

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное автономное образовательное учреждение
высшего образования
«ТЮМЕНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»
Тобольский педагогический институт им. Д.И.Менделеева
(филиал) Тюменского государственного университета

21

УТВЕРЖДАЮ
Директор _____ Шидлов С.П.
« 28 » _____ 2020 г.



ОСНОВЫ КОРПУСНОЙ ЛИНГВИСТИКИ
Рабочая программа
для обучающихся по направлению подготовки
45.03.02 ЛИНГВИСТИКА
Профиль Теория и методика преподавания иностранных языков и культур
Форма обучения: очная

Шитиков Петр Михайлович. Основы корпусной лингвистики. Рабочая программа для обучающихся по направлению подготовки 45.03.02 Лингвистика, профиль Теория и методика преподавания иностранных языков и культур, форма обучения очная. Тобольск, 2020

Рабочая программа дисциплины (модуля) опубликована на сайте ТюмГУ: Основы корпусной лингвистики [электронный ресурс] / Режим доступа: <https://tobolsk.utmn.ru/sveden/education/#>

21

1. Пояснительная записка

Курс «Основы корпусной лингвистики» имеет значимость в трех направлениях: во-первых, знание основ этой ветви науки о языке входит в требования к кругозору любого лингвиста с высшим образованием. Эту цель можно назвать «информирующей», во-вторых, методы корпусной лингвистики являются методами прикладной лингвистики, в-третьих, могут быть чрезвычайно полезны для научного исследования в любой области прикладной лингвистики, а также для перевода текстов.

Целью дисциплины является ознакомить студентов с основами корпусной лингвистики в рамках эмпирического подхода к изучению языка.

Для достижения этой цели нужно решить следующие **задачи**:

- ознакомить студентов с предметами, задачами, методами и понятийным (терминологическим) аппаратом корпусной лингвистики;
- сформировать знания о компьютерном инструментарии корпусной лингвистики;
- дать первичное представление о возможных применениях корпусов в лингвистических исследованиях.

1.1. Место дисциплины (модуля) в структуре образовательной программы

Данная дисциплина «Основы корпусной лингвистики» относится к циклу Б1. дисциплины (модули), к дисциплинам по выбору вариативной части ОП.

Курс способствует формированию лингвистического мировоззрения, т.е. понимания законов существования языка, практическому освоению одного из наиболее перспективных направлений в современной теоретической и прикладной лингвистике, основанного на использовании языковых электронных корпусов и корпусных технологий. Для освоения данной дисциплины необходимы компетенции, сформированные у студентов в результате изучения курсов, «Математическое и компьютерное моделирование в лингвистике». Дисциплина также логически связана с курсами «Методы лингвистического исследования», «Актуальные проблемы прикладной лингвистики», «Введение в нейронные сети и машинный перевод» и другими курсами соответствующего профиля подготовки, изучаемыми одновременно и в последующих семестрах.

Знания, полученные студентами в ходе изучения курса, будут необходимы при подготовке к преддипломной практике, а также к государственной итоговой аттестации и в процессе ее.

1.2. Компетенции обучающегося, формируемые в результате освоения данной дисциплины (модуля).

Код и наименование компетенции (из ФГОС ВО)	Компонент (знаниевый/функциональный)
ПК-23 способность использовать понятийный аппарат философии, теоретической и прикладной лингвистики, переводоведения, лингводидактики и теории межкультурной коммуникации для решения профессиональных задач	Знает информацию о широком круге проблем, составляющих в настоящее время предмет изучения в области лингвистики, а также в сопряженных науках
	Умеет устанавливать межпредметные связи и использовать знания, полученные в других гуманитарных науках, для осмысления, объяснения и интерпретации проблем

	корпусной лингвистики
ПК-24 способность выдвигать гипотезы и последовательно развивать аргументацию в их защиту	Знает основные принципы формирования научной гипотезы и ее аргументации в области корпусной лингвистики
	Умеет формировать и аргументировать гипотезу исследования в области корпусной лингвистики
ПК-25 владение основами современных методов научного исследования, информационной и библиографической культурой	Знает методологию лингвистического исследования в области корпусной лингвистики
	Умеет использовать методологию лингвистического исследования в области корпусной лингвистики
ПК-27 способность оценить качество исследования в данной предметной области, соотнести новую информацию с уже имеющейся, логично и последовательно представить результаты собственного исследования	Знает основные критерии качества научного исследования
	Умеет соотнести материалы собственного исследования с уже имеющейся научной информацией

