

ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ УЧРЕЖДЕНИЕ НАУКИ
КАЛМЫЦКИЙ ИНСТИТУТ ГУМАНИТАРНЫХ ИССЛЕДОВАНИЙ
РОССИЙСКОЙ АКАДЕМИИ НАУК
(КИГИ РАН)



РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ
Б.1.В.ДВ.1.1 КОРПУСНАЯ ЛИНГВИСТИКА

Направленность подготовки:
45.06.01 Языкознание и литературоведение

Профиль: Языки народов зарубежных стран Европы,
Азии, Африки, аборигенов Америки и Австралии (монгольские языки)

Квалификация (степень) выпускника:
Исследователь. Преподаватель-исследователь

Курс обучения: 3
Форма обучения: очная

Составители: Куканова В.В., кандидат филологических наук, старший научный сотрудник отдела языкознания; Бембеев Е.В., кандидат филологических наук, старший научный сотрудник отдела языкознания.

Рабочая программа дисциплины «Корпусная лингвистика» основной профессиональной образовательной программы высшего образования, реализуемой в КИГИ РАН по направлению подготовки 45.06.01 «Языкознание и литературоведение», направленность (профиль) – «Языки народов зарубежных стран Европы, Азии, Африки, аборигенов Америки и Австралии (монгольские языки)» разработана в соответствии с Федеральным государственным образовательным стандартом высшего образования по направлению подготовки 45.06.01 «Языкознание и литературоведение», утвержденным Приказом Министерства образования и науки Российской Федерации от «30» июля 2014 г. № 903 (с изменениями и дополнениями от 30 апреля 2015 г.). Рабочая программа утверждена на расширенном заседании отдела языкознания «11» января 2016 г., протокол № 1, и соответствует учебному плану.

Аннотация к рабочей программе по дисциплине «Корпусная лингвистика»

Данная дисциплина относится к дисциплинам учебного плана программы аспирантуры по направлению подготовки 45.06.01 «Языкознание и литературоведение», профиль – «Языки народов зарубежных стран Европы, Азии, Африки, аборигенов Америки и Австралии (монгольские языки)». Дисциплина «Корпусная лингвистика» является дисциплиной по выбору вариативной части дисциплин (Б1.В.ДВ.1.1) согласно учебному плану ОПОП ВО.

Квалификация выпускника, освоившего программу аспирантуры, – Исследователь. Преподаватель-исследователь.

1. Компетенции, формируемые в результате освоения дисциплины (в соответствии с ФГОС ВО и учебным планом):

ОПК-1 – способность самостоятельно осуществлять научно-исследовательскую деятельность в соответствующей профессиональной области с использованием современных методов исследования и информационно-коммуникационных технологий.

УК-1 – способность к критическому анализу и оценке современных научных достижений, генерированию новых идей при решении исследовательских и практических задач, в том числе в междисциплинарных областях.

УК-4 – готовность использовать современные методы и технологии научной коммуникации на государственном и иностранном языках.

ПК-1 – способность к самостоятельному проведению научно-исследовательской работы и получению научных результатов, удовлетворяющих установленным требованиям к содержанию диссертаций на соискание ученой степени кандидата наук по направленности (научной специальности).

ПК-3 – владение базовыми навыками сбора и анализа различных типов информации с использованием традиционных методов и современных информационных технологий.

2. Содержание дисциплины

Тема 1. Вводная лекция. Понятие лингвистического корпуса (2 ч)

Тема 2. Цели и основные направления корпусной лингвистики. Соотношение корпусной лингвистики и компьютерной лингвистики (2 ч)

Тема 3. Предмет исследования. Общая характеристика лингвистических корпусов. Машинные фонды, базы данных, корпус (2 ч)

Тема 4. Виды корпусов: устные и письменные, одноязычные и многоязычные (2 ч)

Тема 5. Разметка. Виды разметок. Лингвистическая аннотация и метаданные (2 ч)

Тема 6. Методы извлечения информации из корпуса (2 ч)

Тема 7. Создание формализованных описаний для разработки корпуса. Словарный и бессловарный способ (2 ч)

Тема 8. Создание своего корпуса. Планирование. Сбор и оцифровка данных. Кодировка текста (2 ч)

Тема 9. Аннотирование корпуса. Хранение, публикация и обновление корпусов (2 ч)

Тема 10. Лингвистические исследования на базе корпуса: изучение лексики (2 ч)

Тема 11. Лингвистические исследования на базе корпуса: составление словарей (2 ч)

Тема 12. Лингвистические исследования на базе корпуса: изучение грамматики (2ч)

Тема 13. Обзор корпусных систем по монгольским языкам (2 ч)

Тема 14. Актуальные проблемы корпусной лингвистики монгольских языков (2 ч)

3. Общая трудоемкость дисциплины составляет 4 зачетные единицы, 144 ч. Из них общий объем аудиторных часов 28, в том числе лекций – 28 ч.

