

ФГБОУ ВО «БАШКИРСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»  
НАИМЕНОВАНИЕ ФИЛИАЛА  
НАИМЕНОВАНИЕ ФАКУЛЬТЕТА (ИНСТИТУТА)

Утверждено:  
на заседании кафедры  
протокол № 7 от «22» января 2021 г.  
Зав. кафедрой \_\_\_\_\_ / Самигуллина А.С.

Согласовано:  
Председатель УМК факультета /института

\_\_\_\_\_ / \_\_\_\_\_

**РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ (МОДУЛЯ)**

дисциплина **Корпусные методы исследования**  
*(наименование дисциплины)*

Часть, формируемая участниками образовательных отношений

*(указать часть (обязательная часть или часть, формируемая участниками образовательных отношений, факультатив))*

**программа бакалавриата**

Направление подготовки (специальность)

45.03.03 Фундаментальная и прикладная лингвистика

*(указывается код и наименование направления подготовки (специальности))*

Направленность (профиль) подготовки


Языковые технологии

*(указывается наименование направленности (профиля) подготовки)*

Квалификация

Бакалавр

*(указывается квалификация)*

Разработчик (составитель) доц., к. филол. н. Кошеварова Ю. А. (должность, ученая степень, ученое звание)	 _____ / Кошеварова Ю. А. (подпись, Фамилия И.О.)
--	---

Для приема: 2021 г.

Уфа 2021 г.

Составитель / составители: Кошеварова Ю.А.

Рабочая программа дисциплины *утверждена* на заседании кафедры протокол от «22» января 2021 г. № 7

Дополнения и изменения, внесенные в рабочую программу дисциплины, утверждены на заседании кафедры \_\_\_\_\_

\_\_\_\_\_  
\_\_\_\_\_  
протокол № \_\_\_\_ от « \_\_\_\_ » \_\_\_\_\_ 20 \_ г.

Заведующий кафедрой \_\_\_\_\_ / \_\_\_\_\_ Ф.И.О./

Дополнения и изменения, внесенные в рабочую программу дисциплины, утверждены на заседании кафедры \_\_\_\_\_

\_\_\_\_\_  
\_\_\_\_\_  
протокол № \_\_\_\_ от « \_\_\_\_ » \_\_\_\_\_ 20 \_ г.

Заведующий кафедрой \_\_\_\_\_ / \_\_\_\_\_ Ф.И.О./

Дополнения и изменения, внесенные в рабочую программу дисциплины, утверждены на заседании кафедры \_\_\_\_\_

\_\_\_\_\_  
\_\_\_\_\_  
протокол № \_\_\_\_ от « \_\_\_\_ » \_\_\_\_\_ 20 \_ г.

Заведующий кафедрой \_\_\_\_\_ / \_\_\_\_\_ Ф.И.О./

Дополнения и изменения, внесенные в рабочую программу дисциплины, утверждены на заседании кафедры \_\_\_\_\_

\_\_\_\_\_  
\_\_\_\_\_  
протокол № \_\_\_\_ от « \_\_\_\_ » \_\_\_\_\_ 20 \_ г.

Заведующий кафедрой \_\_\_\_\_ / \_\_\_\_\_ Ф.И.О./

## Список документов и материалов

1. Перечень планируемых результатов обучения по дисциплине, соотнесенных с установленными в образовательной программе индикаторами достижения компетенций 4
2. Цель и место дисциплины в структуре образовательной программы 4
3. Содержание рабочей программы (объем дисциплины, типы и виды учебных занятий, учебно-методическое обеспечение самостоятельной работы обучающихся) 4
4. Фонд оценочных средств по дисциплине 5
  - 4.1. Перечень компетенций и индикаторов достижения компетенций с указанием соотнесенных с ними запланированных результатов обучения по дисциплине. Описание критериев и шкал оценивания результатов обучения по дисциплине. 5
  - 4.2. Типовые контрольные задания или иные материалы, необходимые для оценивания результатов обучения по дисциплине, соотнесенных с установленными в образовательной программе индикаторами достижения компетенций. Методические материалы, определяющие процедуры оценивания результатов обучения по дисциплине. 5
5. Учебно-методическое и информационное обеспечение дисциплины 11
  - 5.1. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины 11
  - 5.2. Перечень ресурсов информационно-телекоммуникационной сети «Интернет» и программного обеспечения, необходимых для освоения дисциплины, включая профессиональные базы данных и информационные справочные системы 11
6. Материально-техническая база, необходимая для осуществления образовательного процесса по дисциплине 13

# 1. Перечень планируемых результатов обучения по дисциплине, соотнесенных с установленными в образовательной программе индикаторами достижения компетенций

По итогам освоения дисциплины обучающийся должен достичь следующих результатов обучения:

Категория (группа) компетенций	Формируемая компетенция (с указанием кода)	Код и наименование индикатора достижения компетенции	Результаты обучения по дисциплине
Информационно-коммуникационная культура.	ПК-10. Способен владеть принципами создания электронных языковых ресурсов (текстовых, речевых и мультимодальных корпусов; словарей, тезаурусов, онтологий; фонетических, лексических, грамматических и иных баз данных и баз знаний) и умением пользоваться такими ресурсами.	ПК-10.1. Знать: разновидности электронных языковых ресурсов (текстовых, речевых и мультимодальных корпусов; словарей, тезаурусов, онтологий; фонетических, лексических, грамматических и иных баз данных и баз знаний).	Знает: разновидности электронных языковых ресурсов (текстовых, речевых и мультимодальных корпусов; словарей, тезаурусов, онтологий; фонетических, лексических, грамматических и иных баз данных и баз знаний).
		ПК-10.2. Уметь: создавать и пользоваться электронными языковыми ресурсами (текстовые, речевые и мультимодальные корпусы; словари, тезаурусы, онтологий; фонетические, лексические, грамматические и иные базы данных и базы знаний).	Умеет: создавать и пользоваться электронными языковыми ресурсами (текстовые, речевые и мультимодальные корпусы; словари, тезаурусы, онтологий; фонетические, лексические, грамматические и иные базы данных и базы знаний).
		ПК-10.3. Владеть: навыками отбора материала и составления электронных языковых ресурсов (текстовых, речевых и мультимодальных корпусов; словарей, тезаурусов, онтологий; фонетических, лексических, грамматических и иных баз данных и баз знаний).	Владеет: навыками отбора материала и составления электронных языковых ресурсов (текстовых, речевых и мультимодальных корпусов; словарей, тезаурусов, онтологий; фонетических, лексических, грамматических и иных баз данных и баз знаний).

## 2. Цель и место дисциплины в структуре образовательной программы

Дисциплина «Корпусные методы исследования» входит в часть, формируемую участниками образовательных отношений. Дисциплина по выбору

Дисциплина изучается на 3 курсе в 2 семестре.

Целью данной учебной дисциплины является освоение студентами знаний в области методологии корпусных исследований.

Для освоения дисциплины необходимы знания, умения и компетенции, сформированные у обучающихся в результате прохождения теоретических дисциплин на 1 и 2 курсах обучения. Изучение дисциплины необходимо для дальнейшего развития филологической компетенции обучающихся на уровне теоретического осмысления сущности языка, а также прикладных навыков работы с корпусами языковых данных.

