



МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
федеральное государственное бюджетное образовательное учреждение
высшего образования
«ИРКУТСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»
Институт филологии, иностранных языков и медиакоммуникации
Факультет теоретической и прикладной филологии
Кафедра русского языка и общего языкознания



Рабочая программа дисциплины

Б1.В.ДВ.03.02 Технологии корпусной лингвистики

Направление подготовки **45.03.01 Филология**

Направленность (профиль) подготовки
**«Прикладная филология: обеспечение эффективной письменной и
устной коммуникации»**

Квалификация (степень) выпускника – **бакалавр**
Форма обучения – **очная**

Согласовано с УМК ИФИЯМ
Протокол № 7 от «15» марта 2022 г.
Председатель О.Л. Михалёва

Рекомендовано кафедрой
русского языка и общего языкознания
Протокол № 6 от «10» марта 2022 г.
Зав. кафедрой М.Б. Ташлыкова

Иркутск 2022 г.

Содержание

I.		Цели и задачи дисциплины	3
II.		Место дисциплины в структуре ОПОП	3
III.		Требования к результатам освоения дисциплины	3
IV.		Содержание и структура дисциплины	6
	4.1.	Содержание дисциплины, структурированное по темам, с указанием видов учебных занятий и отведенного на них количества академических часов	6
	4.2.	План внеаудиторной самостоятельной работы обучающихся по дисциплине	7
	4.3.	Содержание учебного материала	12
	4.3.1.	Перечень семинарских, практических занятий и лабораторных работ	13
	4.3.2.	Перечень тем (вопросов), выносимых на самостоятельное изучение в рамках самостоятельной работы студентов	15
	4.4.	Методические указания по организации самостоятельной работы студентов	16
V.		Учебно-методическое и информационное обеспечение дисциплины	17
	а)	перечень литературы	
	б)	периодические издания	
	в)	список авторских методических разработок	
	г)	базы данных, поисково-справочные и информационные системы	
VI.		Материально-техническое обеспечение дисциплины	19
	6.1.	Учебно-лабораторное оборудование	
	6.2.	Программное обеспечение	
	6.3.	Технические и электронные средства обучения	
VII.		Образовательные технологии	20
VIII.		Оценочные материалы для текущего контроля и промежуточной аттестации	20

I. ЦЕЛИ И ЗАДАЧИ ДИСЦИПЛИНЫ

Цель дисциплины – познакомить студентов с концепциями корпусной лингвистики, помочь им освоить основы корпусных технологий, приобрести навыки работы с корпусами, определить место дисциплины и собственно корпусов в ряду информационно-лингвистических технологий.

Задачи дисциплины:

- ознакомить студентов с новой парадигмой в лингвистических исследованиях;
- ознакомить студентов с историей корпусных исследований;
- ознакомить студентов с языковыми и программными средствами корпусной лингвистики;
- сформировать у студентов навыки работы с программными средствами и информационными ресурсами корпусной лингвистики;
- ознакомить студентов с конкретными лингвистическими исследованиями, основанными на корпусных данных.

II. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОПОП

2.1. Учебная дисциплина «Технологии корпусной лингвистики» относится к обязательной части учебного плана.

2.2. Для изучения данной учебной дисциплины необходимы знания, умения и навыки, формируемые предшествующими дисциплинами:

Б1.О.01 Введение в языкознание

Б1.О.11.01. Фонетика и фонология

Б1.О.11.02. Лексикология

Б1.О.11.03 Морфология

Б1.В.ДВ.01.02 Математические методы в лингвистике

2.3. Перечень последующих учебных дисциплин, для которых необходимы знания, умения и навыки, формируемые данной учебной дисциплиной:

Б1.В.13 Функциональная стилистика

Б1.В.17 Основы дискурс-анализа

Б1.В.22 Лингвистическая экспертиза текста

Б1.В.ДВ.03.01 Технологии обработки звучащей речи

Б1.В.ДВ.08.01 Курсовая работа по языкознанию-2

Полученные в ходе освоения дисциплины «Основы корпусной лингвистики» знания, умения и навыки составляют основу для понимания роли корпусных методов при анализе текстов значительного объема, знания о возможности использования корпусных данных в качестве доказательной базы в рамках лингвистической экспертизы.

III. ТРЕБОВАНИЯ К РЕЗУЛЬТАТАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Процесс изучения дисциплины направлен на формирование следующих компетенций:

• ПК-1 – способен использовать систематизированные теоретические и практические знания в области языкознания и литературоведения для постановки и решения научных и образовательных задач;

• ПК-6 – способен использовать систематизированные теоретические и практические знания в области языкознания и литературоведения для постановки и решения научных и образовательных задач.

**Перечень планируемых результатов обучения,
соотнесенных с индикаторами достижения компетенций**

Компетенция	Индикаторы компетенций	Результаты обучения
<p><i>ПК-1</i> – способен применять широкий спектр методов понимания и создания устных и письменных сообщений, включая электронные</p>	<p><i>ИДК ПК-1.1</i> Анализирует и структурирует информацию, сопоставляет различные сообщения, выделяя ключевые для решения профессиональных задач аспекты</p>	<p>Знать:</p> <ul style="list-style-type: none"> • принципы работы корпусов разных типов и информационно-поисковых систем. <p>Уметь:</p> <ul style="list-style-type: none"> • выделять в сообщении ключевые для решения профессиональных задач аспекты; • сопоставлять разные сообщения с точки зрения их эффективности при поиске информации. <p>Владеть: методами корпусного анализа устной, письменной и виртуальной коммуникации</p>
<p><i>ПК-5</i> – способен использовать систематизированные теоретические и практические знания в области языкознания и литературоведения для постановки и решения научных и образовательных задач</p>	<p><i>ИДК ПК-5.1</i> Применяет полученные знания в области теории и истории основного изучаемого языка (языков) и литературы (литератур), теории коммуникации, филологического анализа и интерпретации текста в собственной научно-исследовательской деятельности.</p>	<p>Знать</p> <ul style="list-style-type: none"> • основные принципы работы морфологических анализаторов и корпус-менеджеров. <p>Уметь</p> <ul style="list-style-type: none"> • применять корпусные методы для описания и анализа языковых фактов; • использовать современные корпусные инструменты. <p>Владеть навыками поиска в лингвистических базах данных и других информационно-поисковых системах.</p>
	<p><i>ИДК ПК-5.3</i> Способен формулировать аргументированные умозаключения и выводы.</p>	<p>Знать:</p> <ul style="list-style-type: none"> • теоретические основания предобработки языкового материала; • корпусные методы, которые могут быть использованы для анализа языковых единиц. <p>Уметь:</p> <ul style="list-style-type: none"> • использовать языковые корпуса для получения надежных выводов о том или ином языковом явлении, процессе. <p>Владеть:</p>