2. Структура и объем дисциплины

Таблица 1

Вид учебной работы	Всего часов (академические часы)	Часов в семестре (академические часы)
		7
Общий объем зач. ед. час	4	4
	144	144
Из них:		
Часы аудиторной работы (всего):	64	64
Лекции	16	16
Практические занятия	-	-
Лабораторные / практические занятия по подгруппам	48	48
Часы внеаудиторной работы, включая самостоятельную работу обучающегося	80	80
Вид промежуточной аттестации (зачет, диф. зачет, экзамен)	зачет	зачет

3. Система оценивания

Оценивание осуществляется в рамках балльной системы, разработанной преподавателем и доведенной до сведения обучающихся на первом занятии

№	Виды оцениваемой работы	Количество баллов	
		Текущий контроль	Промежуточный контроль

1.	Посещение занятия	0/2	-
2.	Реферат	0-8	-
3.	Устный ответ	0-5	0-20
4.	Тестирование	0-45	0-10

Промежуточная аттестация может быть выставлена по совокупности баллов, полученных обучающимся в рамках текущего контроля.

Перевод баллов в оценки (зачет)

№	Баллы	Оценки
1.	0-60	не зачтено
2.	61-100	зачтен

4. Содержание дисциплины

4.1. Тематический план дисциплины

Таблица 2

№ п/п	Наименование тем и/или разделов	Объем дисциплины (модуля), час.				
		Всего	Виды аудиторной работы (академические часы)			Иные виды контактной работы
			Лекции	Практические занятия	Лабораторные / практические занятия по подгруппам	
1	2	3	4	5	6	7
1	Вводная лекция	4	1	-	-	-
2	Задачи и основные направления корпусной лингвистики	14	1	-	2	-
3.	Предмет исследования	14	2	-	4	-
4.	Типы корпусов. 1	16	2	-	6	-
5.	Типы корпусов. 2	16	2	-	6	-
6.	Лингвистические исследования на базе корпуса. 1	16	2	-	6	-
7.	Лингвистические исследования на базе корпуса. 2	16	2	-	6	-
8.	Методы извлечения информации из корпуса.	14	2	-	4	-
9.	Создание своего корпуса.	20	2	-	6	-
10	Аннотирование корпуса.	14		-	8	-
	Итого (часов)	144	16	-	48	

4.2. Содержание дисциплины (модуля) по темам

4.2.1 Лекционный курс

Вводная лекция	1. Понятие лингвистического корпуса.
----------------	--------------------------------------

	<ol style="list-style-type: none"> 2. История корпусной лингвистики. 3. Корпусный (эмпирический) подход в сравнении с хомскианской лингвистикой
Задачи и основные направления корпусной лингвистики	<ol style="list-style-type: none"> 1. Задачи корпусной лингвистики 2. Направления корпусной лингвистики 3. Межпредметные связи
Предмет исследования	<ol style="list-style-type: none"> 1. Типологизация корпусов 2. Два поколения корпусов 3. Развитие корпусных методик
Типы корпусов. 1	<ol style="list-style-type: none"> 1. Описание устных и письменных корпусов 2. Принципы работы с разными типами корпуса
Типы корпусов. 2	<ol style="list-style-type: none"> 1. Корпусы Аннотированные и неаннотированные 2. Принципы работы с типами корпуса
Лингвистические исследования на базе корпуса. 1	<ol style="list-style-type: none"> 1. Лингводидактические аспекты корпусных методов 2. Стилистические аспекты корпусных методов
Лингвистические исследования на базе корпуса. 2	<ol style="list-style-type: none"> 1. Изучение лексики, грамматики посредством корпуса 2. Изучение лексической нормы/аномалий
Методы извлечения информации из корпуса.	<ol style="list-style-type: none"> 1. Обзор программного обеспечения для корпусных исследований 2. Основные этапы анализа 3. Типология извлекаемой информации
Создание своего корпуса.	<ol style="list-style-type: none"> 1. Подготовка к созданию корпуса 2. Планирование и анализ результатов

Лабораторное занятие № 1

Тема: Задачи и основные направления корпусной лингвистики.

План занятия

1. Задачи корпусной лингвистики
2. компьютерная лингвистика.

Лабораторное занятие № 3,4

Тема: Предмет исследования.

