода. Статистические системы машинного перевода.	1	8-9		1	3	Выполнение практического задания: работа с доступными в Интернете статистическими системами перевода и



								критический анализ систем лексических функций
16.	Краткий обзор смежных задач прикладной лингвистики. Информационный поиск. Интеллектуальный анализ данных. Синонимическое перифразирование высказываний и его прикладное значение.	1	9		1	3		
17.	Некоторые современные цифровые лингвистические ресурсы и их роль в задачах автоматической обработки текстов. Аннотированные корпуса текстов. Word Net, Frame Net, Treebanks (включая SynTagRus), Semantic Web).	1	10		1	3		Выполнение практического задания: выработка навыков пользования конкретными цифровыми лингвистическими ресурсами
18.	Зачет	1				2		Реферат; коллоквиум по контрольным вопросам с выполнением практического задания
	Итого:			4	16	56		

### 3. Содержание дисциплины

1. Лингвистика как наука о языке. Представление об уровнях представления языка – фонетика, морфология, синтаксис, семантика. Лингвистика и прагматика.
2. Лингвистическое моделирование. Действующие модели языка. Теория «Смысл – Текст» и возможности ее использования для построения систем автоматической обработки текста.
3. Грамматика и словарь естественного языка. Представление об интегральном описании языка.
4. Краткий обзор формальных грамматик. Порождающие грамматики. Грамматики составляющих и грамматики зависимостей. Гибридные грамматики.
5. Автоматический анализ и синтез текста. Морфологический анализ текста. Синтаксический анализ текста (парсинг). Различные подходы к синтаксическому анализу: анализ «сверху вниз» и «снизу вверх». Правильные и статистические подходы к автоматической обработке текста.

6. Языковая неоднозначность как принципиальное свойство языка и методы ее разрешения при автоматической обработке текста. Типы неоднозначности. Интерактивное разрешение лексической и синтаксической неоднозначности. Применение онтологии для разрешения неоднозначности. Методы машинного обучения в применении к разрешению неоднозначности.
7. Задача машинного перевода в кругу задач автоматической обработки текста на естественном языке. Система машинного перевода как механизм обратной связи и источник новых лингвистических знаний.
8. Типы систем машинного перевода. Автоматический и автоматизированный перевод. Память переводов. Интерлингва. UNL (универсальный сетевой язык) как тип интерлингвы и его использование в задаче машинного перевода.
9. Морфологический компонент системы автоматической обработки текстов. Морфологическая структура слова и предложения.
10. Синтаксический компонент системы автоматической обработки текстов. Синтаксическая структура предложения. Алгоритм синтаксического анализа. Синтаксические отношения. Синтагмы.
11. Словарь системы автоматической обработки текстов. Словарь системы машинного перевода. Структура словарной статьи. Синтаксические признаки. Семантические признаки (дескрипторы). Теория валентностей. Модель управления.
12. Правила межъязыкового перевода в узком смысле (трансфер).
13. Лексические функции в машинном переводе.
14. Краткий обзор действующих систем машинного перевода. Статистические системы машинного перевода.
15. Краткий обзор смежных задач прикладной лингвистики. Информационный поиск. Интеллектуальный анализ данных. Синонимическое перифразирование высказываний и его прикладное значение.
16. Некоторые современные цифровые лингвистические ресурсы и их роль в задачах автоматической обработки текстов.. Аннотированные корпуса текстов. Word Net, Frame Net, Treebanks (включая SynTagRus), Semantic Web).

#### **4. Образовательные технологии**

Интерактивные формы обучения в данном курсе предполагают:

1. систематическое использование компьютерных презентаций (как преподавателем в установочной части занятия, так и студентом, выступающим с критическим разбором реферируемого научного сочинения);
2. он-лайн демонстрации работы с лингвистическими базами данных и энциклопедическими интернет-ресурсами (информационно-справочный ресурс по языкам мира Etnolog <http://www.ethnologue.com> (частично-платный ресурс, социолингвистическая, информация, ареал распространения, карты); типологическая база данных WALS, The World Atlas of Language Structures <http://wals.info>; информационно-справочный ресурс по языкам мира <http://glottolog.org> (генеалогическая аффилиация и обширная библиография); многоязычная платформа для составления конкордансов и исследования грамматической и лексической дистрибуции WordSketchEngine <https://www.sketchengine.co.uk/> ; лексикологическая база данных WordNet (<http://wordnet.princeton.edu/>; электронная энциклопедия по лингвистике серии Oxford Research Encyclopedia <http://linguistics.oxfordre.com/> ; и др.;