		<ul style="list-style-type: none">• навыками оценки результатов, полученных при поиске в языковых корпусах.
--	--	---

IV. СОДЕРЖАНИЕ И СТРУКТУРА ДИСЦИПЛИНЫ

Трудоемкость дисциплины составляет 2 зачетных единицы, 72 часа. Форма промежуточной аттестации – экзамен.

4.1. Содержание дисциплины, структурированное по темам, с указанием видов учебных занятий и отведенного на них количества академических часов

№ п/п	Раздел дисциплины/темы	семестр	Всего часов	Из них – практическая подготовка обучающихся	Виды учебной работы, включая самостоятельную работу обучающихся и трудоемкость (в часах)				Формы текущего контроля успеваемости
					Контактная работа преподавателя с обучающимися			Самостоятельная работа	
					Лекции	Семинарские (практические занятия)	Консультации		
1.	Корпусная лингвистика. История корпусной лингвистики	4	6			4		2	устный опрос
2.	Классификация корпусов	4	6			2		4	устный опрос
3.	Корпусная разметка	4	8			4		4	устный опрос
4.	Этапы развития автоматического анализа текста	4	6			2		4	устный опрос
5.	Морфологическая разметка	4	10			6		4	проверочная работа
6.	Современные корпуса русского языка	4	10			6		4	проверочная работа
7.	Создание собственного корпуса	4	8			4		4	анализ текста
8.	Корпусная лингвистика: сферы применения	4	8			4		4	устный опрос
			72 (в том числе 2 часа КСР)			32		30	8

4.2. План внеаудиторной самостоятельной работы обучающихся по дисциплине

Семес тр	Название раздела / темы	Самостоятельная работа обучающихся			Оценочное средство	Учебно- методическое обеспечение СР
		Вид СР	Сроки	Трудоем кость (в часах)		
4	Корпусная лингвистика. История корпусной лингвистики	Работа с основной и дополнительной литературой, конспектирование. Подготовка к устному опросу.	2 неделя семестра	2	устный опрос	Копотев, М. В. Введение в корпусную лингвистику [Текст] : учеб. пособие для студ. филол. и лингв. специальностей университетов / М. В. Копотев; Animedia Company, 2014. – 230 с. Захаров, В. П. Корпусная лингвистика. / В. П. Захаров, С. Ю. Богданова. 3-е изд. ред. СПб. : Издательство Санкт-Петербургского университета, 2020. – 234 стр. Николаев, И.С. Прикладная и компьютерная лингвистика [Текст] / И.С.Николаев, О.В Митренина., Т.М. Ландо (Ред.) ; - М: URSS, 2017. – 320 с.
4	Классификация корпусов	Работа с основной и дополнительной литературой, конспектирование. Подготовка к устному опросу.	3 неделя семестра	4	устный опрос	Копотев, М. В. Введение в корпусную лингвистику [Текст] : учеб. пособие для студ. филол. и лингв. специальностей университетов / М. В. Копотев; Animedia Company, 2014. – 230 с. Захаров, В. П. Корпусная

						лингвистика. / В. П. Захаров, С. Ю. Богданова. 3-е изд. ред. СПб. : Издательство Санкт-Петербургского университета, 2020. – 234 стр.
4	Корпусная разметка	Работа с основной и дополнительной литературой, конспектирование. Подготовка к устному опросу.	4-5 неделя семестра	4	устный опрос	Копотев, М. В. Введение в корпусную лингвистику [Текст] : учеб. пособие для студ. филол. и лингв. специальностей университетов / М. В. Копотев; Animedia Company, 2014. – 230 с. Захаров, В. П. Корпусная лингвистика. / В. П. Захаров, С. Ю. Богданова. 3-е изд. ред. СПб. : Издательство Санкт-Петербургского университета, 2020. – 234 стр. Ляшевская, О. Н. Корпусные инструменты в грамматических исследованиях русского языка [Текст] / О. Н. Ляшевская. - Москва : Рукописные памятники Древней Руси, 2016. - 518 с.
4	Этапы развития автоматического анализа текста	Работа с основной и дополнительной литературой, конспектирование. Подготовка к устному опросу.	6 неделя семестра	4	устный опрос	Копотев, М. В. Введение в корпусную лингвистику [Текст] : учеб. пособие для студ. филол. и лингв. специальностей университетов / М. В. Копотев; Animedia Company, 2014. – 230 с.