План занятия

1. Развитие лингвистических корпусов в мире.
2. Первое и второе поколение корпусов.
3. Знакомство с корпусами

Лабораторное занятие № 5-7

Тема: Типы корпусов. 1.

План занятия

1. Устные корпуса
2. Письменные корпуса
3. Одноязычные корпуса
4. Многоязычные корпуса

Лабораторное занятие № 8-10

Тема: Типы корпусов. 2.

План занятия

1. Аннотированные корпусы
2. Неаннотированные корпусы
3. Лингвистическая аннотация и метаданные.

Лабораторное занятие № 10-12

Тема: Лингвистические исследования на базе корпуса. 1.

План занятия

1. Изучение лексики.
2. Изучение грамматики
3. Исследование нормы/узуса
4. Социолингвистические исследования

Лабораторное занятие № 13-15

Тема: Лингвистические исследования на базе корпуса. 2.

План занятия

1. Возможности преподавания иностранного языка на основе корпусной лингвистики
2. Стилистические исследования на основе корпусной лингвистики
3. Практические сферы применения: установление плагиата и скрытого цитирования;
4. Практические сферы применения: решение задач лингвистической экспертизы

Лабораторное занятие № 16, 17

Тема: Методы извлечения информации из корпуса.

План занятия

1. Типы извлекаемой информации.
2. Конкорданс.
3. Программы для работы с корпусом: конкордансеры и другие.
4. AntConc, WordSmith, Dialing Concordancer, Corsis.

Лабораторное занятие № 18-20

Тема: Создание своего корпуса.

План занятия

1. Планирование.
2. Сбор и цифровка данных.
3. Кодировка текста.

Лабораторное занятие № 21-24

Тема: Аннотирование корпуса.

1. Хранение корпусов
2. публикация корпусов

3. обновление корпусов.

5. Учебно-методическое обеспечение и планирование самостоятельной работы обучающихся

Таблица 3

№ темы	Темы	Формы СРС, включая требования к подготовке к занятиям
1.	Вводная лекция	Конспектирование первоисточников, изучение научной, учебно-методической литературы
2.	Задачи и основные направления корпусной лингвистики	Доработка конспектов лекций: подбор, изучение, анализ и конспектирование рекомендованной литературы.
3.	Предмет исследования	Самостоятельное изучение отдельных тем учебной программы (или отдельных вопросов, проблем темы) с последующим контролем выполнения задания: предоставлением рефератов, комментариев или в виде устного ответа
4.	Типы корпусов. 1	Работа с литературой по заданной теме, составление вопросов для анализа, опорных конспектов
5.	Типы корпусов. 2	Самостоятельное изучение отдельных тем учебной программы (или отдельных вопросов, проблем темы) с последующим контролем выполнения задания: предоставлением рефератов, комментариев или в виде устного ответа
6.	Лингвистические исследования на базе корпуса. 1	Конспектирование первоисточников, изучение научной, учебно-методической литературы
7.	Лингвистические исследования на базе корпуса. 2	Предоставление реферата по заданной теме
8.	Методы извлечения информации из корпуса.	Подготовка презентаций
9.	Создание своего корпуса.	Подготовка учебного корпуса
10.	Аннотирование корпуса.	Конспектирование первоисточников, изучение научной, учебно-методической литературы

6. Промежуточная аттестация по дисциплине (модулю)

6.1 Оценочные материалы для проведения промежуточной аттестации по дисциплине

Примерные вопросы к зачету:

Вопросы к зачету.

- 1) Количественные методы в языкознании.
- 2) Понятие лингвистического корпуса.
- 3) Основные свойства лингвистических корпусов.
- 4) Репрезентативность и сбалансированность лингвистического корпуса.

- 5) Разметка в лингвистическом корпусе.
- 6) Виды корпусной разметки, метатекстовая и лингвистическая разметка.
- 7) Виды лингвистической разметки.
- 8) Корпус как информационно-поисковая система.
- 9) Критерии классификации корпусов.
- 10) Различные виды корпусов.
- 11) История появления корпусов языков мира.
- 12) Современные корпуса языков мира.
- 13) Национальный корпус русского языка.
- 14) Корпусы европейских языков.
- 15) Корпус английского языка, его структурные, технологические и функциональные особенности.
- 16) Научные исследования на основе корпусов русского и английского языков.