## 3. Содержание рабочей программы (объем дисциплины, типы и виды учебных занятий, учебно-методическое обеспечение самостоятельной работы обучающихся)

Содержание рабочей программы представлено в Приложении № 1.

#### 4. Фонд оценочных средств по дисциплине

##### 4.1. Перечень компетенций индикаторов достижения компетенций с указанием соотнесенных с ними запланированных результатов обучения по дисциплине. Описание критериев и шкал оценивания результатов обучения по дисциплине.

Код и формулировка компетенции: ПК-10 – Способен владеть принципами создания электронных языковых ресурсов и умением пользоваться такими ресурсами.

Код и наименование индикатора достижения компетенции	Результаты обучения по дисциплине	Критерии оценивания результатов обучения	
		«Не зачтено»	«Зачтено»
ПК-10. Способен владеть принципами создания электронных языковых ресурсов (текстовых, речевых и мультимодальных корпусов; словарей, тезаурусов, онтологий; фонетических, лексических, грамматических и иных баз данных и баз знаний).	Знает: разновидности электронных языковых ресурсов (текстовых, речевых и мультимодальных корпусов; словарей, тезаурусов, онтологий; фонетических, лексических, грамматических и иных баз данных и баз знаний).	Знает разновидности электронных языковых ресурсов (текстовых, речевых и мультимодальных корпусов; словарей, тезаурусов, онтологий; фонетических, лексических, грамматических и иных баз данных и баз знаний).	Не знает разновидности электронных языковых ресурсов (текстовых, речевых и мультимодальных корпусов; словарей, тезаурусов, онтологий; фонетических, лексических, грамматических и иных баз данных и баз знаний).
	Умеет: создавать и пользоваться электронными языковыми ресурсами (текстовые, речевые и мультимодальные корпусы; словари, тезаурусы, онтологий; фонетические, лексические, грамматические и иные базы данных и базы знаний).	Совсем не умеет создавать и пользоваться электронными языковыми ресурсами (текстовые, речевые и мультимодальные корпусы; словари, тезаурусы, онтологий; фонетические, лексические, грамматические и иные базы данных и базы знаний).	Удовлетворительно умеет создавать и пользоваться электронными языковыми ресурсами (текстовые, речевые и мультимодальные корпусы; словари, тезаурусы, онтологий; фонетические, лексические, грамматические и иные базы данных и базы знаний).
	Владет: навыками отбора материала и составления электронных языковых ресурсов (текстовых, речевых и мультимодальных корпусов; словарей, тезаурусов, онтологий; фонетических, лексических, грамматических и иных баз данных и баз знаний).	Совсем не владеет навыками отбора материала и составления электронных языковых ресурсов (текстовых, речевых и мультимодальных корпусов; словарей, тезаурусов, онтологий; фонетических, лексических, грамматических и иных баз данных и баз знаний).	Удовлетворительно владеет навыками отбора материала и составления электронных языковых ресурсов (текстовых, речевых и мультимодальных корпусов; словарей, тезаурусов, онтологий; фонетических, лексических, грамматических и иных баз данных и баз знаний).

##### 4.2. Типовые контрольные задания или иные материалы, необходимые для оценивания результатов обучения по дисциплине, соотнесенных с установленными в образовательной программе индикаторами достижения компетенций. Методические материалы, определяющие процедуры оценивания результатов обучения по дисциплине.

Код и наименование индикатора достижения компетенции	Результаты обучения по дисциплине	Оценочные средства
ПК-10. Способен владеть принципами создания электронных языковых ресурсов (текстовых, речевых и мультимодальных корпусов; словарей, тезаурусов, онтологий; фонетических, лексических, грамматических и иных баз данных и баз знаний) и умением пользоваться такими ресурсами.	ПК-10.1. Знать: разновидности электронных языковых ресурсов(текстовых, речевых и мультимодальных корпусов; словарей, тезаурусов, онтологий; фонетических, лексических, грамматических и иных баз данных и баз знаний).	Групповой опрос, письменные ответы на вопросы, индивидуальное задание.
	ПК-10.2. Уметь: создавать и пользоваться электронными языковыми ресурсами (текстовые, речевые и мультимодальные корпусы; словари, тезаурусы, онтологий; фонетические, лексические, грамматические и иные базы данных и базы знаний).	Ситуационные задачи.
	ПК-10.3. Владеть: навыками отбора материала и составления электронных языковых ресурсов (текстовых, речевых и мультимодальных корпусов; словарей, тезаурусов, онтологий; фонетических, лексических, грамматических и иных баз данных и баз знаний).	Групповой опрос, письменные ответы на вопросы, индивидуальное задание.

### Показатели сформированности компетенций.

Критериями оценивания являются баллы, которые выставляются преподавателем за виды деятельности (оценочные средства) по итогам изучения модулей (разделов дисциплины), перечисленных в рейтинг-плане дисциплины (текущий контроль – максимум 50 баллов; рубежный контроль – максимум 50 баллов, поощрительные баллы – максимум 10).

**Шкалы оценивания:** зачтено – от 60 до 110 рейтинговых баллов (включая 10 поощрительных баллов), не зачтено – от 0 до 59 рейтинговых баллов.

## Рейтинг – план дисциплины

### Корпусные методы исследования

(название дисциплины согласно рабочему учебному плану)

направление/специальность 45.03.03 Фундаментальная и прикладная лингвистика  
курс 3, семестр 2

Виды учебной деятельности студентов	Балл за конкретное задание	Число заданий за семестр	Баллы	
			Минимальный	Максимальный
<b>Модуль 1. Сущность и метаязык корпусных исследований</b>				
<b>Текущий контроль</b>				
1. Аудиторная работа (групповой опрос)	2	5	0	10
2. Выполнение домашних заданий (индивидуальное задание)	5	3	0	15
<b>Рубежный контроль</b>				
1. Письменная контрольная работа № 1	25	1	0	25
<b>Итоговый балл за модуль 1</b>			<b>0</b>	<b>50</b>
<b>Модуль 2. Виды корпусов и основные методики работы с ними</b>				
<b>Текущий контроль</b>				
1. Аудиторная работа (групповой опрос)	2	5	0	10
2. Выполнение домашних заданий (индивидуальное задание)	5	3	0	15
<b>Рубежный контроль</b>				
1. Письменная контрольная работа № 2	25	1	0	25
<b>Итоговый балл за модуль 2</b>			<b>0</b>	<b>50</b>
<b>Итого (Модули 1-2)</b>			<b>0</b>	<b>100</b>
<b>Поощрительные баллы</b>				
Выступление с докладом, презентацией	5	2	0	10
<b>Итого</b>			<b>0</b>	<b>110</b>
<b>Посещаемость (баллы вычитаются из общей суммы набранных баллов)</b>				
Посещение лекционных занятий			0	-6