						Захаров, В. П. Корпусная лингвистика. / В. П. Захаров, С. Ю. Богданова. 3-е изд. ред. СПб. : Издательство Санкт-Петербургского университета, 2020. – 234 стр.
4	Морфологическая разметка	Работа с основной и дополнительной литературой, конспектирование. Подготовка к проверочной работе.	8-9 неделя семестра	4	проверочная работа	Копотев, М. В. Введение в корпусную лингвистику [Текст] : учеб. пособие для студ. филол. и лингв. специальностей университетов / М. В. Копотев; Animedia Company, 2014. – 230 с. Захаров, В. П. Корпусная лингвистика. / В. П. Захаров, С. Ю. Богданова. 3-е изд. ред. СПб. : Издательство Санкт-Петербургского университета, 2020. 234 стр. Ляшевская, О. Н. Корпусные инструменты в грамматических исследованиях русского языка [Текст] / О. Н. Ляшевская. - Москва : Рукописные памятники Древней Руси, 2016. - 518 с.
4	Современные корпуса русского языка	Работа с основной и дополнительной литературой, конспектирование. Подготовка к проверочной работе.	10-11 неделя семестра	4	проверочная работа	Копотев, М. В. Введение в корпусную лингвистику [Текст] : учеб. пособие для студ. филол. и лингв. специальностей университетов / М. В. Копотев; Animedia

						<p>Companu, 2014. – 230 с. Захаров, В. П. Корпусная лингвистика. / В. П. Захаров, С. Ю. Богданова. 3-е изд. ред. СПб. : Издательство Санкт-Петербургского университета, 2020. 234 стр. Ляшевская, О. Н. Корпусные инструменты в грамматических исследованиях русского языка [Текст] / О. Н. Ляшевская. - Москва : Рукописные памятники Древней Руси, 2016. - 518 с. Пиперски, А. Ч. Русский язык и корпусное разнообразие // Компьютерная лингвистика и интеллектуальные технологии. По материалам международной конференции «Диалог». Вып. 19 (26). С. 84–95.</p>
4	Создание собственного корпуса	Работа с основной и дополнительной литературой, конспектирование. Подготовка к анализу текста с использованием морфологического анализатора и корпус-менеджера .	12-14 неделя семестра	4	анализ текста	<p>Копотев, М. В. Введение в корпусную лингвистику [Текст] : учеб. пособие для студ. филол. и лингв. специальностей университетов / М. В. Копотев; Animedia Companu, 2014. – 230 с. Захаров, В. П. Корпусная лингвистика. / В. П. Захаров,</p>

						С. Ю. Богданова. 3-е изд. ред. СПб. : Издательство Санкт-Петербургского университета, 2020. – 234 стр.
4	Корпусная лингвистика: сферы применения	Работа с основной и дополнительной литературой, конспектирование. Подготовка к устному опросу.	15-16 неделя семестра	4	устный опрос	Копотев, М. В. Введение в корпусную лингвистику [Текст] : учеб. пособие для студ. филол. и лингв. специальностей университетов / М. В. Копотев; Animedia Company, 2014. – 230 с. Захаров, В. П. Корпусная лингвистика. / В. П. Захаров, С. Ю. Богданова. 3-е изд. ред. СПб. : Издательство Санкт-Петербургского университета, 2020. – 234 стр.
Общий объем самостоятельной работы по дисциплине (час) – 30						
Из них объем самостоятельной работы с использованием электронного обучения и дистанционных образовательных технологий (час) 30						

4.3. Содержание учебного материала

1. Корпусная лингвистика. История корпусной лингвистики

1.1. Определение и особенности языкового корпуса. Основные свойства лингвистического корпуса (репрезентативность, сбалансированность, разметка).

1.2. Предшественники корпусных инструментов. Конкордансы. Словари на основе реального словоупотребления. Грамматики на основе реального словоупотребления. Частотные списки и словари. Современные корпуса.

2. Классификация корпусов

Основания для классификации корпусов: язык текстов, тип текстов, представленность языкового материала, объем и полнота данных, тип доступа, тип разметки.

3. Корпусная разметка

3.1. Типы разметки. Экстралингвистическая (метатекстовая) и лингвистическая разметка. Теоретические основания корпусной разметки.

3.2. Типы лингвистического аннотирования. Фонетическая и просодическая разметка. Морфологическая разметка. Семантическая разметка. Синтаксическая разметка.

3.3. Принципы разметки общедоступного корпуса.

4. Этапы развития автоматического анализа текста

Токенизация. Лемматизация. Частеречная разметка. Полная морфологическая разметка. Семантические сети.

5. Морфологическая разметка

5.1. Морфологические анализаторы. Алгоритмы, лежащие в основе морфологических анализаторов.

5.2. Бессловарные и словарные методы. Представление со словарем основ. Анализ, построенный на словаре окончаний. Словарь словоформ.

5.3. Проблемы морфологического аннотирования русского языка.

6. Современные корпуса русского языка

6.1. Национальный корпус русского языка. Национальный корпус русского литературного языка.

6.2. Открытый корпус (opencorpora). Генеральный интернет-корпус русского языка (webcorpora). Русскоязычный корпус проекта Arapea. RuTenTen. Русскоязычный корпус коротких текстов (RuTweetCorp).

6.3. Рассказы о свидениях и другие корпуса звучащей речи. Русский мультимедийный дискурс.

7. Создание собственного корпуса

7.1. Выбор материала в соответствии с основными характеристиками лингвистического корпуса. Инструменты для создания корпуса.

7.2. Конкордансеры и корпусные менеджеры. Морфологический анализатор Mystem. Основные опции, примеры вызова, расшифровка грамем.

7.3. Конкордансеры и корпусные менеджеры. Корпус-менеджер AntConc. Списки коллокатов к опорному элементу, статистики обрабатываемых текстов, набор ключевых слов текста.

7.4. Форматы представления данных.

8. Корпусная лингвистика: сферы применения

8.1. Дескриптивное описание языка. Исследования лексики. Исследования грамматических явлений.

8.2. Исследования устной речи. Исследования художественных текстов.

8.3. Социолингвистические исследования.

8.4. Преподавание русского языка как родного, неродного, иностранного.

8.5. Лингвистическая экспертиза текста.