Примерные темы рефератов

1. Метаразметка в Национальном корпусе русского языка. Возможности отбора текстов для поиска лингвистической информации по заранее заданным критериям
2. Проект TEI (Text Encoding Initiative)
3. Рекомендации EAGLES (Expert Advisory Group on Language Engineering Standards)
4. Стандарт CES (Corpus Encoding Standard)
5. Стандарт XCES (Corpus Encoding Standard for XML)
6. Проект ISLE (International Standard for Language Engineering)
7. Стандарт CDIF (Corpus Document Interchange Format)
8. Частеречная разметка (POS-tagging) в НКРЯ
9. Акцентологическая разметка в НКРЯ
10. Синтаксическая разметка в НКРЯ
11. Семантическая разметка в НКРЯ
12. Виды разметки в мультимодальном подкорпусе НКРЯ
13. Проблема выравнивания и разметки текстов в параллельном корпусе
14. Основные трудности частеречной разметки в корпусах английского, русского и других языков.
15. Трудности семантической разметки и варианты решения проблем, принятые в НКРЯ

Примерный тест для текущей аттестации

1. Какие источники данных являются традиционными для лингвистики?
 - а) _____
 - б) _____
 - в) _____
 - г) _____
2. Языковой корпус – это
 - а) электронная коллекция текстов, которые отобраны и обработаны по определенным критериям;
 - б) электронная библиотека художественных текстов;
 - в) набор файлов с текстами разных жанров;
 - г) всё перечисленное выше.

3. С появлением корпусов лингвисты получили возможность исследовать
- а) языковую норму;
 - б) реальное употребление языковых единиц;
 - в) ошибки в речи носителей языка;
 - г) частотность грамматических конструкций.

4. Самый первый корпус содержал
- а) 100 млн. словоупотреблений;
 - б) 1 млн. словоупотреблений;
 - в) 1 млн. текстов;
 - г) 1 млн. предложений.

5. Чтобы выводы, полученные на основе корпусного анализа, могли распространяться на использование языка в определенном языковом сообществе в конкретный период времени, корпус должен быть

- а) размеченным;
- б) репрезентативным;
- в) однородным;
- д) современным.

6. Метаразметка – это
- а) грамматический анализ предложения;
 - б) информация о свойствах словоформ;
 - в) информация о свойствах текста;
 - д) информация о частях речи.

7. Морфологическая и синтаксическая разметка обеспечивает
- а) возможность грамматического анализа предложения;
 - б) автоматический поиск грамматической информации;
 - в) поиск точных форм слов;
 - д) все перечисленное выше.

8. Лемма – это
- а) начальная форма слова;
 - б) одна из возможных словоформ лексемы;
 - в) информация о грамматических свойствах словоформы;
 - д) информация о частеречной принадлежности слова.

9. Символ * в поисковом запросе заменяет
- а) один символ;
 - б) одну морфему;
 - в) любое количество символов в словоформе;
 - д) два символа.

10. Сбалансированность – это
- а) равномерная представленность в корпусе текстов разных жанров;
 - б) наличие в корпусе текстов разных авторов;
 - в) наличие в корпусе текстов одинаковой длины;
 - д) наличие в корпусе параллельных текстов.

6.2 Критерии оценивания компетенций:

Таблица 4

Карта критериев оценивания компетенций

№ п/п	Код и наименование компетенции	Компонент (знаниевый/функциональный)	Оценочные материалы	Критерии оценивания
1	ПК-23 способность использовать понятийный аппарат философии, теоретической и прикладной лингвистики, переводоведения, лингводидактики и теории межкультурной коммуникации для решения профессиональных задач	Знает информацию о широком круге проблем, составляющих в настоящее время предмет изучения в области лингвистики, также сопряженных наук Умеет устанавливать межпредметные связи и использовать знания, полученные в других гуманитарных науках, для осмысления, объяснения и интерпретации проблем корпусной лингвистики	Реферат Тестирование Вопросы к зачету	1. Знает проблемы, входящие в круг лингвистических исследований с применением компьютерных технологий 2. Умеет устанавливать межпредметные связи и использовать полученные в ходе обучения знания 3. Демонстрирует навык применения знаний, полученных в других гуманитарных науках, для осмысления, объяснения и интерпретации проблем корпусной лингвистики
2	ПК-24 способность выдвигать гипотезы и последовательно развивать аргументацию в их защиту	Знает основные принципы формирования научной гипотезы и ее аргументации в области корпусной лингвистики Умеет формировать и аргументировать гипотезу исследования в области корпусной лингвистики	Реферат Тестирование Вопросы к зачету	1. Знает принципы формирования научной гипотезы в области корпусной лингвистики 2. Умеет применять на практике принципы аргументации выдвинутых гипотез исследования 3. Демонстрирует способность к аргументации

