



**МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ
ФЕДЕРАЦИИ**

федеральное государственное бюджетное образовательное учреждение
высшего образования

**«ИРКУТСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»
ФГБОУ ВО «ИГУ»**

Педагогический институт

Кафедра иностранных языков и лингводидактики



Директор _____ А.В. Семиров

«21» мая 2020 г.

Рабочая программа дисциплины

Б1.В.ДВ.01.01 Корпусная лингвистика

Направление подготовки: 44.03.05 Педагогическое образование (с двумя профилями подготовки)

Направленность подготовки: Иностранный язык (английский) – Иностранный язык (китайский)

Квалификация выпускника: Бакалавр

Форма обучения: Очная

Согласована с УМС ПИ ИГУ

Протокол № 4 от «29» апреля 2020 г.

Председатель _____ М.С. Павлова

Рекомендовано кафедрой:

Протокол № 9 от «21» апреля 2020 г.

Зав. кафедрой _____ М.В. Носкова

Иркутск 2020 г.

I. Цели и задачи дисциплины:

Цель: овладеть современными технологиями поиска и обработки лингвистических данных.

Задачи:

- введение в проблематику корпусной лингвистики, ознакомление с различными направлениями в исследовании языка, основанными на корпусных данных;
- овладение основными методами и приемами построения и использования электронных корпусов текстов;
- формирование у обучающихся практических навыков работы с существующими корпусами.

II. Место дисциплины в структуре ОПОП:

2.1. Учебная дисциплина Б1.В.ДВ.01.01 Корпусная лингвистика относится к Блоку 1 к части, формируемой участниками образовательных отношений.

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

Б1.В.01 Практика устной и письменной речи (английский язык)

Б1.О.19 Профессиональная ИКТ- компетентность педагога

Б1.О.20 Методика обучения и воспитания (уровень общего образования)

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

Б1.О.20 Методика обучения и воспитания (уровень общего образования)

Б1.О.27 Решение практических задач

Б2.О.05(II) Педагогическая практика

III. Требования к результатам освоения дисциплины (модуля):

Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с индикаторами достижения компетенций

Компетенция	Индикаторы компетенций	Результаты обучения
<i>ПК-1 Способен осваивать и использовать базовые научно-теоретические знания и практические умения по предмету иностранный язык в профессиональной деятельности</i>	<i>ИДК ПК1.1 анализирует содержание, сущность, закономерности, принципы и особенности изучаемых явлений и про ессов, базовые теории в иноязычной предметной области; основы теоретических дисциплин в объеме, необходимом для решения педагогических и методических задач</i>	Знать: – основы классификации лингвистических корпусов по разным основаниям; – современные представления о человеческой и человеко-машинной коммуникации; – принципы интерпретации количественных данных. Уметь: – анализировать полученные с помощью корпусных технологий языковые данные; – осуществлять подбор программного обеспечения для решения практических задач в области преподавания иностранного языка.

		<p>Владеть:</p> <ul style="list-style-type: none"> – навыками поиска и получения информации из корпусов текстов; – навыками обращения с системами управления базами данных и знаний различного назначения; – навыками анализа возможностей использования систем управления базами данных и знаний различного назначения.
	<p><i>ИДК ПК1.2 использует базовые предметные научнотеоретические представления о сущности, закономерностях, принципах и особенностях изучаемых языковых явлений и про ессов</i></p>	<p>Знать:</p> <ul style="list-style-type: none"> – основы поиска языковых данных в корпусах современного английского языка, – распределение языковых единиц по регистрам; – принципы использования информации о коллокатах ключевого слова. <p>Уметь:</p> <ul style="list-style-type: none"> – использовать корпусную методологию для определения сочетаемости слов при коммуникации в письменной форме; – осуществлять поиск в устном подкорпусе текстов на английском языке для формирования навыков устной разговорной речи. <p>Владеть:</p> <ul style="list-style-type: none"> – навыками поиска и получения информации о языковых явлениях из корпусов текстов.