4.3.1. Перечень семинарских, практических занятий и лабораторных работ

№ п/п	№ раздела и темы дисциплины	Наименование семинаров, практических и лабораторных работ	Трудоемкость (час.)		Оценочные средства	Формируемые компетенции и индикаторы
			Всего часов	Из них – практическая подготовка		
1	2	3	4	5	6	7
1.	1.1.	Определение и особенности языкового корпуса. Основные свойства лингвистического корпуса (репрезентативность, сбалансированность, разметка).	2		устный опрос	ПК-1. ПК-1.1 ПК-5. ПК-5.1
2.	1.2.	Предшественники корпусных инструментов. Конкордансы. Словари на основе реального словоупотребления. Грамматики на основе реального словоупотребления. Частотные списки и словари. Современные корпуса.	2		устный опрос	ПК-1. ПК-1.1 ПК-5. ПК-5.1
3.	2.	Основания для классификации корпусов: язык текстов, тип текстов, представленность языкового материала, объем и полнота данных, тип доступа, тип разметки.	2		устный опрос	ПК-1. ПК-1.1 ПК-5. ПК-5.1
4.	3.1	Типы разметки. Экстралингвистическая (метатекстовая) и лингвистическая разметка. Теоретические основания корпусной разметки.	2		устный опрос	ПК-5. ПК-5.1 ПК-5. ПК-5.3
5.	3.2	Типы лингвистического аннотирования. Фонетическая и просодическая разметка.	2		устный опрос	ПК-5. ПК-5.1 ПК-5. ПК-5.3

		Морфологическая разметка. Семантическая разметка. Синтаксическая разметка				
6.	4.	Токенизация. Лемматизация. Частеречная разметка. Полная морфологическая разметка. Семантические сети.	2		устный опрос	ПК-1. ПК-1.1 ПК-5. ПК-5.1
7.	5.1.	Морфологические анализаторы. Алгоритмы, лежащие в основе морфологических анализаторов.	2		проверочная работа	ПК-5. ПК-5.1 ПК-1. ПК-1.1
8.	5.2	Бессловарные и словарные методы. Представление со словарем основ. Анализ, построенный на словаре окончаний. Словарь словоформ.	2		устный опрос	ПК-1. ПК-1.1 ПК-5. ПК-5.1
9.	5.3	Проблемы морфологического аннотирования русского языка.	2		устный опрос	ПК-1. ПК-1.1 ПК-5. ПК-5.1
10.	6.1	Национальный корпус русского языка. Национальный корпус русского литературного языка.	2		проверочная работа	ПК-5. ПК-5.1 ПК-1. ПК-1.1
11.	6.2	Открытый корпус (opencorpus). Генеральный интернет-корпус русского языка (webcorpus). Русскоязычный корпус проекта Arapea. RuTenTen. Русскоязычный корпус коротких текстов (RuTweetCorp).	2		проверочная работа	ПК-5. ПК-5.1 ПК-1. ПК-1.1
12.	6.3	Рассказы о сновидениях и другие корпуса звучащей речи. Русский мультимедийный дискурс.	2		проверочная работа	ПК-5. ПК-5.1 ПК-5. ПК-5.3
13.	7.1, 7.2	Выбор материала в соответствии с основными характеристиками лингвистического корпуса. Инструменты для создания корпуса. Конкордансеры и корпусные менеджеры. Морфологический анализатор Mystem. Основные опции, примеры вызова, расшифровка грамем.	2		анализ текста	ПК-5. ПК-5.1 ПК-5. ПК-5.3

14.	7.3,7.4	Корпус-менеджер AntConc. Списки коллокатов к опорному элементу, статистики обрабатываемых текстов, набор ключевых слов текста. Форматы представления данных.	2		анализ текста	ПК-5. ПК-5.1 ПК-1. ПК-1.1
15.	8.1., 8.2.	Дескриптивное описание языка. Исследования лексики. Исследования грамматических явлений. Исследования устной речи. Исследования художественных текстов.	2		устный опрос	ПК-1. ПК-1.1 ПК-5. ПК-5.1
16.	8.3., 8.4., 8.5.	Социолингвистические исследования. Преподавание русского языка как родного, неродного, иностранного. Лингвистическая экспертиза текста	2		устный опрос	ПК-1. ПК-1.1 ПК-5. ПК-5.1

4.3.2. Перечень тем (вопросов), выносимых на самостоятельное изучение студентами в рамках самостоятельной работы (СРС)

№ п/п	Тема	Задание	Формируемая компетенция	ИДК
1.	Предшественники корпусных инструментов. Словари на основе реального словоупотребления.	Сравните иллюстративную зону словарных статей, представленную в «Словаре Академии Российской» и «Толковом словаре живого великорусского языка» В. И. Даля. Какой из словарей составлен на основании реального словоупотребления?	ПК-5.	ПК-5.1
2.	Современные корпуса.	Поищите в Гугле или Яндекске сочетание «лингвистический корпус» с ограничением за последние 5 лет. Составьте список корпусов, появившихся за это время.	ПК-5.	ПК-5.1
3.	Открытый корпус (орпсорога). Генеральный интернет-корпус русского языка (webcorpora). Русскоязычный корпус проекта Агапеа. RuTenTen. Русскоязычный корпус коротких текстов (RuTweetCorp).	Сравните интерфейс корпусов орпсорога, webcorpora, RuTenTen. Сформулируйте, в чем заключаются основные отличия системы поиска в перечисленных корпусах. Какие ограничения, связанные с характеристиками текстов составляющих перечисленные корпусы, должен учитывать исследователь?	ПК-5.	ПК-5.1

4.4. Методические указания по организации самостоятельной работы студентов

Раздел 1. В результате работы с основной и дополнительной литературой и конспектирования источников студент должен уметь раскрыть содержание следующих тем: 1. Определение и особенности языкового корпуса. 2. Основные свойства лингвистического корпуса (репрезентативность, сбалансированность, разметка). 3. Предшественники корпусных инструментов. 4. Конкордансы. 5. Словари на основе реального словоупотребления. Грамматики на основе реального словоупотребления. 5. Частотные списки и словари. 6. Современные корпуса.