				поставленной в исследовании задачи
3	ПК-25 владение основами современных методов научного исследования, информационной и библиографической культурой	Знает методологию лингвистического исследования в области корпусной лингвистики Умеет использовать методологию лингвистического исследования в области корпусной лингвистики	Реферат Тестирование Вопросы к зачету	1. Знает методологию корпусного лингвистического исследования 2. Умеет использовать на практике изученные методы корпусных исследований 3. Демонстрирует навык самостоятельного анализа на основании изученных методов
4	ПК-27 способность оценить качество исследования в данной предметной области, соотнести новую информацию с уже имеющейся, логично и последовательно представить результаты собственного исследования	Знает основные критерии качества научного исследования Умеет соотнести материалы собственного исследования с уже имеющейся научной информацией	Реферат Тестирование Вопросы к зачету	1. Знает главные критерии качества научного исследования 2. Умеет применять данные критерии на практике 3. Демонстрирует способность реализации собственного исследования на основании имеющейся научной информацией

7. Учебно-методическое и информационное обеспечение дисциплины (модуля)

7.1 Основная литература:

1. Захаров В. П. Корпусная лингвистика: учебник для студентов гуманитарных вузов / В. П. Захаров, С. Ю. Богданова. – Иркутск: Иркутский государственный лингвистический университет, 2011. – 161 с. – URL: <http://www.iprbookshop.ru/21088.html>

7.2 Дополнительная литература:

1. Методы когнитивного анализа семантики слова: компьютерно-корпусный подход / В. И. Заботкина, Е. Е. Голубкова, М. А. Кронгауз [и др.] ; под редакцией В. И.

Заботкиной. — 2-е изд. — Москва : Издательский Дом ЯСК, 2019. — 342 с. — URL: <http://www.iprbookshop.ru/97629.html>

2. Рассказы о сновидениях. Корпусное исследование устного русского дискурса / О. А. Савельева-Трофимова, В. Л. Цуканова, А. О. Литвиненко [и др.] ; под редакцией А. А. Кибрик, В. И. Подлеская. — Москва : Языки славянских культур, 2009. — 736 с. — URL: <http://www.iprbookshop.ru/15133.html>

- 7.3. Интернет-ресурсы: <http://www.natcorp.ox.ac.uk/>
<https://ruscorporara.ru/new/>

7.3. Современные профессиональные базы данных и информационные справочные системы

Знаниум - <https://new.znaniium.com/>

Лань - <https://e.lanbook.com/>

IPR Books - <http://www.iprbookshop.ru/>

Elibrary - <https://www.elibrary.ru/>

Национальная электронная библиотека (НЭБ) - <https://rusneb.ru/>

Межвузовская электронная библиотека (МЭБ) - <https://icdlib.nspu.ru/>

"ИВИС" (БД периодических изданий) - <https://dlib.eastview.com/browse>

Электронная библиотека Тюмгу - <https://library.utmn.ru/>

8. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю)

Платформа для электронного обучения Microsoft Teams.

Бесплатное программное обеспечение, установленное в аудиториях: 7-Zip, Adobe Acrobat Reader, Mozilla Firefox, OpenOffice.org.

Лицензионное программное обеспечение, установленное в аудиториях: Microsoft Office 2003, Microsoft Office 2007, Microsoft Office 2010, Windows, Конструктор тестов 2.5 (Keepsoft).

9. Технические средства и материально-техническое обеспечение дисциплины

- Учебные аудитории для проведения лекций и практических занятий укомплектованы специализированной мебелью и техническими средствами обучения, служащими для представления учебной информации большой аудитории: мультимедийные аудитории, укомплектованные таким оборудованием, как проектор, документ камера, проекционный экран.
- Помещения для самостоятельной работы обучающихся (компьютерные классы) оснащены компьютерной техникой с возможностью подключения к сети «Интернет» и обеспечены доступом к электронной информационно-образовательной среде вуза.
- Лаборатории, оснащенные лабораторным оборудованием, для реализации данной дисциплины не предусмотрены.