3. использование открытых он-лайн аудио- и видео обучающих ресурсов (для академического английского: [www.cambridge.org/elt/english-for-academics](http://www.cambridge.org/elt/english-for-academics));
4. практическую работу с электронными анализаторами речи и платформами лингвистического документирования (платформы E-Language Archiving Technology, ELAN <http://www.lat-mpi.eu/tools/elan>; коллекция электронных ресурсов на сайте Summer Institute of Linguistics ([www.sil.org](http://www.sil.org)), в том числе, программа для акустического анализа речи Speech analyzer, программа IPA Help для обучения и пользования Международной фонетической транскрипцией, комплекс программ для полевой работы; программа для акустического анализа речи PRAAT (<http://www.fon.hum.uva.nl/praat>); открытое он-лайн издание по документированию и консервации языков Language Documentation & Conservation, LD&C <http://www.nflrc.hawaii.edu/ldc>).
5. 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.

## 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	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:

- 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, сдаёт зачёт по всему курсу и предъявляет преподавателю собственноручно написанные конспекты специальной литературы и выполненные домашние задания ко всем семинарам.

### 5.2. Критерии выставления оценок

Баллы/ Шкала ECTS	Оценка по дисциплине	Критерии оценки результатов обучения по дисциплине
100-83/ А,В	«отлично»/ «зачтено (отлично)»/ «зачтено»	<p>Выставляется обучающемуся, если он глубоко и прочно усвоил теоретический и практический материал, может продемонстрировать это на занятиях и в ходе промежуточной аттестации.</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>Оценка по дисциплине выставляется обучающемуся с учётом результатов текущей и промежуточной аттестации.</p> <p>Компетенции, закреплённые за дисциплиной, сформированы на уровне – «высокий».</p>
82-68/ С	«хорошо»/ «зачтено (хорошо)»/ «зачтено»	<p>Выставляется обучающемуся, если он знает теоретический и практический материал, грамотно и по существу излагает его на занятиях и в ходе промежуточной аттестации, не допуская существенных неточностей.</p> <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>Оценка по дисциплине выставляется обучающемуся с учётом результатов текущей и промежуточной аттестации.</p> <p>Компетенции, закреплённые за дисциплиной, сформированы на уровне – «хороший».</p>
67-50/ D,Е	«удовлетвори- тельно»/ «зачтено (удовлетвори- тельно)»/ «зачтено»	<p>Выставляется обучающемуся, если он знает на базовом уровне теоретический и практический материал, допускает отдельные ошибки при его изложении на занятиях и в ходе промежуточной аттестации.</p> <p>Обучающийся испытывает определённые затруднения в применении теоретических положений при решении практических задач профессиональной направленности</p>

Баллы/ Шкала ECTS	Оценка по дисциплине	Критерии оценки результатов обучения по дисциплине
		<p>стандартного уровня сложности, владеет необходимыми для этого базовыми навыками и приёмами. Демонстрирует достаточный уровень знания учебной литературы по дисциплине. Оценка по дисциплине выставляются обучающемуся с учётом результатов текущей и промежуточной аттестации. Компетенции, закреплённые за дисциплиной, сформированы на уровне – «достаточный».</p>
49-0/ F,FX	«неудовлетворительно»/ не зачтено	<p>Выставляется обучающемуся, если он не знает на базовом уровне теоретический и практический материал, допускает грубые ошибки при его изложении на занятиях и в ходе промежуточной аттестации. 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. Демонстрирует фрагментарные знания учебной литературы по дисциплине. Оценка по дисциплине выставляются обучающемуся с учётом результатов текущей и промежуточной аттестации. 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: подготовка к семинарским занятиям (8 занятий – 5 баллов максимум каждое); реферат (20 баллов максимум); коллоквиум по контрольным вопросам с выполнением практического задания (максимально - 40 баллов). Для получения удовлетворительной оценки необходимо набрать минимум 60 баллов. Самостоятельная работа: при подготовке к семинару студент заблаговременно знакомится с источниками, указанными в плане семинарских занятий в качестве основных. Кроме того, каждый студент по согласованию с преподавателем выбирает из списка дополнительной литературы научный текст для реферирования; реферат представляется в форме подробного хэндаута и презентации. При подготовке к коллоквиуму предполагается знакомство с полным списком обязательной литературы.

В качестве домашних заданий предлагаются задания следующих типов

Реферирование научного сочинения из рекомендованного списка

Выполнение практического задания: работа с конкретной системой машинного перевода ЭТАП-3. Наблюдение за алгоритмом синтаксического анализа.

Выполнение практического задания: работа с конкретной системой машинного перевода ЭТАП-3. Наблюдение за решением задачи разрешения неоднозначностей разных типов.