Раздел 2. В результате работы с основной и дополнительной литературой и конспектирования источников студент должен уметь раскрыть содержание следующих тем: 1. Основания для классификации корпусов. Язык текстов. 2. Основания для классификации корпусов. Тип текстов. 3. Основания для классификации корпусов. Представленность языкового материала. 4. Основания для классификации корпусов. Объем и полнота данных. 5. Основания для классификации корпусов. Тип доступа. 6. Основания для классификации корпусов. Тип разметки.

Раздел 3. В результате работы с основной и дополнительной литературой и конспектирования источников студент должен уметь раскрыть содержание следующих тем: 1. Типы разметки. 2. Экстралингвистическая (метатекстовая) и лингвистическая разметка. 3. Теоретические основания корпусной разметки. 4. Фонетическая и просодическая разметка. 5. Морфологическая разметка. 6. Семантическая разметка. 7. Синтаксическая разметка. 8. Принципы разметки общедоступного корпуса.

Раздел 4. В результате работы с основной и дополнительной литературой и конспектирования источников студент должен уметь раскрыть содержание следующих тем: 1. Токенизация. 2. Лемматизация. 3. Частеречная разметка. 4. Полная морфологическая разметка. 5. Семантические сети.

Раздел 5. В результате работы с основной и дополнительной литературой и конспектирования источников студент должен уметь раскрыть содержание следующих тем: 1. Морфологические анализаторы. Алгоритмы, лежащие в основе морфологических анализаторов. 2. Бессловарные и словарные методы. 3. Представление со словарем основ. 4. Анализ, построенный на словаре окончаний. 5. Словарь словоформ. 6. Проблемы морфологического аннотирования русского языка.

В результате подготовки к проверочной работе студент должен уметь оценивать ограничения, которые влечет за собой использование того или иного морфологического анализатора, комментировать причины, по которым возникают проблемы, связанные с морфологическим аннотированием русского языка

Раздел 6. В результате работы с основной и дополнительной литературой и конспектирования источников студент должен уметь раскрыть содержание следующих тем: 1. Национальный корпус русского языка. Национальный корпус русского литературного языка. 2. Открытый корпус (opencorpus). Генеральный интернет-корпус русского языка (webcorpus). Русскоязычный корпус проекта Aranea. RuTenTen. Русскоязычный корпус коротких текстов (RuTweetCorp). 3. Рассказы о свидениях и другие корпуса звучащей речи. Русский мультимедийный дискурс.

В результате подготовки к проверочной работе студент должен уметь применять основные методы поиска в предложенных корпусах на практике.

Раздел 7. В результате работы с основной и дополнительной литературой и конспектирования источников студент должен уметь раскрыть содержание следующих тем: 1. Выбор материала в соответствии с основными характеристиками лингвистического корпуса. 2. Инструменты для создания корпуса. 3. Конкордансеры и корпусные менеджеры. 4. Морфологический анализатор Mystem. Основные опции, примеры вызова, расшифровка граммем. 5. Корпус-менеджер AntConc. Списки коллокатов к опорному элементу, статистики обрабатываемых текстов, набор ключевых слов текста. 6. Форматы представления данных.

В результате подготовки к проверочной работе студент должен уметь использовать основные опции морфологического анализатора Mystem, способы предобработки текста для работы с корпус-менеджером AntConc, специфику представления результатов автоматического анализа текста.

Раздел 8. В результате работы с основной и дополнительной литературой и конспектирования источников студент должен уметь раскрыть содержание следующих тем: 1. Дескриптивное описание языка. 2. Исследования лексики. 3. Исследования грамматических явлений. 4. Исследования устной речи. Исследования художественных текстов. 5. Социолингвистические исследования. 6. Преподавание русского языка как родного, неродного, иностранного. 7. Лингвистическая экспертиза текста.

4.5. Примерная тематика курсовых работ (проектов)

Курсовые работы в рамках данной дисциплины не предусмотрены.

V. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ:

а) перечень литературы

основная литература

Копотев, М. В. Введение в корпусную лингвистику [Текст] : учеб. пособие для студ. филол. и лингв. специальностей университетов / М. В. Копотев; Animedia Company, 2014. – 230 с.

Захаров, В. П. Корпусная лингвистика. / В. П. Захаров, С. Ю. Богданова. 3-е изд. ред. СПб. : Издательство Санкт-Петербургского университета, 2020. – 234 стр.

дополнительная литература

Ляшевская, О. Н. Корпусные инструменты в грамматических исследованиях русского языка [Текст] / О. Н. Ляшевская. - Москва : Рукописные памятники Древней Руси, 2016. - 518 с.

Большакова, Е.И. Автоматическая обработка текстов на естественном языке и анализ данных : учеб. пособие / Большакова Е.И., Воронцов К.В., Ефремова Н.Э., Клышинский Э.С., Лукашевич Н.В., Сапин А.С. — М.: Изд-во НИУ ВШЭ, 2017. – 269 с.

Николаев, И.С. Прикладная и компьютерная лингвистика [Текст] / И.С.Николаев, О.В Митренина., Т.М. Ландо (Ред.) ; - М: URSS, 2017. – 320 с.

Базарова, Б.Б. Введение в корпусную лингвистику [Текст] : учебно-методическое пособ. / Б. Б. Базарова ; Министерство образования и науки Российской Федерации, Бурятский государственный университет. – Улан-Удэ : Изд-во Бурятского госуниверситета, 2016. – 75 с.