4. Формы текущего контроля и промежуточной аттестации

Формы текущего контроля – дискуссия.

Форма промежуточной аттестации – зачет.

1. Цели и задачи дисциплины

Данный курс направлен на формирование у аспирантов совокупности универсальных, профессиональных и общепрофессиональных компетенций, составляющих необходимую основу для изучения основных проблем современной филологии и лингвистики.

1.1. Цель дисциплины: ознакомить с основами корпусной лингвистики и сформировать навыки и умения работы в корпусных системах в рамках эмпирического подхода к изучению языка. Кроме того, познакомить аспиранта с основами разработки информационно-справочных систем, актуальными проблемами в корпусной лингвистике монгольских языков.

1.2. Виды и задачи профессиональной деятельности по дисциплине:

- ознакомить аспирантов с предметами, задачами, методами и понятийным (терминологическим) аппаратом корпусной лингвистики;
- освоить материал по основам корпусной лингвистики, различными корпусными системами;
- сформировать знания о компьютерном инструментарии корпусной лингвистики;
- дать аспирантам первичное представление о возможных применениях корпусов в лингвистических исследованиях, в частности в монгольском языкознании.

2. Место дисциплины в структуре ОПОП ВО

Данная дисциплина относится к дисциплинам учебного плана программы аспирантуры по направлению подготовки 45.06.01 «Языкознание и литературоведение», профиль – «Языки народов зарубежных стран Европы, Азии, Африки, аборигенов Америки и Австралии (монгольские языки)». Дисциплина «Корпусная лингвистика» является дисциплиной по выбору вариативной части дисциплин (Б1.В.ДВ.1.1) согласно учебному плану ОПОП ВО.

Дисциплина «Корпусная лингвистика» изучается в 5-м семестре 3-го года обучения.

Компетенции обучающегося, формируемые в результате освоения дисциплины. Процесс изучения дисциплины «Корпусная лингвистика» направлен на формирование следующих компетенций:

ОПК-1 – способность самостоятельно осуществлять научно-исследовательскую деятельность в соответствующей профессиональной области с использованием современных методов исследования и информационно-коммуникационных технологий.

УК-1 – способность к критическому анализу и оценке современных научных достижений, генерированию новых идей при решении исследовательских и практических задач, в том числе в междисциплинарных областях.

УК-4 – готовность использовать современные методы и технологии научной коммуникации на государственном и иностранном языках.

ПК-1 – способность к самостоятельному проведению научно-исследовательской работы и получению научных результатов, удовлетворяющих установленным требованиям к содержанию диссертаций на соискание ученой степени кандидата наук по направленности (научной специальности).

ПК-3 – владение базовыми навыками сбора и анализа различных типов информации с использованием традиционных методов и современных информационных технологий.

В результате освоения дисциплины аспирант осваивает следующие компетенции:

Компетенция	Код компетенции по ФГОС	Структура компетенции	Формы и методы обучения, способствующие формированию и
-------------	-------------------------	-----------------------	--

			развитию компетенции
Способность к критическому анализу и оценке современных научных достижений, генерированию новых идей при решении исследовательских и практических задач, в том числе в области корпусной лингвистики	УК-1	<p>Знать: основные методы научно-исследовательской деятельности.</p> <p>Уметь: выделять и систематизировать основные идеи в научных текстах; критически оценивать любую поступающую информацию, вне зависимости от источника; избегать автоматического применения стандартных формул и приемов при решении задач.</p> <p>Владеть: навыками сбора, обработки, анализа и систематизации информации по теме исследования; навыками выбора методов и средств решения задач исследования.</p>	Лекции, индивидуальные консультации, самостоятельная работа с привлечением Интернет-ресурсов
Готовность использовать современные методы и технологии научной коммуникации на государственном и иностранном языках.	УК-4	<p>Знать: виды и особенности письменных текстов и устных выступлений; понимать общее содержание сложных текстов на абстрактные и конкретные темы, в том числе узкоспециальные тексты.</p> <p>Уметь: подбирать литературу по теме, составлять двуязычный словарь, переводить и реферировать специальную литературу, подготавливать научные доклады и презентации на базе прочитанной специальной литературы, объяснить свою точку зрения и рассказать о своих планах.</p> <p>Владеть: навыками анализа основных мировоззренческих и методологических проблем, в т. ч. междисциплинарного характера возникающих в науке на современном этапе ее развития, владеть технологиями планирования</p>	Лекции, индивидуальные консультации, самостоятельная работа с привлечением Интернет-ресурсов