### Перечень вопросов для письменного опроса:

1. Понятие лингвистического корпуса. История корпусной лингвистики. Корпусный (эмпирический) подход в сравнении с хомскианской лингвистикой.
2. Задачи и основные направления корпусной лингвистики. Корпусная лингвистика и компьютерная лингвистика. Предмет исследования.
3. Определение лингвистического корпуса. Отличие корпуса от электронной библиотеки. История создания лингвистических корпусов. Основные понятия. Лингвистические и нелингвистические корпуса. Проблема репрезентативности.
4. Развитие лингвистических корпусов в мире. Первое и второе поколение корпусов.
5. Экстралингвистическая метаразметка. Лингвистическая разметка. Автоматическая морфологическая и синтаксическая разметка. Снятие неоднозначности.
6. Лемматизация. Графематическая, дискурсивная, семантическая разметка. Жанровая разметка и другие виды разметки.
7. Языковые средства представления размеченных текстов. Международные стандарты и проекты (TEI, EAGLES, CDIF, XCES). Проблемы, возникающие при разметке текстов.
8. Создание и типология корпусов. Типы корпусов: устные и письменные, одноязычные и многоязычные, аннотированные и неаннотированные.
9. Национальный корпус русского языка. Корпусы английского языка. Практическое использование национальных корпусов. Проблемы и перспективы развития национальных корпусов.

10. Технология создания корпусов. Стадии работы
11. Корпус как особый тип информационно-поисковой системы. Разбиение на подсистемы.
12. Языки запросов корпусных менеджеров. Выходные интерфейсы
13. Сравнительный анализ корпусов и корпусных менеджеров.
14. Поиск в корпусах: конкордансер и корпусный менеджер как поисковая система.
15. Сравнение корпусных менеджеров с архитектурой поисковых систем в сети Интернет. Поисковые системы Интернет как корпусные менеджеры.
16. Специализированные корпусные интерфейсы к индексам глобальных поисковых систем
17. Параллельные корпуса. Конкорданс. Многоязычие. Проблема выравнивания. Примеры параллельных корпусов
18. Использование корпусов в обучении языкам и гуманитарных исследованиях. Лингвистические исследования на базе корпуса: изучение лексики.
19. Лингвистические исследования на базе корпуса: изучение грамматики.
20. Использование корпусов в настройке и самообучении лингвистических автоматов.
21. Методы извлечения информации из корпуса. Типы извлекаемой информации.
22. Пользуясь Поэтическим корпусом, докажите, что это не единственный в 19 веке
23. Назовите первый пример в Поэтическом корпусе, где с уверенностью можно сказать, что в корне слова произносится Е, а не Ё.
24. Из социальной сети Facebook нередко приходит сообщение: «Интересуюсь о Вас на Facebook. Здравствуйте, Natasha! Я Вас приглашаю посмотреть мою личную страницу на Facebook...».
25. Воспользовавшись корпусом, оцените, насколько конструкция «интересоваться о ком-либо» распространена, в каких текстах она встречается. Сравните ее частотность с частотностью конструкции «интересоваться кем-либо».
26. Найдите в Корпусе все случаи употребления сочетания "скучать по вас" и "скучать по вам". Сколько их? Какие примеры для каждой конструкции самые ранние? Какие самые поздние? Приведите эти примеры. Основываясь на этих данных, прокомментируйте ответ Справочной службы ИРЯ.
27. Посчитайте соотношение частоты употребления слов «этот» и «сей» в следующие периоды: а) 1770-1799 б) 1800-1829 в) 1830-1869
28. Когда в Корпусе появляются первые примеры глагола «реагировать»? Приведите первые два примера. Что они значат? Когда появляются первые примеры глагола «реагировать» с какой-либо приставкой? Приведите первые два примера.
29. Какая разметка из приведенных ниже является морфологической, а какая семантической?

1) \_\_\_\_\_

```

<?xml version="1.0" encoding="windows-1251" ?>
    <text>
        <p>
            <s>
                <w>Звонили<ana lemma="ЗВОНИТЬ" pos="Г" gram="мн,нс,нп,дст,прш," /></w>
                <w>к<ana lemma="К" pos="ПРЕДЛ" gram="" /></w>
            </s>
            <w>вечерне
                <ana lemma="ВЕЧЕРНЯ" pos="С" gram="жр,ед,дт,пр,но," />
                <ana lemma="ВЕЧЕРНИЙ" pos="П" gram="ср,ед,кр," /></w>
                <pun>.</pun> </s>
            <s><w>Торжественный<ana lemma="ТОРЖЕСТВЕННЫЙ" pos="П"
                gram="мр,ед,им,вн," /></w>
                <w>гул<ana lemma="ГУЛ" pos="С" gram="мр,ед,им,вн,но," /></w>
            </s>
            <w>колоколов
                <ana lemma="КОЛОКОЛ" pos="С" gram="мр,мн,рд,но," />
                <ana lemma="КОЛОКОЛОВ" pos="С" gram="мр,фам,ед,им,од," /></w>
            .....<pun>.</pun> </s></p></text>

```



2)

```
<?xml version="1.0" encoding="cp866"?>
  <МЕТА-ДОКУМЕНТ>
    <ДОКУМЕНТ>
<НАЗВАНИЕ>Сообщение газеты "Мир Кино"</НАЗВАНИЕ>
<ДАТА>20.03.2000</ДАТА>
  <УЧАСТНИК_ОБЩЕНИЯ Имя="Мир Кино" Тип="Автор">
    <СОЦИАЛЬНЫЙ_СТАТУС>
<ХАРАКТЕРИСТИКА>Печатное издание</ХАРАКТЕРИСТИКА>
    </СОЦИАЛЬНЫЙ_СТАТУС>
  </УЧАСТНИК_ОБЩЕНИЯ>
</ДОКУМЕНТ>
```

29. Создайте морфологическую разметку предложения: Модуль семантического анализа текстов предназначен для нормализации синтаксической структуры, распознавания терминов, классификации терминов по семантическим признакам.
30. Объясните, что такое корпус? Что такое «Корпус как идеология» в представлении акад. Плунгяна.
31. Виды разметки в корпусе и средства их автоматизации.
32. Виды корпусов, их назначение. Приведите примеры 3-4 видов корпусов, дайте описание корпусных менеджеров (ИПС).
33. Способы использования корпусов в лингвистических и филологических исследованиях.
34. Национальный корпус русского языка: назначение, принципы создания, возможности поиска.

**Вопросы для группового устного и письменного опроса, включая рубежную письменную контрольную работу, коррелируют с темами лекционных занятий.**

**Примерные вопросы для группового устного и письменного опроса.**

**По разделу 1 (Сущность и метаязык корпусных исследований).**

1. Понятие лингвистического корпуса. История корпусной лингвистики. Корпусный (эмпирический) подход в сравнении с хомскианской лингвистикой. Задачи и основные направления корпусной лингвистики. Корпусная лингвистика и компьютерная лингвистика. Предмет исследования.
2. Определение лингвистического корпуса. Отличие корпуса от электронной библиотеки. История создания лингвистических корпусов. Основные понятия. Лингвистические и нелингвистические корпуса. Проблема репрезентативности. Развитие лингвистических корпусов в мире. Первое и второе поколение корпусов.
3. Экстралингвистическая метаразметка. Лингвистическая разметка. Автоматическая морфологическая и синтаксическая разметка. Снятие неоднозначности. Лемматизация. Графематическая, дискурсивная, семантическая разметка. Жанровая разметка и другие виды разметки. Процедуры лемматизации и разметка корпуса. Обзор автоматических средств разметки.
4. Языковые средства представления размеченных текстов. Международные стандарты и проекты (TEI, EAGLES, CDIF, XCES). Проблемы, возникающие при разметке текстов.
5. Разметка корпуса. Виды разметок. Создание корпуса в MOODLE. Метаразметка корпуса.