б) периодические издания

Пиперски, А. Ч. Русский язык и корпусное разнообразие // Компьютерная лингвистика и интеллектуальные технологии. По материалам международной конференции «Диалог». Вып. 19 (26). С. 84–95.

в) список авторских методических разработок

Методические материалы по курсу «Основы корпусной лингвистики» размещены на образовательном портале belca.isu.ru

г) базы данных, поисково-справочные и информационные системы

Корпуса русского языка

1. <http://www.ruscorpora.ru/>

<http://www.ruscorpora.ru/corpora-other.html>

Национальный корпус русского языка – информационно-справочная система, основанная на собрании русских текстов в электронной форме

2. <http://www.slaviska.uu.se/korpus.htm>

Упсальский корпус русского языка

3. <http://www.sfb441.uni-tuebingen.de/b1/rus/korpora.html>

Тюбингенский корпус русского языка

4. <http://cfrl.ru/>

Машинный фонд русского языка

5. <http://www.philol.msu.ru/~lex/corpus/>

Компьютерный корпус текстов русских газет конца XX века

6. <http://opencorpora.org/>

Открытый корпус (Opencorpora)

7. <http://h248.it.helsinki.fi/hanco/index.html>

Хельсинкский аннотированный корпус (ХАНКО)

8. <http://narusco.ru/resources.htm>

Национальный корпус русского литературного языка (НКРЛЯ)

9. <http://unesco.uniba.sk/guest/index.html>

Русскоязычный корпус проекта Aranea

10. <http://www.webcorpora.ru/contacts>

Генеральный интернет-корпус русского языка (ГИКРЯ)

11. <https://www.sketchengine.eu/rutenten-russian-corpus/>

ruTenTen – Russian corpus from the web

12. <https://study.mokoron.com/>

Русскоязычный корпус коротких текстов RuTweetCorp

Электронные библиотеки

1. [Научная электронная библиотека eLIBRARY.RU](#)

2. [Коллекция журналов издательства «НАУКА»](#)

3. [Журналы Института научной информации по общественным наукам \(ИНИОН РАН\)](#)

4. [Университетская информационная система РОССИЯ \(УИС РОССИЯ\)](#)

5. [Журнал Science издательства American Association for the Advancement of Science \(AAAS\)](#)

6. <http://www.philology.ru/>

Библиотека филологических текстов (статей, монографий)

7. <http://philologos.narod.ru/>

Материалы по теории языка и литературы

8. <http://www.scribd.com/>

База письменных документов (научных статей, монографий, художественных текстов etc.) на разных языках

9. <http://ellib.library.isu.ru>

Электронная библиотека «Труды ученых ИГУ»

10. <http://library.cjes.org>

Библиотека Центра экстремальной журналистики

Другие сайты

10. <http://www.gramota.ru>

Словари и справочники, представленные на портале

11. <http://www.grammar.ru>

Словари и справочники, представленные на портале

12. <http://linguistlist.org/>
«The world's largest online linguistic resource»: информация о конференциях, публикации, каталог ссылок на другие лингвистические ресурсы etc.
13. <http://www.ruthenia.ru/web/rusweb.html>
«Русистика на Вебе»: ссылки на сайты филологических факультетов, отделений русистики и славистики
14. <http://educa.isu.ru/>; <http://belca.isu.ru/>
Образовательный портал Иркутского государственного университета, образовательный портал ИФИЯМ

VI. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

6.1. Учебно-лабораторное оборудование.

<p>Специальные помещения</p> <p>Учебная аудитория (компьютерный класс) для проведения занятий лекционного и семинарского типа.</p>	<p>Оборудование аудитории</p> <p>Мебель:</p> <ul style="list-style-type: none"> • специализированная учебная мебель на 17 посадочных мест; • стол-кафедра – 1 шт.; • стол преподавателя – 1 шт.; стул – 1 шт. <p>Технические средства обучения, служащие для представления учебной информации большой аудитории по дисциплине:</p> <ul style="list-style-type: none"> • стационарный ПК с неограниченным доступом к сети Интернет – 20 шт.; • принтер – 1; • сканер – 1; • колонки Sven. <p>Программное обеспечение: офисный пакет приложений Microsoft Office</p>
<p>Специальные помещения</p> <p>Компьютерный класс (учебная аудитория) для групповых и индивидуальных консультаций, организации самостоятельной работы, в том числе научно-исследовательской.</p>	<p>Оборудование аудитории</p> <p>Мебель:</p> <ul style="list-style-type: none"> • специализированная учебная мебель на 17 посадочных мест; • стол-кафедра – 1 шт.; • стол преподавателя – 1 шт.; стул – 1 шт. <p>Технические средства обучения, служащие для представления учебной информации большой аудитории по дисциплине:</p> <ul style="list-style-type: none"> • стационарный ПК с неограниченным доступом к сети Интернет – 20 шт.; • принтер – 1; • сканер – 1; • колонки Sven. <p>Программное обеспечение: офисный пакет приложений Microsoft Office</p>

Специализированное учебное оборудование не используется.

6.2. Программное обеспечение

Университет обеспечен необходимым комплектом лицензионного и свободно распространяемого программного обеспечения, в том числе отечественного производства (подробная информация размещена на сайте ИГУ)

№ п/п	Наименование
1.	Adobe Reader DC 2019.008.20071
2.	IrfanView 4.42
3.	Foxit PDF Reader 8.0
4.	Google Chrome
5.	Java 8
6.	Kaspersky Endpoint Security для бизнеса- Стандартный Russian Edition. 250-499.
7.	Microsoft Office Enterprise 2007 Russian Academic OPEN No Level
8.	Mozilla Firefox
9.	Office 365 ProPlus for Students
10.	OpenOffice 4.1.3
11.	Opera 45
12.	PDF24Creator 8.0.2
13.	Skype 7.30.0
14.	VLC Player 2.2.4
15.	WinPro 10 RUS Upgrd OLP NL Acdmc.
16.	7zip 18.06

6.3. Технические и электронные средства:

Авторские презентации к лекциям и практическим занятиям, фрагменты открытых лекций по проблематике.