Выполнение практического задания: работа с системой UNL.

Выполнение практического задания: построение морфологических структур предложений разных типов на примере различных естественных языков.

Выполнение практических заданий: 1) ручное построение синтаксических структур предложений разных типов на примере различных естественных языков, 2) самостоятельное построение синтаксических правил на специальном формальном языке.

Выполнение практического задания: создание словарных статей словарей разных языков для машинного перевода.

Выполнение практического задания: написание пробных правил трансфера для перевода с русского языка на английский и обратно.

Выполнение практического задания: работа с базой данных лексических функций.

Выполнение практического задания: работа с доступными в Интернете статистическими системами перевода и критический анализ систем лексических функций.

Выполнение практического задания: выработка навыков пользования конкретными цифровыми лингвистическими ресурсами.

#### Зачет ориентирован на следующие контрольные вопросы

1. Действующая модель языка «Смысл – Текст»: основные положения.
2. Определение предмета лингвистических дисциплин - фонетики, морфологии, синтаксиса, семантики – в свете подхода «Смысл – Текст»
3. Интегральное описание языка в смысле Ю.Д.Апресяна – основные положения.
4. Грамматики составляющих и грамматики зависимостей в свете задачи автоматической обработки текстов.
5. Гибридные грамматики и возможности их компьютерной реализации.
6. Морфологический анализ текста. Морфологическая структура слова и предложения.
7. Синтаксический анализ текста. Правильные и статистические подходы к автоматической обработке текста.
8. Синтаксическая структура предложения. Синтаксические отношения. Синтагмы.
9. Языковая неоднозначность и методы ее разрешения при автоматической обработке текста. Типы неоднозначности. Интерактивное разрешение лексической и синтаксической неоднозначности.
10. Типы систем машинного перевода. Автоматический и автоматизированный перевод. Память переводов. Интерлингва.
11. UNL (универсальный сетевой язык) как тип интерлингвы и его использование в задаче машинного перевода.
12. Словарь системы машинного перевода. Структура словарной статьи.
13. Синтаксические признаки.
14. Семантические признаки (дескрипторы).
15. Теория валентностей. Модель управления.

16. Правила межъязыкового перевода в узком смысле (трансфер).
17. Лексические функции в машинном переводе.
18. Статистические системы машинного перевода: общий подход.
19. Некоторые современные цифровые лингвистические ресурсы и их роль в задачах автоматической обработки текстов.. Аннотированные корпуса текстов. Word Net, Frame Net, Treebanks (включая SynTagRus), Semantic Web).

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

### **6.1. Список источников и литературы**

#### Основная литература

1. Bahdanau D., Cho K., Bengio Y., Neural machine translation by Jointly learning to align and translate. 2015 <https://arxiv.org/pdf/1409.0473.pdf>
2. Birch A. et al. HUME: Human UCCA-Based Evaluation of Machine Translation 2016 <https://arxiv.org/pdf/1607.00030.pdf>
3. Denkowski M., Lavie A., Meteor Universal: Language Specific Translation Evaluation for Any Target Language – Workshop on statistical machine translation, 2014 <http://www.aclweb.org/anthology/W14-3348>
4. Koehn et al. Moses: open source toolkit for statistical machine translation. 2007 <https://aclanthology.info/pdf/P/P07/P07-2045.pdf>
5. Papineni K. et al. BLEU: a method for automatic Evaluation of machine translation. 2002 <https://aclanthology.info/pdf/P/P02/P02-1040.pdf>
6. Papineni K. et al. BLEU: a method for automatic Evaluation of machine translation. 2002 <https://aclanthology.info/pdf/P/P02/P02-1040.pdf>
7. The Routledge Encyclopedia of Translation Technology. Chan Sin-Wai (Ed.) Routledge, 2015.
8. Wu Y. Et al. Google’s Neural Machine Translation System: Bridging the Gap between Human and Machine Translation 2016 [https://arxiv.org/pdf/1609.08144.pdf%20\(7\).pdf](https://arxiv.org/pdf/1609.08144.pdf%20(7).pdf)

#### Рекомендованная литература

1. 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).
2. Carl Pollard, Ivan A. Sag: Head-Driven Phrase Structure Grammar. Chicago: University of Chicago Press. 1994.
3. Nirenburg, Sergei, Harold L. Somers, and Yorick Wilks (eds.) Readings in Machine Translation. , Cambridge, Massachusetts: MIT Press. 2003
4. Boguslavsky I, Iomdin L, Nivre J. Parsing the Russian Dependency Treebank. Proceedings of COLING-2008. Manchester, 2008.
5. Philipp Koehn. Statistical Machine Translation. Cambridge University Press. 2009.
6. 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.