**По разделу 2 (Виды корпусов и основные методики работы с ними).**

1. Создание и типология корпусов. Типы корпусов: устные и письменные, одноязычные и многоязычные, аннотированные и неаннотированные.
2. Национальный корпус русского языка. Корпусы английского языка. Практическое использование национальных корпусов. Проблемы и перспективы развития национальных корпусов.
3. Технология создания корпусов. Стадии работы
4. Создание и разметка корпуса на основе WIKI. Средства разметки корпуса. Анализ программных средств. Создание корпуса в MOODLE. Метаразметка корпуса. Возможности тегирования записей в корпусе.

5. Языки запросов корпусных менеджеров. Выходные интерфейсы.

**Критерии оценки (в баллах) для группового устного и письменного опроса:**

- 2 балла выставляется студенту, показавшему всесторонние, глубокие знания темы, свободное и правильное обоснование проблемных ситуаций;
- 1 балл выставляется студенту, показавшему не достаточно глубокие знания темы, не умеющему свободно и в полной мере аргументировано обосновать проблемную ситуацию;
- 0 баллов выставляется студенту, который не знает большей части основного содержания темы, допускает грубые ошибки в формулировках основных понятий и не способен решать типовые практические задачи или проблемные ситуации.

**Пример рубежной письменной контрольной работы по Разделу 1 (Модулю 1).**

1. Понятие лингвистического корпуса. История корпусной лингвистики. Корпусный (эмпирический) подход в сравнении с хомскианской лингвистикой.
2. Задачи и основные направления корпусной лингвистики. Корпусная лингвистика и компьютерная лингвистика. Предмет исследования.
3. Определение лингвистического корпуса.
4. Отличие корпуса от электронной библиотеки.
5. История создания лингвистических корпусов. Основные понятия. Лингвистические и нелингвистические корпуса.
6. Проблема репрезентативности.
7. Развитие лингвистических корпусов в мире. Первое и второе поколение корпусов.
8. Экстралингвистическая метаразметка.
9. Лингвистическая разметка.
10. Автоматическая морфологическая и синтаксическая разметка.
11. Снятие неоднозначности.

**Критерии оценки (в баллах) рубежной письменной контрольной работы:**

- 25 баллов выставляется студенту, показавшему всесторонние, систематизированные, глубокие знания учебной программы модуля и умение применять их на практике при решении конкретных задач, свободное и правильное обоснование проблемных ситуаций;
- 15 баллов выставляется студенту, показавшему не достаточно глубокое знание теории межкультурной коммуникации, не умеющему в полной мере аргументировано обосновать решение конкретных задач;
- 0 баллов выставляется студенту, который не знает большей части основного содержания учебной программы модуля, допускает грубые ошибки в формулировках основных понятий дисциплины и не умеет использовать полученные знания при решении типовых практических задач.

**Темы индивидуальных заданий студентов:**

1. Сравнительный анализ корпусов и корпусных менеджеров. Поиск в корпусах: конкордансер и корпусный менеджер как поисковая система. Поисковые системы Интернет как корпусные менеджеры.
2. Сравнение корпусных менеджеров с архитектурой поисковых систем в сети Интернет. Поисковые системы Интернет как корпусные менеджеры. Специализированные корпусные интерфейсы к индексам глобальных поисковых систем
3. Параллельные корпуса. Конкорданс. Многоязычие. Проблема выравнивания. Примеры параллельных корпусов. Параллельные корпуса. Конкорданс. Анализ программных средств создания конкордансов и существующих параллельных корпусов.
4. Использование корпусов в обучении языкам и гуманитарных исследованиях. Лингвистические исследования на базе корпуса: изучение лексики. Лингвистические исследования на базе корпуса НКРЯ и BNC: изучение грамматики, исследование семантики (метафора, этимология).
5. Способы использования корпусов в лингвистических исследованиях.
6. Исследование способов использования корпусов в лексикографии.
7. Изучение средств обработки корпусных данных, представленных на языке XML.

8. Создание электронной хрестоматии по корпусной лингвистике.
9. Исследование механизмов взаимодействия корпуса текстов и электронной картотеки (корпусы цитат).
10. Создание веб-сайта по корпусной лингвистике.
11. Британский национальный корпус.
12. Национальный корпус русского языка: история создания, перспективы развития.
13. Технологии морфологической и синтаксической разметки НКРЯ.
14. Программа MyStem Yandex: разработка, использование, перспективы.

Индивидуальные задания выполняются в форме краткого устного сообщения (доклада) или медиапрезентации и оцениваются поощрительными баллами.

#### **Критерии оценки (в баллах) для индивидуальных заданий:**

- 5 баллов выставляется студенту, продемонстрировавшему самостоятельное, всестороннее, систематизированное, глубокое раскрытие темы, свободное владение материалом;
- 2 балла выставляется студенту, показавшему не достаточно самостоятельное и глубокое понимание темы, не умеющему свободно представить материал;
- 0 баллов выставляется студенту, который не понимает большей части основного содержания сообщения, не способен ясно, последовательно представить материал.

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

#### **5.1. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины**

##### **Основная литература:**

1. Захаров В.П., Богданова С.Ю. Корпусная лингвистика: учебник для студентов гуманитарных вузов. – Иркутск: ИГЛУ, 2011. – 161 с. ISBN: 978-5-88267-316-0, [http://lib.biblioclub.ru/book\\_89753\\_Korpusnaya\\_lingvistika\\_Uchebnik\\_dlya\\_studentov\\_gumanitarnykh\\_vuzov](http://lib.biblioclub.ru/book_89753_Korpusnaya_lingvistika_Uchebnik_dlya_studentov_gumanitarnykh_vuzov)
2. Копотев М. Введение в корпусную лингвистику: учебник для вузов. – Animedia Company, 2014. – 195 с. ISBN: 978-80-7499-067-0 [http://lib.biblioclub.ru/book\\_375463\\_vvedenie\\_v\\_korpusnyu\\_lingvistiku](http://lib.biblioclub.ru/book_375463_vvedenie_v_korpusnyu_lingvistiku)