С подробным перечнем можно ознакомиться по ссылке:

<https://belca.isu.ru/course/view.php?id=3136>

VII. ОБРАЗОВАТЕЛЬНЫЕ ТЕХНОЛОГИИ:

Преподавание дисциплины осуществляется в форме лекционных и практических занятий, а также включает самостоятельную работу студентов по освоению теоретического и дидактического материала. Промежуточная отчетность предполагает анализ текстов с использованием автоматических методов, выполнение проверочных работ, для проверки усвоения пройденной темы.

В ходе проведения практических занятий используются кейс-технологии, дискуссии, саморедактирование и взаимное редактирование, с использованием справочных ресурсов интернета, Национального корпуса русского языка и т.п.

Используются интерактивные формы проведения занятий (компьютерные симуляции, тренинги) с целью формирования и развития профессиональных навыков студентов. Широко применяются мультимедийные средства. Сетевые компьютерные технологии включают использование ресурсов Интернет и виртуального университета.

VIII. ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ДЛЯ ТЕКУЩЕГО КОНТРОЛЯ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ

Оценочные средства для входного контроля

Для обеспечения входного контроля используется самостоятельная работа, ориентированная на проверку остаточных знаний, полученных во время обучения на первом курсе и первом семестре второго курса.

Примерная программа, на основе которой формулируются вопросы для самостоятельной работы

1. Основные единицы языка: предложение, слово, морфема, фонема.
2. Различия в пределах общенародного языка, связанные с внутренним членением языкового коллектива.
3. Литературный язык.
4. Функционально-стилистическая дифференциация языка.
5. Нелитературные типы речи.
6. Территориальная дифференциация языка.
7. Просторечие. Элементы социальной дифференциации в языке.
8. Членение высказывания и основные единицы языка: предложение, слово, морфема, фонема. Основные функции единиц языка.
9. Системные отношения между языковыми единицами: парадигматические, синтагматические, иерархические.
10. Проблемы определения слова. Лексическое и грамматическое значения. Слова знаменательные и служебные.
11. Моносемия и полисемия. Отличия омонимии от полисемии. Трудности их разграничения.
12. Понятие о морфеме. Экспонент и содержание морфемы. Варьирование морфемы.
13. Слова изменяемые и неизменяемые. Слово и словоформа. Основа слова и основа словоформы.
14. Грамматические значения и их формальные показатели. Способы выражения грамматических значений: аффиксация, супплетивизм, чередования (внутренняя флексия), редупликация, служебные слова, порядок слов, ударение, интонация.
15. Части речи и критерии их выделения. Система частей речи.
16. Единицы синтаксиса. Понятие синтаксических отношений и синтаксической связи. Сочинительные и подчинительные связи.
17. Средства выражения синтаксических связей: аффиксы и служебные слова, порядок слов и интонация, синтаксическое основосложение (инкорпорация). Синтаксические функции.
18. Понятие о языковом корпусе.
19. Информационно-поисковые системы.
20. Понятие о квантитативной лингвистике.

Оценочные средства текущего контроля

Текущий контроль за усвоением содержания дисциплины осуществляется в следующих формах:

- тематический опрос студентов на практических занятиях;
- выполнение заданий, ориентированных на усвоение требований к системе и структуре текста;
- изучение и анализ публичных выступлений;
- анализ кейсов;
- учебные тренинги и деловые игры;
- выступления (с презентацией) реферативного характера;
- выступления (с презентацией) творческого характера.
- обсуждение отдельных вопросов курса в консультационном режиме.

Проверочная работа №1 (демонстрационный вариант)

1. Бессловарный метод. Словоформа оканчивается «-алось». Укажите, какие морфологические признаки глагола по этой финали определяются: 1) с точностью в 100 %, 2) с точностью в 50 %, 3) не определяются вообще.
2. Предложит ли алгоритм, использующий словарь основ, разбор формы *стола* как глагольной формы (прош. вр., жен. р., ед. ч.)? Обоснуйте ответ, указав, какие основы в

словаре основ и какие окончания в словаре окончаний должны существовать для такого решения.

Проверочная работа №2 (демонстрационный вариант)

1. Пользуясь любым корпусом русского языка, расположите падежи в соответствии с частотой их употребления в корпусе. Совпадает ли частотный порядок падежей для существительных и прилагательных? Соотносится ли частотный порядок с традиционным порядком падежей?

2. Используя НКРЯ, кратко охарактеризуйте ударение в русских глаголах движения оканчивающихся на *-ить* в настоящем времени 2 л. ед. числа.

Анализ текста (демонстрационный вариант)

Используя Mystem, лемматизируйте текст, укажите количество лемм, использованных в тексте. Проведите морфологический анализ словоформ, снимите случаи омонимии. Используя корпус-менеджер AntConc, составьте список частотности частей речи для данного текста, выявите ключевые слова. Какие слова оказались в десятке первых в списке частотности? К какому стилю относится предложенный текст? Аргументируйте свой ответ, подкрепляя его данными, полученными с помощью корпусного менеджера.

Женщина с самого сотворения мира считается существом вредным и злокачественным. Она стоит на таком низком уровне физического, нравственного и умственного развития, что судить ее и зубоскалить над ее недостатками считает себя вправе всякий, даже лишенный всех прав прохвост и сморкающийся в чужие платки губошлеп.