		профессиональной деятельности в сфере научных исследований	
Способность самостоятельно осуществлять научно-исследовательскую деятельность в соответствующей профессиональной области с использованием современных методов исследования и информационно-коммуникационных технологий	ОПК-1	<p>Знать:</p> <p>цели и задачи научных исследований по направлению деятельности, базовые принципы и методы их организации; основные источники научной информации и требования к представлению информационных материалов.</p> <p>Уметь:</p> <p>составлять общий план работы по заданной теме, предлагать методы исследования и способы обработки результатов, проводить исследования по согласованному с руководителем плану, представлять полученные результаты</p> <p>Владеть:</p> <p>систематическими знаниями по направлению деятельности; углубленными знаниями по выбранной направленности подготовки, базовыми навыками проведения научно-исследовательских работ по предложенной теме.</p>	Лекции; индивидуальные консультации; самостоятельная работа по составлению текстов рефератов (докладов, лекций), подготовки семинарских занятий по заданной тематике с использованием Интернет-ресурсов
Углубленное знание современных проблем филологии, умение предлагать и аргументированно обосновывать способы их решения	ПК-1	<p>Знать:</p> <p>фундаментальные основы науки о теории языка и специальных дисциплин.</p> <p>Уметь:</p> <p>аргументировать и составлять план работы по заданной теме, обосновывать получаемые результаты; умеет работать с основными информационно-поисковыми системами, системами представления знаний, синтаксического и морфологического анализа, автоматического синтеза и распознавания речи, обработки лексикографической информации и автоматизированного перевода;</p>	Лекции; индивидуальные консультации; самостоятельная работа по составлению текстов рефератов (докладов, лекций), подготовки семинарских занятий по заданной тематике с использованием Интернет-ресурсов

		<p>Владеть: глубокими знаниями в избранной конкретной области филологии; владеет методами формального и когнитивного моделирования естественного языка и методами создания метаязыков; владеет основными математико-статистическими методами обработки лингвистической информации с учетом элементов программирования и автоматической обработки лингвистических корпусов.</p>	
<p>Системное владение методами научного исследования, способность формулировать новые цели и достигать новых результатов в соответствующей предметной области.</p>	ПК-3	<p>Знать: современные методики и методологии научных исследований, результаты научных исследований в области филологии в России и за рубежом. Уметь: формулировать новые цели и связанные с целью задачи, анализировать, комментировать, реферировать и обобщать результаты научных исследований, проведенных другими специалистами, с использованием современных методик и методологий, передового отечественного и зарубежного опыта. Владеть: системой методов для достижения новых результатов в избранной конкретной области филологии</p>	<p>Лекции; индивидуальные консультации; самостоятельная работа с привлечением Интернет-ресурсов</p>

3. Объем дисциплины и виды учебной работы

3.1. График изучения дисциплины

Общая трудоемкость дисциплины составляет 4 зачетные единицы, 144 ч. Из них общий объем аудиторных часов – 28 ч, в том числе лекций – 28 ч. Самостоятельная работа – 116 ч.

Вид учебной работы	Часы						
	Всего час/зет	Год/семестр					
		1 год		2 год		3 год	
		1	2	3	4	5	6

Общая трудоемкость дисциплины	144/ 4 ЗЕ					144/ 4 ЗЕ	
Аудиторные занятия	28					28	
Лекции (Л)						28	
Практические занятия (ПЗ)							
Самостоятельная работа	116					116	
Зачет						зачет	