IV. СОДЕРЖАНИЕ И СТРУКТУРА ДИСЦИПЛИНЫ
4.1. Объем дисциплины (модуля) и виды учебной работы

Вид учебной работы	Всего часов / зачетных единиц	Семестры			
		7			
Аудиторные занятия (всего)	64	64			
В том числе:			-	-	-
Лекции	32	32			
Практические занятия (ПЗ)	32	32			
Лабораторные работы (ЛР)					
Самостоятельная работа (всего)	44	44			
Вид промежуточной аттестации (зачет)					
Контактная работа (всего)*	64	64			
Общая трудоемкость	часы	108	108		
	зачетные единицы	3	3		

4.2. Содержание учебного материала дисциплины (модуля)

Раздел 1. Введение в корпусную лингвистику

Тема 1. Основные понятия корпусной лингвистики

Определение корпуса и корпусной лингвистики. Основные положения корпусной лингвистики. Цели корпусной лингвистики. Традиционные и современные методы лингвистических исследований. Принятие принципиальных теоретических решений составителями корпусов, зависимость валидности данных от предоставляемой информации.

Тема 2. Типы корпусов

Принцип репрезентативности. Его применение на корпусах разных типов. Первые компьютерные корпуса. Брауновский корпус США и LOB корпус в Европе. Основные типы корпусов. Корпусы английского и русского языков. Классификация корпусов по хронологии, по жанрам, по аннотированности, по целям и задачам, по объему, по языкам.

Раздел 2. Создание корпусов

Тема 3. Технологии корпусной лингвистики

Конструирование корпуса. Предварительная обработка. Токенизация, лемматизация, парсинг. Принципы разметки текстов. Основные процедуры предварительной обработки корпусов.

Раздел 3. Использование корпусов

Тема 4. Сферы применения корпусной лингвистики

Пользователи корпусов. Национальный корпус русского языка, British National Corpus, Longman-Lancaster Corpus, London-Lund Corpus. Международные организации, связанные с корпусной лингвистикой. Лексикографические исследования, основанные на корпусах. Грамматические исследования, основанные на корпусах. Исследования дискурса, основанные на корпусах.

4.3. Разделы и темы дисциплин (модулей) и виды занятий

№ п/п	Наименование раздела	Наименование темы	Типы занятий в часах			
			Лек	Практ. зан.	СРС	Всего
1.	Раздел 1. Введение в корпусную лингвистику	Тема 1. Основные понятия корпусной лингвистики	8	8	10	26
		Тема 2. Типы корпусов	8	8	10	26
2.	Раздел 2. Создание корпусов	Тема 3. Технологии корпусной лингвистики	8	8	14	30
3	Раздел 3. Использование корпусов	Тема 4. Сферы применения корпусной лингвистики	8	8	10	26
	Всего		32	32	44	108

4.4. Методические указания по организации самостоятельной работы студентов

Методические рекомендации и учебно-методические материалы по организации СРС по дисциплине *Б1.В.ДВ.01.01 Корпусная лингвистика* представляют собой комплекс материалов и подходов, позволяющих студенту оптимальным образом организовать процесс изучения данной дисциплины в рамках СРС, а также рекомендации преподавателям по различным аспектам организации СРС.

Выполнение самостоятельной работы предусмотрено во внеаудиторной форме, без участия преподавателя, то есть вне наличия непосредственной обратной связи, корректирующих, контролирующих и оценивающих действий преподавателя. Преподаватель определяет цель, содержание деятельности, сроки выполнения заданий. Самостоятельность студента заключается в индивидуальном подходе при осуществлении заданного преподавателем объема и формата работы, в разработке методов решения поставленных проблем.

По каждой изучаемой теме предусмотрена самостоятельная работа студента, результаты которой напрямую или опосредованно подвергаются проверке и инкорпорированы в требования к уровню освоения дисциплины.

Внеаудиторная самостоятельная работа предполагает не только работу с текущим лекционным материалом и первоисточниками, но и индивидуальную учебно-исследовательскую работу, нацеленную на формирование навыков самостоятельного изучения исследуемой лингвистической проблематики. В этой связи выполнение заданий в ходе самостоятельной работы предусматривает разноаспектный формат работ, который включает в себя работу в научной библиотеке – в том числе чтение научных лингвистических журналов, а также работу с корпусами текстов в Интернете.