##### **Дополнительная литература:**

1. Захарова, Т.В. Практические основы компьютерных технологий в переводе: учебное пособие/ Т.В. Захарова, Е.В. Турлова; Министерство образования и науки Российской Федерации, Оренбургский Государственный Университет. - Оренбург: Оренбургский государственный университет, 2017. – 109 с.: табл., граф., ил. - Библиогр.: с. 104 - ISBN 978-5-7410-1736-4; То же [Электронный ресурс]. - URL: <http://biblioclub.ru/index.php?page=book&id=481823>
2. Актуальные проблемы современной лингвистики: учебное пособие / сост. Л.Н. Чурилина. - 10-е изд., стереотип. – М.: Издательство «Флинта», 2017. - 412 с. - ISBN 978-5-89349-892-9; То же [Электронный ресурс]. URL: <http://biblioclub.ru/index.php?page=book&id=103797> (07.04.2019)
3. Грудева, Е.В. Корпусная лингвистика [Электронный ресурс] : учебное пособие / Е.В. Грудева. – Электрон. дан. – М.: ФЛИНТА, 2017. – 165 с. - Режим доступа: <https://e.lanbook.com/book/106859>
4. Колпакова Г.В. Корпусная лингвистика и лексикография Грани познания – 2011г. №2 <https://e.lanbook.com/reader/journalArticle/108607/#1>

#### **5.2. Перечень ресурсов информационно-телекоммуникационной сети «Интернет» и программного обеспечения, необходимых для освоения дисциплины, включая профессиональные базы данных и информационные справочные системы**

1. Ресурсы электронной библиотеки «Гумер». – <http://www.gumer.info>.
2. Ресурсы Электронно-библиотечной системы "Университетская библиотека online". – <http://www.biblioclub.ru>.

3. Академик (словари и энциклопедии). – < <http://dic.academic.ru/> >
4. Всероссийская государственная библиотека иностранной литературы им. М.И. Рудомино). – < <http://www.libfl.ru/> >
5. Российская государственная библиотека. – < <http://www.rsl.ru/> >
6. Фундаментальная электронная библиотека. – < <http://feb-web.ru/>>
7. The Library of Congress. – < <http://www.loc.gov/> >
8. Project Gutenberg. A digital library of free books. – <<http://www.gutenberg.org/>>
9. Questia online library. – < <http://www.questia.com/>>
10. Windows 8 Russian. Windows Professional 8 Russian Upgrade. Договор № 104 от 17.06.2013 г. Лицензии бессрочные.
11. Microsoft Office Standard 2013 Russian. Договор № 114 от 12.11.2014 г. Лицензии бессрочные.

## 6. Материально-техническая база, необходимая для осуществления образовательного процесса по дисциплине

Компьютерный класс, мультимедийный проектор, экран, доска, оргтехника, аудиоаппаратура (в стандартной комплектации для лекционных занятий и самостоятельной работы студентов); доступ к сети Интернет (во время самостоятельной подготовки и на лекционных занятиях).

<i>Наименование специализированных аудиторий, кабинетов, лабораторий</i>	<i>Вид занятий</i>	<i>Наименование оборудования, программного обеспечения</i>
<i>1</i>	<i>2</i>	<i>3</i>
<p><b>1. Учебная аудитория для проведения занятий лекционного типа:</b> аудитория №17, аудитория № 31 (ул. Коммунистическая, д. 19, лит. А, А1)</p>	<p><i>Лекции</i></p>	<p><b>Аудитория №31</b> Учебная мебель, учебно-наглядные пособия, доска, мультимедиа-проектор Mitsubishi EX320U XGA, экран настенный Classic Norma 244*183, ноутбук ASUS X51RL (место хранения деканат ФРГФ, ауд. № 6а).</p>
<p><b>2. Учебная аудитория для проведения занятий семинарского типа:</b> аудитория №17, аудитория №11, аудитория №31 (Мультимедийный класс), аудитория № 09, аудитория №0 4 (ул. Коммунистическая, д. 19, лит. А, А1)</p>	<p><i>Семинарские занятия</i></p>	<p><b>Аудитория №17</b> Учебная мебель, учебно-наглядные пособия, доска, переносной проектор и ноутбук Lenovo (фактическое место хранения – деканат, ауд. 6а) <b>Аудитория №09</b> Учебная мебель, учебно-наглядные пособия, доска.</p>
<p><b>3. Учебная аудитория для проведения групповых и индивидуальных консультаций, учебная аудитория для текущего контроля и промежуточной аттестации:</b> аудитория №17, аудитория №11, аудитория № 31, аудитория №9, аудитория №4 (Лингафонный кабинет) (ул. Коммунистическая, д. 19, лит. А, А1)</p>	<p><i>Групповые и индивидуальные консультации, текущий контроль промежуточная аттестация</i></p>	<p><b>Аудитория №11</b> Учебная мебель, учебно-наглядные пособия, доска <b>Аудитория №04</b> Учебная мебель, учебно-наглядные пособия, доска, моноблоки – 12 шт. с выходом в Интернет, обеспечивающие доступ к электронной информационно-образовательной среде (ЭИОС) вуза. <b>Аудитория №13</b> Учебная мебель, учебно-наглядные пособия, доска, учебно-методическая литература, многофункциональное устройство – 1 шт., моноблоки – 2 шт. с выходом в Интернет, обеспечивающие доступ к электронной информационно-образовательной среде (ЭИОС) вуза, книжный фонд читального зала ФРГФ.</p>
<p><b>4. Помещения для самостоятельной работы:</b> аудитория №13 (читальный зал) (ул. Коммунистическая, д. 19, лит. А, А1)</p>	<p><i>Самостоятельная работа</i></p>	<p>1. Windows 8 Russian. Windows Professional 8 Russian Upgrade. Договор № 104 от 17.06.2013 г. Лицензии бессрочные 2. Microsoft Office Standard 2013 Russian. Договор № 114 от 12.11.2014 г. Лицензии бессрочные</p>

ФГБОУ ВО «БАШКИРСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»  
 НАИМЕНОВАНИЕ ФИЛИАЛА  
 НАИМЕНОВАНИЕ ФАКУЛЬТЕТА (ИНСТИТУТА)

**СОДЕРЖАНИЕ РАБОЧЕЙ ПРОГРАММЫ**

дисциплины «Корпусные методы исследования» на 6 семестр

Форма обучения: очная

Вид работы	Объем дисциплины
Общая трудоемкость дисциплины (з.е. / часов)	3/108
Учебных часов на контактную работу с преподавателем:	32,2
лекций	16
практических/ семинарских	16
лабораторных	0
других (групповая, индивидуальная консультация и иные виды учебной деятельности, предусматривающие работу обучающихся с преподавателем) (ФКР)	0,2
из них, предусмотренные на выполнение курсовой работы / курсового проекта	0
Учебных часов на самостоятельную работу обучающихся (СР)	75,8
из них, предусмотренные на выполнение курсовой работы / курсового проекта	0
Учебных часов на подготовку к экзамену/зачету/дифференцированному зачету (Контроль)	0

Форма(ы) контроля:

экзамен   0   семестр

зачет   6   семестр

курсовая работа / курсовой проект   0   семестр.