3.2. Содержание лекций

№ темы	Название раздела/темы	Технология проведения	Формируемые компетенции (код)	Форма оценочных средств	Трудоемкость (час.)
1	Вводная лекция. Понятие лингвистического корпуса	Организационно ориентированная лекция	ПК-1 ПК-3 УК-1	дискуссия	2
2	Цели и основные направления корпусной лингвистики. Соотношение корпусной лингвистики и компьютерной лингвистики	Информационно-обзорная лекция, проблемная лекция	ПК-1 ПК-3 УК-1	дискуссия	2
3	Предмет исследования. Общая характеристика лингвистических корпусов. Машинные фонды, базы данных, корпус	Проблемная лекция	УК-1 ПК-1 ПК-3	дискуссия	2
4	Виды корпусов: устные и письменные, одноязычные и многоязычные	Информационно-обзорная лекция	ПК-1 ПК-3 УК-1	дискуссия	2
5	Разметка. Виды разметок. Лингвистическая аннотация и метаданные	Информационно-обзорная лекция	УК-1 ПК-1 ПК-3 ОПК-1	дискуссия	2
6	Методы извлечения информации из корпуса	Информационно-обзорная лекция	ПК-1 ПК-3 УК-1 УК-4	дискуссия	2
7	Создание формализованных	Проблемная лекция	УК-1 УК-4	дискуссия	2

	описаний для разработки корпуса. Словарный и бессловарный способ		ПК-1 ПК-3 ОПК-1		
8	Создание своего корпуса. Планирование. Сбор и оцифровка данных. Кодировка текста	Лекция-диалог	УК-1 УК-4 ПК-1 ПК-3 ОПК-1	дискуссия	2
9	Аннотирование корпуса. Хранение, публикация и обновление корпусов	Проблемная лекция	УК-1 УК-4 ПК-1 ПК-3	дискуссия	2
10	Лингвистические исследования на базе корпуса: изучение лексики	Лекция-диалог	УК-1 УК-4 ПК-1 ПК-3 ОПК-1	дискуссия	2
11	Лингвистические исследования на базе корпуса: составление словарей	Лекция-диалог	УК-1 УК-4 ПК-1 ПК-3 ОПК-1	дискуссия	2
12	Лингвистические исследования на базе корпус: изучение грамматики	Лекция-диалог	УК-1 УК-4 ПК-1 ПК-3 ОПК-1	дискуссия	2
13	Обзор корпусных систем по монгольским языкам	Информационно- обзорная лекция	УК-1 УК-4 ПК-1 ПК-3 ОПК-1	дискуссия	2
14	Актуальные проблемы корпусной лингвистики монгольских языков	Проблемная лекция	УК-1 УК-4 ПК-1 ПК-3 ОПК-1	дискуссия	2

Тема 1. Вводная лекция. Понятие лингвистического корпуса. Определение лингвистического (языкового) корпуса. Лингвистические направления, предвосхитившие появление корпусной лингвистики: историческая лингвистика, написание грамматик, лексикография и обучение языку, социолингвистика, формализм в языке. Сопоставление корпусной и традиционной лингвистики. Корпусный подход и его специфика.

Тема 2. Цели и основные направления корпусной лингвистики. Соотношение корпусной лингвистики и компьютерной лингвистики. Становление корпусной лингвистики как раздела компьютерной лингвистики. Ее цели, задачи, предмет и объект исследования. Основные направления корпусной лингвистики: лексикографические исследования, квантитативная лингвистика, машинный перевод, лингводидактика, лингвистика текста и др.

Тема 3. Предмет исследования. Общая характеристика лингвистических корпусов. Машинные фонды, базы данных, корпус. Основные характеристики корпусов. Понятие репрезентативности корпусов. Понятие сбалансированности корпусов. Машинные фонды, базы данных, корпус. Общие черты и отличия. Предпосылки создания корпусов на основе машинных фондов, баз данных. Машинный фонд vs Корпус. База данных vs корпус.

Тема 4. Виды корпусов: устные и письменные, одноязычные и многоязычные. Виды корпусов: исследовательский, иллюстративный, статический, динамический, устный, письменный, одноязычный, многоязычный и др. Специфика каждого корпуса. Демонстрация разных типов корпусов.

Тема 5. Разметка. Виды разметок. Лингвистическая аннотация и метаданные. Понятие разметки. Программы автоматической разметки корпусов. Лингвистическая разметка: морфологическая, синтаксическая, семантическая, анафорическая, просодическая, дискурсная. Понятие метаданных. Экстралингвистическая разметка: библиографическая, типологическая, тематическая, социологическая, формальная структурная, технико-технологическая. Глубина разметки.

Тема 6. Методы извлечения информации из корпуса. Языки запросов. Выходные интерфейсы. Типы извлекаемой информации. Автоматизированный поиск информации по жанровым разновидностям текстов, представленных в корпусе языковых данных. Конкорданс, коллокация, кластеры, чанки, коллигация.