7. The Routledge Encyclopedia of Translation Technology. Chan Sin-Wai (Ed.) Routledge, 2015.
8. Мельчук И.А. Опыт теории лингвистических моделей класса «Смысл – Текст».– М.: Наука, 1995.
9. Апресян Ю.Д., Богуславский И.М., Иомдин Л.Л. и др. Лингвистический процессор для сложных информационных систем. М.: Наука, 1992.
10. 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.

### 6.2 Перечень ресурсов информационно-телекоммуникационной сети «Интернет»

№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

### 7. Материально-техническое обеспечение дисциплины

Занятия по курсу можно проводить с максимальной эффективностью, если проводить их в компьютерном классе с доступом в Интернет, проектором и экраном для презентаций. Необходимые программы:

№п /п	Наименование ПО	Производитель	Способ распространения
1	Adobe Master Collection CS4	Adobe	лицензионное
2	Microsoft Office 2010	Microsoft	лицензионное
3	Windows 7 Pro	Microsoft	лицензионное
7	Microsoft Share Point 2010	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



## **8. Обеспечение образовательного процесса для лиц с ограниченными возможностями здоровья**

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**

Дисциплина реализуется в Институте лингвистики УНЦ лингвистической типологии.

Цель дисциплины:

Предметом дисциплины является система описания языка с особенностями, присущими такому описанию, когда оно осуществляется в целях автоматической обработки текстов, в первую очередь машинного перевода, основанного на правилах.

Основное внимание уделяется 1) тому, как текст на входном языке анализируется на разных уровнях (морфологическом, синтаксическом, семантическом), как соотносятся грамматическое описание языка для целей автоматической обработки текста и словарь языка, 2) тому, как осуществляется преобразование глубинной структуры входного языка в структуру выходного языка и 3) тому, как синтезируется текст на выходном языке. Курс предполагает интенсивную, практическую работу магистрантов, которые должны освоить основные методы и подходы, применяемые лингвистами, разрабатывающими прикладные системы автоматической обработки текстов, и научиться самостоятельно создавать адекватные правила такой обработки.

Задачи дисциплины:

- дать представление о месте теоретической лингвистики в задачах, решаемых компьютерной лингвистикой;
- познакомить магистрантов с важнейшими областями междисциплинарных исследований на стыке лингвистики со смежными дисциплинами, в первую очередь с компьютерной наукой;
- научить магистрантов обнаруживать типологические сходства и различия естественных языков, между которыми осуществляется машинный перевод, и

- творчески использовать полученные знания;
- научить магистрантов пользоваться методами обратной связи, т.е. применять полученные при разработке автоматических систем результаты для извлечения новых знаний о естественном языке;
- познакомить магистрантов с современными подходами к решению задачи машинного перевода, в том числе с гибридными и статистическими подходами и приемами машинного обучения.

Дисциплина направлена на формирование следующих компетенций:

<b>Компетенция</b>	<b>Индикаторы компетенций</b>	<b>Результаты обучения</b>
УК-1 Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий	УК-1.2	Умеет определять круг задач и решать поставленные задачи в рамках избранных видов профессиональной деятельности, планировать собственную деятельность исходя из имеющихся ресурсов
УК-2 Способен управлять проектом на всех этапах его жизненного цикла	УК-2.2	Умеет соотносить разнородные явления и систематизировать их в рамках избранных видов профессиональной деятельности
УК-6 Способен определять и реализовывать приоритеты собственной деятельности и способы ее совершенствования на основе самооценки	УК-6.2	Умеет планировать свое рабочее время и время для саморазвития, формулировать цели личного и профессионального развития и условия их достижения, исходя из тенденций развития области профессиональной деятельности, индивидуально-личностных особенностей
ПК-2 Владеет 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-2.3	Имеет практический опыт разработки электронных языковых ресурсов; опыт применения основных методов, способов и средств получения, хранения, переработки информации
ПК-3 Способен использовать лингвистические технологии для проектирования систем автоматической обработки звучащей речи и письменного текста на естественном языке, лингвистических компонентов интеллектуальных и информационных электронных систему	ПК-3.3	Умеет пользоваться существующими системами автоматической обработки текста и звучащей речи, интеллектуальными и информационными электронными системами; проводить их сравнительный анализ; проектировать модули данных систем, составлять технические задания

По дисциплине предусмотрена промежуточная аттестация в форме зачета.  
Общая трудоемкость освоения дисциплины составляет 2 зачетных единиц.

**ЛИСТ ИЗМЕНЕНИЙ**

№	Текст актуализации или прилагаемый к РПД документ, содержащий изменения	Дата	№ протокола
1	Приложение №		