В процессе самостоятельной работы студенты выполняют все типы заданий: познавательные, практические, обобщающие, исследовательские.

Специфика дисциплины требует тщательного конспектирования лекций, что существенно облегчит самостоятельную работу студента с рекомендованной литературой при подготовке к практическим занятиям, тестам и зачету.

При освоении дисциплины «Корпусная лингвистика» в процессе самостоятельной подготовки к практическим занятиям студенту необходимо:

- тщательно проработать материал лекций и разделы рекомендованных учебников и научной литературы по соответствующим темам, составить краткий план выступления, уточнить необходимые термины в справочной литературе;
- обратить особое внимание на определения основных научных понятий, формулировки проблем и примеры решения практических задач, приводимых на лекциях и семинарах;
- при подготовке к практическим и семинарским занятиям использовать

- мультимедийные средства;
- следить за выполнением графика самостоятельной работы, не нарушать сроки отчетности.

V. Учебно-методическое и информационное обеспечение дисциплины (модуля):

а) основная литература

1. Захаров, В. П. Корпусная лингвистика: учебник для вузов / В. П. Захаров, С. Ю. Богданова. – Иркутск: ИГЛУ, 2011. – 161 с. (40 экз)

б) дополнительная литература

1. Баранов А.Н. Введение в прикладную лингвистику [Текст]: учебник для вузов / А. Н. Баранов; МГУ. – 3-е изд. – М.: ЛКИ, 2007. – 360 с. (24 экз.)
2. Потапова, Р.К. Новые информационные технологии и лингвистика [Текст] /Р.К. Потапова. – М.: Эдиториал УРСС, 2005. – 368 с. (253 экз)
3. Tognini Bonelli E. Corpus Classroom Currency. – Darbai ir Dienos, 2000. – Pp. 205–243.
<http://donelaitis.vdu.lt/publikacijos/bonelli.pdf>

в) программное обеспечение

Интернет-ресурсы

1. <http://biblioclub.ru>
2. <http://www.ruscorpora.ru>
3. <http://corpus.byu.edu>
4. <http://www.csa.ru/BAN>
5. <http://www.magister.msk.ru/library/>
6. <http://www.eupublishing.com/series/ETOTEL Advanced/Lindquist>
7. <http://www.linguistlist.Org/issues/6/6-74.html>

г) базы данных, информационно-справочные и поисковые системы

- Электронная научная библиотека – <http://elibrary.ru/defaultx.asp>
Научная электронная библиотека «КИБЕРЛЕНИНКА» – <http://cyberleninka.ru/>
Научная электронная библиотека ГПНТБ России – <http://ellib.gpntb.ru/>
Электронная библиотека – <http://elibrary.rsl.ru/>
Электронная библиотека – <http://www.aspirantura.ru/bibl.php>

VI. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

6.1. Помещения и оборудование

В институте имеются в наличии:

Специальные помещения:

- учебная аудитория для проведения занятий лекционного типа с набором демонстрационного оборудования;
- учебная аудитория для проведения занятий семинарского типа, для групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации;
- помещение для самостоятельной работы, оснащенное компьютерной техникой, подключенной к сети Интернет, обеспечивающей доступ в электронно-информационную образовательную среду организации – образовательный портал ИГУ
<http://www.educa.isu.ru>.

Оборудование:

Экран – 1, компьютеров Formoza AMD SDA 3100 – 12, мультимедийный переносной проектор BenQ MPG 11C DLP.

6.2. Лицензионное и программное обеспечение

ОС: Windows XP (Номер Лицензии Microsoft 19683056), Антивирус Kaspersky (Форус Контракт № 04-114-16 от 14 ноября 2016 г. KES Счет № РСЦЗ-000147 и АКТ от 23 ноября 2016 г. Лиц. № 1В08161103014721370444).

VII.