Тема 7. Создание формализованных описаний для разработки корпуса. Словарный и бессловарный способ. Языковой корпус как компьютерная база данных. Процессы токенизации, лемматизации, стемминга и парсинга: специфика, цели, результаты. Корпусные менеджеры.

Тема 8. Создание своего корпуса. Планирование. Сбор и оцифровка данных. Кодировка текста. Техника и алгоритм создания корпуса письменных текстов. Техника и алгоритм создания корпуса устных текстов. Техники бутстрэпинга. Проблема авторских прав. Стандартизация. Программы для оцифровки текстов. Расшифровка устной речи. Программы для расшифровки.

Тема 9. Аннотирование корпуса. Хранение, публикация и обновление корпусов. Единица хранения корпуса данных. Формат представления информации. Он-лайн корпуса. Офф-лайн корпуса. Программы аннотирования корпусов. Проблема обновления корпусов.

Тема 10. Лингвистические исследования на базе корпуса: изучение лексики. Частотный словарь. Структурно-вероятностная модель текста. Функции частотного словаря. Единица описания. Понятие токена, леммы, словоформы. Принципы составления частотных словарей. Лингвистические генеральная и выборочная совокупность. Словник. Абсолютная частота. Ранг. I_{pm} . Коэффициент D . Коэффициент R . Покрываемость текста. Лексическое разнообразие текста. Исследование нормы и узуса. Примеры частотных словарей.

Тема 11. Лингвистические исследования на базе корпуса: составление словарей. Краткая характеристика основ идеографического описания языков. Понятие «семантической сети» как одного из способов отражения семантики предметной области и как промежуточной формы представления текстовых данных. Классификации семантических сетей и их критика. Типы отношений в семантических сетях. Тезаурус как особая разновидность словарей. Перечисление значений слов в соответствии с их частотным

распределением в корпусе текстов как один из базовых принципов построения wordnet-тезауруса.

Тема 12. Лингвистические исследования на базе корпуса: изучение грамматики. Исследование нормы и узуса. Исследование морфологии и синтаксиса. Граммема. Частотность грамем. Частотность частей речи. Частотность синтаксических молекул. Структурно-вероятностная модель грамматики.

Тема 13. Обзор корпусных систем по монгольским языкам. История создания и разработки корпусных систем по монгольским языкам. Генеральный корпус монгольского языка. Калмыцкий корпус. Национальный корпус калмыцкого языка. Бурятский корпус.

Тема 14. Актуальные проблемы корпусной лингвистики монгольских языков. Исследования в области грамматики и лексики монгольских языков. Проблемы создания корпуса устной речи монгольских языков. Отсутствие корпусных разработок по некоторым языкам монгольской группы. Создание новых типов разметки для существующих корпусов. Проблема глубины разметки. Проблема омонимии в разработанных корпусных системах.

3.3. Практические (семинарские) занятия не предусмотрены учебным планом

4. Учебно-методическое обеспечение самостоятельной работы аспирантов

4.1. Формы проведения самостоятельной работы

Тема дисциплины	Форма самостоятельной работы	Трудоемкость в часах
Вводная лекция. Понятие лингвистического о корпуса	Выполнение домашней работы. 1. Проанализировать определения корпуса в отечественной и зарубежной корпусной лингвистике. Выявить общие и различные черты. 2. Подготовить реферат об истории развития корпусной лингвистики в России. 3. Подготовить реферат об истории развития корпусной лингвистике за рубежом.	9
Цели и основные направления корпусной лингвистики. Соотношение корпусной лингвистики и компьютерной лингвистики	Выполнение домашней работы. 1) Подготовить письменные ответы на следующие вопросы: 1. Каковы основные цели создания корпусных систем? 2. Каким образом соотносится корпусная и компьютерная лингвистика? 2) Подготовить письменное сообщение об основных направлениях развития корпусной лингвистики.	9
Предмет исследования. Общая характеристика лингвистических корпусов. Машинные фонды, базы	Выполнение домашней работы. Подготовить письменные ответы на следующие вопросы: 1. Охарактеризуйте предмет исследования в корпусной лингвистике. 2. Чем отличаются машинные фонды, базы данных и корпуса? Подготовить индивидуальные сообщения о	9