№ п/п	Тема и содержание	Форма изучения материалов: лекции, практические занятия, семинарские занятия, лабораторные работы, самостоятельная работа и трудоемкость (в часах)				Задания по самостоятельной работе студентов	Форма текущего контроля успеваемости (коллоквиумы, контрольные работы, компьютерные тесты и т.п.)
		ЛК	ПР/СЕМ	ЛР	СР		
1	2	3	4	5	6	8	9
1.	Сущность и метаязык корпусных исследований (см. вопросы к Разделу 1).	8	8	-	35	Изучение обязательной и доп. литературы, использование аудио-, видеоматериалов, Интернет-ресурсов.	Устный опрос, письменные ответы на вопросы, сообщение.
2.	Виды корпусов и основные методики работы с ними.	8	8	-	40,8	Изучение обязательной и доп. литературы, использование аудио-, видеоматериалов, Интернет-ресурсов.	Устный опрос, письменные ответы на вопросы, сообщение, письменная контрольная работа.
<b>Всего часов:</b>		16	16	-	75,8		

Макет фонда оценочных средств  
Приложение к  
рабочей программе дисциплины

ФГБОУ ВО «БАШКИРСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»  
НАИМЕНОВАНИЕ ФИЛИАЛА  
НАИМЕНОВАНИЕ ФАКУЛЬТЕТА (ИНСТИТУТА)  
НАИМЕНОВАНИЕ КАФЕДРЫ

**Фонд оценочных средств**

по учебной дисциплине

**Корпусные методы исследования**

---

наименование дисциплины в соответствии с учебным планом

**программа бакалавриата**

45.03.03 Фундаментальная и прикладная лингвистика

---

шифр и наименование направления

**Языковые технологии**

---

направленность (профиль) подготовки



### **Список документов и материалов**

1. Перечень компетенций и индикаторов достижения компетенций с указанием соотнесенных с ними запланированных результатов обучения по дисциплине. Описание критериев и шкал оценивания результатов обучения по дисциплине.
2. Контрольные задания или иные материалы, необходимые для оценивания результатов обучения по дисциплине. Методические материалы, определяющие процедуры оценивания результатов обучения по дисциплине.

**Основное содержание и структура копируются из РПД, ДОБАВЛЯЮТСЯ ПОЛНЫЕ КОМПЛЕКТЫ ОЦЕНОЧНЫХ СРЕДСТВ: варианты контрольных работ, тестов и т.д.**

**1. Перечень компетенций и индикаторов достижения компетенций с указанием соотнесенных с ними запланированных результатов обучения по дисциплине. Описание критериев и шкал оценивания результатов обучения по дисциплине.**

**Код и формулировка компетенции: ПК-10 – Способен владеть принципами создания электронных языковых ресурсов и умением пользоваться такими ресурсами.**

Код и наименование индикатора достижения компетенции	Результаты обучения по дисциплине	Критерии оценивания результатов обучения	
		«Не зачтено»	«Зачтено»
ПК-10. Способен владеть принципами создания электронных языковых ресурсов (текстовых, речевых и мультимодальных корпусов; словарей, тезаурусов, онтологий; фонетических, лексических, грамматических и иных баз данных и баз знаний); фонетических, лексических, грамматических и иных баз данных и баз знаний) и умением пользоваться такими ресурсами.	Знает: разновидности электронных языковых ресурсов (текстовых, речевых и мультимодальных корпусов; словарей, тезаурусов, онтологий; фонетических, лексических, грамматических и иных баз данных и баз знаний).	Знает разновидности электронных языковых ресурсов (текстовых, речевых и мультимодальных корпусов; словарей, тезаурусов, онтологий; фонетических, лексических, грамматических и иных баз данных и баз знаний).	Не знает разновидности электронных языковых ресурсов (текстовых, речевых и мультимодальных корпусов; словарей, тезаурусов, онтологий; фонетических, лексических, грамматических и иных баз данных и баз знаний).
	Умеет: создавать и пользоваться электронными языковыми ресурсами (текстовые, речевые и мультимодальные корпусы; словари, тезаурусы, онтологий; фонетические, лексические, грамматические и иные базы данных и базы знаний).	Совсем не умеет создавать и пользоваться электронными языковыми ресурсами (текстовые, речевые и мультимодальные корпусы; словари, тезаурусы, онтологий; фонетические, лексические, грамматические и иные базы данных и базы знаний).	Удовлетворительно умеет создавать и пользоваться электронными языковыми ресурсами (текстовые, речевые и мультимодальные корпусы; словари, тезаурусы, онтологий; фонетические, лексические, грамматические и иные базы данных и базы знаний).
	Владеет: навыками отбора материала и составления электронных языковых ресурсов (текстовых, речевых и мультимодальных корпусов; словарей, тезаурусов, онтологий; фонетических, лексических, грамматических и иных баз данных и баз знаний).	Совсем не владеет навыками отбора материала и составления электронных языковых ресурсов (текстовых, речевых и мультимодальных корпусов; словарей, тезаурусов, онтологий; фонетических, лексических, грамматических и иных баз данных и баз знаний).	Удовлетворительно владеет навыками отбора материала и составления электронных языковых ресурсов (текстовых, речевых и мультимодальных корпусов; словарей, тезаурусов, онтологий; фонетических, лексических, грамматических и иных баз данных и баз знаний).

**2. Контрольные задания или иные материалы, необходимые для оценивания результатов обучения по дисциплине. Методические материалы, определяющие процедуры оценивания результатов обучения по дисциплине.**

Код и наименование индикатора достижения компетенции	Результаты обучения по дисциплине	Оценочные средства
ПК-10. Способен владеть принципами создания электронных языковых ресурсов (текстовых, речевых и мультимодальных корпусов; словарей, тезаурусов, онтологий; фонетических, лексических, грамматических и иных баз данных и баз знаний) и умением пользоваться такими ресурсами.	ПК-10.1. Знать: разновидности электронных языковых ресурсов(текстовых, речевых и мультимодальных корпусов; словарей, тезаурусов, онтологий; фонетических, лексических, грамматических и иных баз данных и баз знаний).	Групповой опрос, письменные ответы на вопросы, индивидуальное задание.
	ПК-10.2. Уметь: создавать и пользоваться электронными языковыми ресурсами (текстовые, речевые и мультимодальные корпусы; словари, тезаурусы, онтологий; фонетические, лексические, грамматические и иные базы данных и базы знаний).	Ситуационные задачи.
	ПК-10.3. Владеть: навыками отбора материала и составления электронных языковых ресурсов (текстовых, речевых и мультимодальных корпусов; словарей, тезаурусов, онтологий; фонетических, лексических, грамматических и иных баз данных и баз знаний).	Групповой опрос, письменные ответы на вопросы, индивидуальное задание.

Критериями оценивания при *модульно-рейтинговой системе* являются баллы, которые выставляются преподавателем за виды деятельности (оценочные средства) по итогам изучения модулей (разделов дисциплины), перечисленных в рейтинг-плане дисциплины (*для экзамена*: текущий контроль – максимум 40 баллов; рубежный контроль – максимум 30 баллов, поощрительные баллы – максимум 10; *для зачета*: текущий контроль – максимум 50 баллов; рубежный контроль – максимум 50 баллов, поощрительные баллы – максимум 10).

Шкалы оценивания *для экзамена*:

от 45 до 59 баллов – «удовлетворительно»;

от 60 до 79 баллов – «хорошо»;

от 80 баллов – «отлично».