ОБРАЗОВАТЕЛЬНЫЕ ТЕХНОЛОГИИ

В образовательном процессе используются активные и интерактивные формы проведения занятий (компьютерные симуляции, деловые и ролевые игры, разбор конкретных ситуаций, психологические и иные тренинги, групповые дискуссии), развивающие у обучающихся навыков командной работы, межличностной коммуникации, принятия решений, лидерских качеств и формирующие компетенции.

Наименование тем занятий с использованием активных форм обучения:

Тема занятия	Вид занятия	Форма / Методы интерактивного обучения	Кол-во часов
Тема 1. Основные понятия корпусной лингвистики	Лекция, практическое занятие	- лекция-беседа; - интеллектуальные разминки;	12
Тема 3. Технологии корпусной лингвистики	Лекция, практическое занятие	- презентация докладов с использованием мультимедийных средств; - работа в малых группах;	12
Тема 4. Сферы применения корпусной лингвистики	Лекция, практическое занятие	- обучающие игры (ролевые, имитационные, деловые, образовательные и др.)	12
Итого часов			36

VIII. ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ДЛЯ ТЕКУЩЕГО КОНТРОЛЯ УСПЕВАЕМОСТИ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ

8.1. Оценочные средства для проведения текущего контроля успеваемости

Устный опрос

Вопросы:

- 1) Дайте определения терминов: корпус, корпусная лингвистика, разметка, репрезентативность
- 2) Перечислите типы корпусов.
- 3) Какой корпус текстов был первым? Укажите его основные характеристики.
- 4) Какие два типа корпусов можно выделить по критерию репрезентативности? В чем их отличие?
- 5) Каково основное назначение параллельных корпусов текстов?
- 6) В чем заключается сложность создания корпусов устной речи?

Демонстрационный вариант Теста 1

1. Перечислите технологии в разных отраслях лингвистики, повлиявшие на развитие современной корпусной лингвистики
2. Как осуществляется предварительная обработка текстов при создании корпуса?
3. В чем заключаются сложности составления корпусов устной речи?
4. Как можно различить синонимичные слова с помощью корпусов текстов?

5. Что подразумевается под «репрезентативностью корпуса»?
6. Дайте определения токенизации, лемматизации, парсинга.

Демонстрационный вариант Теста 2

Выберите правильный ответ из правой колонки

1.	Какое определение корпуса является наиболее точным?	<p>а) информационно-справочная система, основанная на собрании текстов на некотором языке в электронной форме.</p> <p>б) репрезентативное собрание текстов, обычно в читаемом машиной формате и включающем информацию о ситуации, в которой текст был произведен, такую как информация о говорящем, авторе, адресате или аудитории.</p> <p>в) своеобразного рода словесное единство, расположенное на машинном носителе, стандартизовано представляющее словесный материал, позволяющее применять стандартные программы его обработки, характеризующееся конечным размером и репрезентативностью как результатом особой процедуры отбора.</p>
2.	Какое определение корпуса является наиболее точным?	<p>а) информационно-справочная система, основанная на собрании текстов на некотором языке в электронной форме.</p> <p>б) репрезентативное собрание текстов, обычно в читаемом машиной формате и включающем информацию о ситуации, в которой текст был произведен, такую как информация о говорящем, авторе, адресате или аудитории.</p> <p>в) своеобразного рода словесное единство, расположенное на машинном носителе, стандартизовано представляющее словесный материал, позволяющее применять стандартные программы его обработки, характеризующееся конечным размером и репрезентативностью как результатом особой процедуры отбора.</p>
3	Что представляет собой корпусная лингвистика?	<p>а) отрасль компьютерной лингвистики, занимающаяся составлением корпусов текстов для лексикографических целей</p> <p>б) деятельность, требующаяся для составления и использования корпуса, направленная на исследование естественного использования языка</p> <p>в) вид деятельности, связанной с использованием текстов для того, чтобы ответить на вопросы о использовании языка</p>
4.	Какой корпус содержит только устную речь?	<p>а) London-Lund Corpus</p> <p>б) Longman-Lancaster Corpus</p> <p>в) British National Corpus</p>
5.	Какой минимальный объем считается репрезентативным для	<p>а) 160 млн. словоупотреблений</p> <p>б) 1 млн. словоупотреблений</p>