данных, корпус	функционирующих корпусных системах.	
Виды корпусов: устные и письменные, одноязычные и многоязычные	Выполнение домашней работы. Подготовить письменные ответы на следующие вопросы: 1. Какие существуют виды корпуса? 2. Какие преимущества и недостатки одноязычных и многоязычных корпусных систем?	11
Разметка. Виды разметок. Лингвистическая аннотация и метаданные	Выполнение домашней работы. Подготовить письменные ответы на следующие вопросы: 1. Что такое разметка? 2. Какие существуют разметки? 3. Какой язык используется при создании разметок? 4. Чем отличаются лингвистическая аннотация и метаданные? 5. Для чего нужна разметка? Придумайте свою систему разметки.	9
Методы извлечения информации из корпуса	Выполнение домашней работы. Подготовить письменные ответы на следующие вопросы: 1. Какие существуют методы извлечения информации из корпуса и типы информации? 2. Какие существуют типы поиска? 3. Какие программы вы знаете для создания конкорданса? 4. Дайте определения терминам <i>конкорданс</i> , <i>коллокация</i> , <i>кластеры</i> , <i>коллигация</i> . Извлеките из Национального корпуса русского языка примеры употребления, коллокации слова <i>корпус</i> . Дайте частотную характеристику терминов <i>я</i> , <i>ты</i> , <i>мы</i> , <i>вы</i> , <i>он</i> , <i>она</i> , <i>оно</i> , <i>они</i> .	8
Создание формализованных описаний для разработки корпуса. Словарный и бессловарный способ	Выполнение домашней работы. Подготовить письменные ответы на следующие вопросы: 1. Дайте характеристику словарному и бессловарному способу создания корпуса. 2. Что понимается под формализованным языком? 3. Формальные языки и естественные языки. 4. Каковы преимущества словарного и бессловарного подхода в создании корпусной системы?	7
Создание своего корпуса. Планирование. Сбор и оцифровка данных. Кодировка текста	Выполнение домашней работы. Подготовить письменные ответы на следующие вопросы: 1. Каковы принципы создания корпусных систем? 2. Какие существуют кодировки символов? 3. Опишите преимущества того или иного вида кодировки. Проведите сбор и оцифровку для своего корпуса.	7
Аннотирование	Выполнение домашней работы.	7

корпуса. Хранение, публикация и обновление корпусов	Подготовить письменные ответы на следующие вопросы: 1. Опишите преимущества ручного и автоматического аннотирования корпуса. 2. Опишите существующие корпусные менеджеры. 3. Опишите преимущества он-лайн и оф-лайн корпуса.	
Лингвистические исследования на базе корпуса: изучение лексики	Выполнение домашней работы. Подготовить письменные ответы на следующие вопросы: 1. Как можно использовать частотные характеристики в лингвистических исследованиях? 2. Какие существуют подходы для изучения сочетаемости слов?	9
Лингвистические исследования на базе корпуса: составление словарей	Выполнение домашней работы. Подготовить письменные ответы на следующие вопросы: 1. Опишите возможности использования корпуса для создания словарей. 2. Какие словари можно создать на основе корпусных исследований? Подготовить словарные статьи на слова <i>эк, эцк, ах, эгч.</i>	7
Лингвистические исследования на базе корпуса: изучение грамматики	Выполнение домашней работы. Подготовить письменные ответы на следующие вопросы: 1. Что может дать исследователю грамматики контекстный поиск в корпусе? 2. Какие работы проводились в проблеме исследования частей речи русского языка в российской корпусной лингвистике? 3. Какие работы проводились в проблеме исследования синтаксических процессов на уровне предложения в российской корпусной лингвистике?	7
Обзор корпусных систем по монгольским языкам	Выполнение домашней работы. Подготовить письменные ответы на следующие вопросы: 1. Каковы способы и инструментарий создания перечисленных корпусных систем? 2. Сделайте сообщение по каждому из корпусов (история и современное состояние)	10
Актуальные проблемы корпусной лингвистики монгольских языков	Выполнение домашней работы. Подготовить письменные ответы на следующие вопросы: 1. Каковы актуальные проблемы корпусной лингвистики монгольских языков? 2. Какие Вам видятся перспективы развития корпусной лингвистики по монгольским языкам?	7
Итого:		116

5. Контроль знаний аспирантов

5.1. Формы текущего контроля работы аспирантов
Дискуссия.

5.2. Промежуточная аттестация по дисциплине
Промежуточная аттестация проводится в форме зачета.