## Рейтинг – план дисциплины

### Корпусные методы исследования

(название дисциплины согласно рабочему учебному плану)

направление/специальность 45.03.03 Фундаментальная и прикладная лингвистика  
курс 3, семестр 6

Виды учебной деятельности студентов	Балл за конкретное задание	Число заданий за семестр	Баллы	
			Минимальный	Максимальный
<b>Модуль 1. Сущность и метаязык корпусных исследований</b>				
<b>Текущий контроль</b>				
1. Аудиторная работа (групповой опрос)	2	5	0	10
2. Выполнение домашних заданий (индивидуальное задание)	5	3	0	15
<b>Рубежный контроль</b>				
1. Письменная контрольная работа № 1	25	1	0	25
<b>Итоговый балл за модуль 1</b>			<b>0</b>	<b>50</b>
<b>Модуль 2. Виды корпусов и основные методики работы с ними</b>				
<b>Текущий контроль</b>				
1. Аудиторная работа (групповой опрос)	2	5	0	10
2. Выполнение домашних заданий (индивидуальное задание)	5	3	0	15
<b>Рубежный контроль</b>				
1. Письменная контрольная работа № 2	25	1	0	25
<b>Итоговый балл за модуль 2</b>			<b>0</b>	<b>50</b>
<b>Итого (Модули 1-2)</b>			<b>0</b>	<b>100</b>
<b>Поощрительные баллы</b>				
Выступление с докладом, презентацией	5	2	0	10
<b>Итого</b>			<b>0</b>	<b>110</b>
<b>Посещаемость (баллы вычитаются из общей суммы набранных баллов)</b>				
Посещение лекционных занятий			0	-6

### Перечень вопросов для письменного опроса:

23. Понятие лингвистического корпуса. История корпусной лингвистики. Корпусный (эмпирический) подход в сравнении с хомскианской лингвистикой.
24. Задачи и основные направления корпусной лингвистики. Корпусная лингвистика и компьютерная лингвистика. Предмет исследования.
25. Определение лингвистического корпуса. Отличие корпуса от электронной библиотеки. История создания лингвистических корпусов. Основные понятия. Лингвистические и нелингвистические корпусы. Проблема репрезентативности.
26. Развитие лингвистических корпусов в мире. Первое и второе поколение корпусов.
27. Экстралингвистическая метаразметка. Лингвистическая разметка. Автоматическая морфологическая и синтаксическая разметка. Снятие неоднозначности.
28. Лемматизация. Графематическая, дискурсивная, семантическая разметка. Жанровая разметка и другие виды разметки.
29. Языковые средства представления размеченных текстов. Международные стандарты и проекты (TEI, EAGLES, CDIF, XCES). Проблемы, возникающие при разметке текстов.
30. Создание и типология корпусов. Типы корпусов: устные и письменные, одноязычные и многоязычные, аннотированные и неаннотированные.
31. Национальный корпус русского языка. Корпусы английского языка. Практическое использование национальных корпусов. Проблемы и перспективы развития национальных корпусов.
32. Технология создания корпусов. Стадии работы

33. Корпус как особый тип информационно-поисковой системы. Разбиение на подсистемы.
34. Языки запросов корпусных менеджеров. Выходные интерфейсы
35. Сравнительный анализ корпусов и корпусных менеджеров.
36. Поиск в корпусах: конкордансер и корпусный менеджер как поисковая система.
37. Сравнение корпусных менеджеров с архитектурой поисковых систем в сети Интернет. Поисковые системы Интернет как корпусные менеджеры.
38. Специализированные корпусные интерфейсы к индексам глобальных поисковых систем
39. Параллельные корпуса. Конкорданс. Многоязычие. Проблема выравнивания. Примеры параллельных корпусов
40. Использование корпусов в обучении языкам и гуманитарных исследованиях. Лингвистические исследования на базе корпуса: изучение лексики.
41. Лингвистические исследования на базе корпуса: изучение грамматики.
42. Использование корпусов в настройке и самообучении лингвистических автоматов.
43. Методы извлечения информации из корпуса. Типы извлекаемой информации.
44. Пользуясь Поэтическим корпусом, докажите, что это не единственный в 19 веке
30. Назовите первый пример в Поэтическом корпусе, где с уверенностью можно сказать, что в корне слова произносится Е, а не Ё.
31. Из социальной сети Facebook нередко приходит сообщение: «Интересуюсь о Васна Facebook. Здравствуйте, Natasha! Я Вас приглашаю посмотреть мою личную страницу на Facebook...».
32. Воспользовавшись корпусом, оцените, насколько конструкция «интересоваться о ком-либо» распространена, в каких текстах она встречается. Сравните ее частотность с частотностью конструкции «интересоваться кем-либо».
33. Найдите в Корпусе все случаи употребления сочетания "скучать по вас" и "скучать по вам". Сколько их? Какие примеры для каждой конструкции самые ранние? Какие самые поздние? Приведите эти примеры. Основываясь на этих данных, прокомментируйте ответ Справочной службы ИРЯ.
34. Посчитайте соотношение частоты употребления слов «этот» и «сей» в следующие периоды: а) 1770-1799 б) 1800-1829 в) 1830-1869
35. Когда в Корпусе появляются первые примеры глагола «реагировать»? Приведите первые два примера. Что они значат? Когда появляются первые примеры глагола «реагировать» с какой-либо приставкой? Приведите первые два примера.
36. Какая разметка из приведенных ниже является морфологической, а какая

1) \_\_\_\_\_

```

<?xml version="1.0" encoding="windows-1251" ?>
    <text>
        <p>
            <s>
                <w>Звонили<ana lemma="ЗВОНИТЬ" pos="Г" gram="мн,нс,нп,дст,прш," /></w>
                <w>к<ana lemma="К" pos="ПРЕДЛ" gram="" /></w>
            </s>
            <w>вечерне
                <ana lemma="ВЕЧЕРНЯ" pos="С" gram="жр,ед,дт,пр,но," />
                <ana lemma="ВЕЧЕРНИЙ" pos="П" gram="ср,ед,кр," /></w>
            <pun>.</pun> </s>
            <s><w>Торжественный<ana lemma="ТОРЖЕСТВЕННЫЙ" pos="П"
                gram="мр,ед,им,вн," /></w>
                <w>гул<ana lemma="ГУЛ" pos="С" gram="мр,ед,им,вн,но," /></w>
            <w>колоколов
                <ana lemma="КОЛОКОЛ" pos="С" gram="мр,мн,рд,но," />
                <ana lemma="КОЛОКОЛОВ" pos="С" gram="мр,фам,ед,им,од," /></w>
            .....<pun>.</pun> </s></p></text>

```

семантической?

2)

```
<?xml version="1.0" encoding="cp866"?>
  <МЕТА-ДОКУМЕНТ>
    <ДОКУМЕНТ>
<НАЗВАНИЕ>Сообщение газеты "Мир Кино"</НАЗВАНИЕ>
<ДАТА>20.03.2000</ДАТА>
    <УЧАСТНИК_ОБЩЕНИЯ Имя="Мир Кино" Тип="Автор">
      <СОЦИАЛЬНЫЙ_СТАТУС>
<ХАРАКТЕРИСТИКА>Печатное издание</ХАРАКТЕРИСТИКА>
      </СОЦИАЛЬНЫЙ_СТАТУС>
    </УЧАСТНИК_ОБЩЕНИЯ>
  </ДОКУМЕНТ>
```

35. Создайте морфологическую разметку предложения: Модуль семантического анализа текстов предназначен для нормализации синтаксической структуры, распознавания терминов, классификации терминов по семантическим признакам.
36. Объясните, что такое корпус? Что такое «Корпус как идеология» в представлении акад. Плунгяна.
37. Виды разметки в корпусе и средства их автоматизации.
38. Виды корпусов, их назначение. Приведите примеры 3-4 видов корпусов, дайте описание корпусных менеджеров (ИПС).
39. Способы использования корпусов в лингвистических и филологических исследованиях.
40. Национальный корпус русского языка: назначение, принципы создания, возможности поиска.