	национального корпуса языка?	в) 100 млн. словоупотреблений
6.	Что обозначает термин «коллокация» в корпусной лингвистике?	а) любое сочетание слов б) устойчивое сочетание слов в) сочетание слов, расположенных рядом в тексте
7.	Области лингвистических исследований, повлиявшие на развитие современной корпусной лингвистики	а) историческая лингвистика, грамматика, социолингвистика б) когнитивная лингвистика, лексикография, социолингвистика в) фонология, лексикография, лингводидактика
8.	Какие данные можно получить из корпусов текстов?	а) суждения о грамматике, экстралингвистическая информация б) экспериментальные данные, данные полевых исследований в) эмпирическая поддержка, информация по частотности, метаинформация (экстралингвистическая информация)
9.	Что такое конкорданс?	а) указатель, связывающий каждое употребление с контекстом б) список встречающихся в тексте словоформ, расположенных в алфавитном порядке в) тип словаря
10	Что такое token?	а) словоформа б) начальная форма слова в) основа слова
11	Какой процесс устанавливает лексему, парадигма которой включает данную словоформу?	а) стемминг б) токенизация в) лемматизация
12	Для чего нужны нормированные подсчеты?	а) они указывают частотность встречаемости определенной словоформы б) они преобразуют количество случаев встречаемости слова по стандартной шкале в) они применяются в лексикографии для выявления новых значений слов
13	Чтобы слова можно было ассоциировать друг с другом, они должны:	а) примыкать друг к другу или располагаться на некотором расстоянии б) располагаться на некотором расстоянии в) примыкать друг к другу
14	В каком из регистров чаще встречаются номинализации с суффиксом –ness?	а) в академической прозе б) в художественной литературе в) в устной речи
15	Для каких исследований можно использовать неаннотированные корпуса?	а) в области морфологии б) в области грамматических категорий в) в области синтаксиса
16	Какой из регистров демонстрирует предпочтение к статической, а не к динамической упаковке информации,	а) устная речь б) художественная литература в) академическая проза

	отраженной в большей частотности номинализаций и более высоком соотношении существительное-глагол?	
--	--	--

8.2. Оценочные средства для промежуточной аттестации

1. Вид промежуточной аттестации - зачет.
2. Форма проведения - устный ответ.

Примерный перечень вопросов к зачету:

1. Определения корпусной лингвистики.
2. Основные понятия корпусной лингвистики: корпус текстов, разметка, конкорданс, токен, лемма, коллокация.
3. Традиционные и современные методы лингвистических исследований корпусов.
4. Зависимость валидности данных от предоставляемой информации.
5. Принцип репрезентативности, его применение на корпусах разных типов.
6. Первые компьютерные корпуса.
7. Брауновский корпус США.
8. LOB корпус в Европе.
9. Основные типы корпусов.
10. Корпусы английского и русского языков
11. Классификация корпусов по хронологии, по жанрам, по аннотированности, по целям и задачам, по объему, по языкам.
12. Технологии в разных отраслях лингвистики, повлиявшие на развитие современной корпусной лингвистики.
13. Конструирование корпуса.
14. Основные процедуры предварительной обработки корпусов. Токенизация, лемматизация, парсинг.
15. Принципы разметки текстов.
16. Конкурирующие технологии (rule-based), статистические, основанные на ограничениях (restriction-based), гибридные).
17. Национальный корпус русского языка.
18. British National Corpus.
19. Longman-Lancaster Corpus, London-Lund Corpus.
20. Международные организации, связанные с корпусной лингвистикой.

Автор: д.филол.н., профессор Богданова С.Ю.

Документ составлен в соответствии с требованиями ФГОС по направлению 44.03.05 Педагогическое образование (с двумя профилями подготовки) утвержденного приказом Минобрнауки РФ об утверждении ФГОС ВО № 125 от 22 февраля 2018 г.

Настоящая программа не может быть воспроизведена ни в какой форме без предварительного письменного разрешения кафедры-разработчика программы.