5.3. Вопросы по дисциплине «Корпусная лингвистика» для зачета:

1. Корпус. База данных. Машинный фонд
2. История лингвистических корпусов: от картотеки к корпусу
3. Классификация (типология) корпусов
4. Корпусная лингвистика: современное состояние
5. Основные понятия корпусной лингвистики
6. Цели и основные направления корпусной лингвистики
7. История создания лингвистических корпусов
8. Основные характеристики корпусов
9. Корпусные менеджеры
10. Классификация корпусов по различным основаниям
11. Особые типы корпусов
12. Понятие разметки. Разметка. Средства разметки корпусов
13. Стандартизация в корпусной лингвистике
14. Обзор существующих корпусов различных типов
15. Зарубежные национальные корпусы
16. Специальные корпусы
17. Корпусные исследования
18. Пользователи корпусов
19. Формализованное описание языка
20. Словарный и бессловарный способ
21. Морфологический и синтаксический анализ
22. Способы использования корпусов
23. Лексикографические исследования, основанные на корпусах
24. История корпусов по монгольским языкам

5.4. Критерии оценки промежуточной аттестации аспирантов по дисциплине «Корпусная лингвистика»

оценка	критерии оценки на экзамене
«зачтено»	–аспирант знает весь изученный материал; –отвечает без особых затруднений на вопросы преподавателя; –аспирант умеет применять полученные знания на практике; –в условных ответах не допускает ошибок, устраняет определенные неточности с помощью дополнительных вопросов преподавателя.
«не зачтено»	У аспиранта имеются отдельные представления об изучаемом материале, но все же большая часть не усвоена

5.5. Фонд оценочных средств

Содержание фонда оценочных средств см. Приложение.

6. Учебно-методическое и информационное обеспечение дисциплины
Основная литература

№	Автор	Название	Место издания	Издательство	Год издания
1	Захаров В. П.	Корпусная лингвистика. Учебник для студентов гуманитарных вузов	Иркутск	Иркутский гос. лингв. ун-т	2011
2		Практикум по полевой лингвистике, или как создавать звуковые корпуса. Часть 1, 2	Санкт-Петербург	Санкт-Петербургский государственный университет	2006
3	Гельбух А.Ф.	Эффективно реализуемая модель морфологии флективного естественного языка	Москва		1994
4	Ножов И.М.	Морфологическая и синтаксическая обработка текста (модели и программы)	Москва		2003
5		Автоматическая обработка текстов на естественном языке и компьютерная лингвистика	Москва	МИЭМ	2011

Дополнительная литература

Национальный корпус русского языка: 2006–2008. Новые результаты и перспективы. СПб.: Нестор-История, 2009.

Национальный корпус русского языка: 2003–2005. Сборник статей. М.: Индрик, 2005.

Труды международной конференции «Корпусная лингвистика – 2004». Санкт-Петербург: Изд-во С.-Петербургского университета, 2004.

Информационные технологии и письменное наследие. IV Международная научная конференция. Петрозаводск, 2012.

Письменное наследие и информационные технологии. V Международная научная конференция. Ижевск, 2014.

Труды международной конференции «Корпусная лингвистика – 2011». Санкт-Петербург: СПб.: С.-Петербургский гос. университет, Филологический факультет, 2011.

Грудева Е.В. Корпусная лингвистика. М.: Флинта, 2012. <https://books.google.ru>

Щипицина Л.Ю. Информационные технологии в лингвистике. М.: Флинта, 2013. <https://books.google.ru>

Методы когнитивного анализа семантики слова. Компьютерно-корпусный подход. М.: Языки славянской культуры, 2015. <https://books.google.ru>

Хроленко А.Т., Денисов А.В. Современные информационные технологии для гуманитария. М.: Наука, 2007. <https://books.google.ru>

The Routledge Handbook of Corpus Linguistics. London, 2010.

Creating and Digitizing Language Corpora. Vol. 1. London, 2007.

English Corpus Linguistics. An Introduction. Cambridge, 2004.

7. Особенности организации образовательного процесса для инвалидов и лиц с ограниченными возможностями здоровья

Обучение инвалидов и лиц с ограниченными возможностями здоровья осуществляется в соответствии с:

1. ст. 79, 273-ФЗ «Об образовании в Российской Федерации»;
2. Раздел IV, п.п. 46-51 приказа Минобрнауки России от 19.11.2013 № 1259 «Об утверждении Порядка организации и осуществления образовательной деятельности по образовательным программам высшего образования – программам подготовки научно-педагогических кадров в аспирантуре (адъюнктуре)»;
3. Федеральным законом РФ от 24.11.1995 г. №181-ФЗ «О социальной защите инвалидов в Российской Федерации»;
4. Методическими рекомендациями по организации образовательного процесса для обучения инвалидов и лиц с ограниченными возможностями здоровья в образовательных организациях высшего образования, в том числе оснащённости образовательного процесса, утвержденными Министерством образования и науки РФ 08.04.2014 г. №АК-44/05вн.