**Вопросы для группового устного и письменного опроса, включая рубежную письменную контрольную работу, коррелируют с темами лекционных занятий.**

**Примерные вопросы для группового устного и письменного опроса.**

**По разделу 1 (Сущность и метаязык корпусных исследований).**

6. Понятие лингвистического корпуса. История корпусной лингвистики. Корпусный (эмпирический) подход в сравнении с хомскианской лингвистикой. Задачи и основные направления корпусной лингвистики. Корпусная лингвистика и компьютерная лингвистика. Предмет исследования.
7. Определение лингвистического корпуса. Отличие корпуса от электронной библиотеки. История создания лингвистических корпусов. Основные понятия. Лингвистические и нелингвистические корпуса. Проблема репрезентативности. Развитие лингвистических корпусов в мире. Первое и второе поколение корпусов.
8. Экстралингвистическая метаразметка. Лингвистическая разметка. Автоматическая морфологическая и синтаксическая разметка. Снятие неоднозначности. Лемматизация. Графематическая, дискурсивная, семантическая разметка. Жанровая разметка и другие виды разметки. Процедуры лемматизации и разметка корпуса. Обзор автоматических средств разметки.
9. Языковые средства представления размеченных текстов. Международные стандарты и проекты (TEI, EAGLES, CDIF, XCES). Проблемы, возникающие при разметке текстов.
10. Разметка корпуса. Виды разметок. Создание корпуса в MOODLE. Метаразметка корпуса.

**По разделу 2 (Виды корпусов и основные методики работы с ними).**

6. Создание и типология корпусов. Типы корпусов: устные и письменные, одноязычные и многоязычные, аннотированные и неаннотированные.
7. Национальный корпус русского языка. Корпусы английского языка. Практическое

использование национальных корпусов. Проблемы и перспективы развития национальных корпусов.

8. Технология создания корпусов. Стадии работы
9. Создание и разметка корпуса на основе WIKI. Средства разметки корпуса. Анализ программных средств. Создание корпуса в MOODLE. Метаразметка корпуса. Возможности тегирования записей в корпусе.
10. Языки запросов корпусных менеджеров. Выходные интерфейсы.

**Критерии оценки (в баллах) для группового устного и письменного опроса:**

- 2 балла выставляется студенту, показавшему всесторонние, глубокие знания темы, свободное и правильное обоснование проблемных ситуаций;
- 1 балл выставляется студенту, показавшему не достаточно глубокие знания темы, не умеющему свободно и в полной мере аргументировано обосновать проблемную ситуацию;
- 0 баллов выставляется студенту, который не знает большей части основного содержания темы, допускает грубые ошибки в формулировках основных понятий и не способен решать типовые практические задачи или проблемные ситуации.

**Пример рубежной письменной контрольной работы по Разделу 1 (Модулю 1).**

12. Понятие лингвистического корпуса. История корпусной лингвистики. Корпусный (эмпирический) подход в сравнении с хомскианской лингвистикой.
13. Задачи и основные направления корпусной лингвистики. Корпусная лингвистика и компьютерная лингвистика. Предмет исследования.
14. Определение лингвистического корпуса.
15. Отличие корпуса от электронной библиотеки.
16. История создания лингвистических корпусов. Основные понятия. Лингвистические и нелингвистические корпуса.
17. Проблема репрезентативности.
18. Развитие лингвистических корпусов в мире. Первое и второе поколение корпусов.
19. Экстралингвистическая метаразметка.
20. Лингвистическая разметка.
21. Автоматическая морфологическая и синтаксическая разметка.
22. Снятие неоднозначности.

**Критерии оценки (в баллах) рубежной письменной контрольной работы:**

- 25 баллов выставляется студенту, показавшему всесторонние, систематизированные, глубокие знания учебной программы модуля и умение применять их на практике при решении конкретных задач, свободное и правильное обоснование проблемных ситуаций;
- 15 баллов выставляется студенту, показавшему не достаточно глубокое знание теории межкультурной коммуникации, не умеющему в полной мере аргументировано обосновать решение конкретных задач;
- 0 баллов выставляется студенту, который не знает большей части основного содержания учебной программы модуля, допускает грубые ошибки в формулировках основных понятий дисциплины и не умеет использовать полученные знания при решении типовых практических задач.

**Темы индивидуальных заданий студентов:**

15. Сравнительный анализ корпусов и корпусных менеджеров. Поиск в корпусах: конкордансер и корпусный менеджер как поисковая система. Поисковые системы Интернет как корпусные менеджеры.
16. Сравнение корпусных менеджеров с архитектурой поисковых систем в сети Интернет. Поисковые системы Интернет как корпусные менеджеры. Специализированные корпусные интерфейсы к индексам глобальных поисковых систем
17. Параллельные корпуса. Конкорданс. Многоязычие. Проблема выравнивания.

Примеры параллельных корпусов. Параллельные корпусы. Конкорданс. Анализ программных средств создания конкордансов и существующих параллельных корпусов.

18. Использование корпусов в обучении языкам и гуманитарных исследованиях. Лингвистические исследования на базе корпуса: изучение лексики. Лингвистические исследования на базе корпуса НКРЯ и VNC: изучение грамматики, исследование семантики (метафора, этимология).
19. Способы использования корпусов в лингвистических исследованиях.
20. Исследование способов использования корпусов в лексикографии.
21. Изучение средств обработки корпусных данных, представленных на языке XML.
22. Создание электронной хрестоматии по корпусной лингвистике.
23. Исследование механизмов взаимодействия корпуса текстов и электронной картотеки(корпусы цитат).
24. Создание веб-сайта по корпусной лингвистике.
25. Британский национальный корпус.
26. Национальный корпус русского языка: история создания, перспективы развития.
27. Технологии морфологической и синтаксической разметки НКРЯ.
28. Программа MyStem Yandex: разработка, использование, перспективы.

Индивидуальные задания выполняются в форме краткого устного сообщения (доклада) или медиапрезентации и оцениваются поощрительными баллами.

**Критерии оценки (в баллах) для индивидуальных заданий:**

- 5 баллов выставляется студенту, продемонстрировавшему самостоятельное, всестороннее, систематизированное, глубокое раскрытие темы, свободное владение материалом;
- 2 балла выставляется студенту, показавшему не достаточно самостоятельное и глубокое понимание темы, не умеющему свободно представить материал;
- 0 баллов выставляется студенту, который не понимает большей части основного содержания сообщения, не способен ясно, последовательно представить материал.