Анатомическое строение ее стоит ниже всякой критики. Когда какой-нибудь солидный отец семейства видит изображение женщины «о натюрель», то всегда брезгливо морщится и сплевывает в сторону. Иметь подобные изображения на виду, а не в столе или у себя в кармане, считается моветонством. Мужчина гораздо красивее женщины. Как бы он ни был жилист, волосат и угреват, как бы ни был красен его нос и узок лоб, он всегда снисходительно смотрит на женскую красоту и женится не иначе, как после строгого выбора. Нет того Квазимодо, который не был бы глубоко убежден, что парой ему может быть только красивая женщина.

Один отставной поручик, обокравший тещу и щеголявший в жениных полусапожках, уверял, что если человек произошел от обезьяны, то сначала от этого животного произошла женщина, а потом уж мужчина. Титулярный советник Слюнкин, от которого жена запирала водку, часто говаривал: «Самое ехидное насекомое в свете есть женский пол».

Ум женщины никуда не годится. У нее волос долог, но ум короток; у мужчины же наоборот. С женщиной нельзя потолковать ни о политике, ни о состоянии курса, ни о чиншевиках. В то время, когда гимназист III класса решает уже мировые задачи, а коллежские регистраторы изучают книгу «30 000 иностранных слов», умные и взрослые женщины толкуют только о модах и военных.

Логика женщины вошла в поговорку. Когда какой-нибудь надворный советник Анафемский или департаментский сторож Дорофей заводят речь о Бисмарке или о пользе наук, то любо послушать их: приятно и умирительно; когда же чья-нибудь супруга, за неимением других тем, начинает говорить о детях или пьянстве мужа, то какой супруг воздержится, чтобы не воскликнуть: «Затарантила таранта! Ну, да и логика же, господи, прости ты меня грешного!» Изучать науки женщина неспособна. Это явствует уже из одного того, что для нее не заводят учебных заведений. Мужчины, даже идиот и кретин, могут не только изучать науки, но даже и занимать кафедры, но женщина — ничтожество ей имя! Она не сочиняет для продажи учебников, не читает рефератов и длинных академических речей, не ездит на казенный счет в ученые командировки и не утилизирует заграничных диссертаций. Ужасно неразвита! Творческих талантов у нее — ни капли. Не только великое и гениальное, но даже пошлое и шантажное пишется мужчинами, ей же

дана от природы только способность заворачивать в творения мужчин пирожки и делать из них папильотки.

Она порочна и безнравственна. От нее идет начало всех зол. В одной старинной книге сказано: «Mulier est malleus, per quem diabolus mollit et malleat univēsum mundum». Когда дьяволу приходит охота учинить какую-нибудь пакость или каверзу, то он всегда норовит действовать через женщин. Вспомните, что из-за Бель Элен вспыхнула Троянская война, Мессалина совратила с пути истины не одного паиньку... Гоголь говорит, что чиновники берут взятки только потому, что на это толкают их жены. Это совершенно верно. Пропивают, в винт проигрывают и на Амалий тратят чиновники только жалованье... Имущества антрепренеров, казенных подрядчиков и секретарей теплых учреждений всегда записаны на имя жены. Распущена женщина донельзя. Каждая богатая барыня всегда окружена десятками молодых людей, жаждущих попасть к ней в альфонсы. Бедные молодые люди!

Отечеству женщина не приносит никакой пользы. Она не ходит на войну, не переписывает бумаг, не строит железных дорог, а запирая от мужа графинчик с водкой, способствует уменьшению акцизных сборов.

Короче, она лукава, болтлива, суетна, лжива, лицемерна, корыстолюбива, бездарна, легкомысленна, зла... Только одно и симпатично в ней, а именно то, что она производит на свет таких милых, грациозных и ужасно умных душек, как мужчины... За эту добродетель простим ей все ее грехи. Будем к ней великодушны все, даже кокетки в пиджаках и те господа, которых бьют в клубах подсвечниками по мордасам.

А.П. Чехов «О женщинах»

Оценочные средства для промежуточной аттестации (в форме зачета)

Список вопросов (программа) к экзамену

1. Определение и особенности языкового корпуса. Основные свойства лингвистического корпуса (репрезентативность, сбалансированность, разметка).
2. Предшественники корпусных инструментов. Конкордансы. Словари на основе реального словоупотребления. Грамматики на основе реального словоупотребления. Частотные списки и словари.
3. Основания для классификации корпусов. Язык текстов.
4. Основания для классификации корпусов. Тип текстов.
5. Основания для классификации корпусов. Представленность языкового материала.
6. Основания для классификации корпусов. Объем и полнота данных.
7. Основания для классификации корпусов. Тип доступа.
8. Основания для классификации корпусов. Тип разметки.
9. Типы разметки. Экстралингвистическая (метатекстовая) и лингвистическая разметка. Теоретические основания корпусной разметки.
10. Типы лингвистического аннотирования. Фонетическая и просодическая разметка.
11. Типы лингвистического аннотирования. Семантическая разметка. Синтаксическая разметка.
12. Принципы разметки общедоступного корпуса.
13. Морфологические анализаторы. Алгоритмы, лежащие в основе морфологических анализаторов.
14. Бессловарные и словарные методы. Представление со словарем основ. Анализ, построенный на словаре окончаний. Словарь словоформ.
15. Проблемы морфологического аннотирования русского языка.
16. Национальные корпуса русского языка
17. Интернет-корпусы русского языка

18. Русские корпуса звучащей речи
19. Конкордансеры и корпусные менеджеры.
20. Сферы применения корпусной лингвистики. Исследования лексики.
21. Сферы применения корпусной лингвистики. Исследования грамматических явлений.
22. Сферы применения корпусной лингвистики. Исследования художественных текстов.
23. Сферы применения корпусной лингвистики. Социолингвистические исследования.
24. Сферы применения корпусной лингвистики. Преподавание русского языка как родного, неродного, иностранного.
25. Сферы применения корпусной лингвистики. Лингвистическая экспертиза текста.