8. Средства обеспечения освоения дисциплины. Материально-техническое обеспечение

В лекциях используются мультимедийные презентации, которые выносятся на большой экран. Также имеются библиотечный фонд КИГИ РАН, компьютерная техника.

Работа в аудитории осуществляется в соответствии с семестровыми рабочими учебными планами, графиками учебного процесса, расписанием аудиторных занятий.

Общие сведения об аудитории: кабинет №137, общая площадь 89 м², высота помещения 3,25 м, площадь, отведенная обучающимся, – 89 м², количество учебных мест – 4.

<http://www.moderna.uu.se/slaviska/ryska/corpus/>

<http://www.narusco.ru/>

<http://cfrl.ru/>

<http://opencorpora.org/>

<http://spokencorpora.ru/>

<http://www-korpus.uni-r.de/diakorp/index.php>

<http://www.lrc-lib.ru/>

<http://mns.udsu.ru/>

<http://smalt.karelia.ru/corpus/index.phtml>

<http://www.americancorpora.org/>

http://web-corpora.net/KalmykCorpus/search/index.php?interface_language=ru

http://web-corpora.net/MongolianCorpus/search/index.php?interface_language=ru

http://web-corpora.net/BuryatCorpus/search/index.php?interface_language=ru

<http://kalmcorpora.ru/>

Содержание фонда оценочных средств

Фонд оценочных средств по дисциплине «Корпусная лингвистика» составляется из контрольных вопросов для оценки владения материалом (см. выше п. 5.3); для оценки практического владения использованием корпусных систем предусмотрены письменные работы следующего содержания:

Вариант 1.

Задание 1. Дайте обоснованный ответ.

При создании корпусов монгольских языков, пользующихся кириллическим письмом, часто возникает необходимость транслитерации кириллической орфографической записи в латиницу. Необходимость эта определяется как потребностями международного лингвистического сообщества, так и чисто техническими возможностями при автоматической обработке некоторых данных. Какие проблемы встают при транслитерации монгольского (калмыцкого) алфавита в латинские символы? Предложите свой (непротиворечивый) вариант латинско-монгольских (латинско-калмыцких) буквенных соответствий.

Задание 2.

Поставьте несколько конкретных лингвистических задач в области морфологии монгольских языков, решение которых без обращения к корпусу невозможно.

Задание 3.

Одной из важных характеристик речи является частота словоупотреблений местоимений в монгольских языках. Как вы думаете, какие типы местоимений чаще употребляются в художественной литературе? Проверьте свои соображения, обратившись к корпусным системам монгольских языков. Дайте лингвистический комментарий своим наблюдениям.

Задание 4.

Пользуясь имеющимися корпусами по монгольским языкам, а также по английскому языку, сравните среднюю длину:

- а) словоформы числительного,
- б) словоформы существительного

в английском и калмыцком языках. Опираясь на знания о строе этих языков, дайте комментарий своим наблюдениям.

Вариант 2.

Задание 1. Дайте обоснованный ответ.

При создании корпусов монгольских языков, пользующихся вертикальным письмом, часто возникает необходимость транслитерации вертикального письма в латиницу, что делается традиционно. Необходимость эта определяется как потребностями международного лингвистического сообщества, так и чисто техническими возможностями при автоматической обработке текстовых данных, до настоящего времени не имеющих кодировки. Какие системы транслитерации Вы знаете? Выявите их преимущества и недостатки. Какие проблемы встают при транслитерации монгольского (калмыцкого) алфавита в латинские символы?

Задание 2.

Поставьте несколько конкретных лингвистических задач в области синтаксиса монгольских языков, решение которых без обращения к корпусу невозможно.

Задание 3.

Одной из важных характеристик речи является частота словоупотреблений местоимений в монгольских языках. Как вы думаете, какие формы местоимений чаще употребляются? Проверьте свои соображения, обратившись к корпусным системам монгольских языков. Дайте лингвистический комментарий своим наблюдениям.

Задание 4.

Пользуясь имеющимися корпусами по монгольским языкам, а также по английскому языку, сравните среднюю длину:

- а) словоформы причастия,
- б) словоформы деепричастия

в английском и калмыцком языках. Опираясь на знания о строе этих языков, дайте комментарий своим наблюдениям.