

МИНОБРНАУКИ РОССИИ



Федеральное государственное бюджетное образовательное учреждение  
высшего образования  
«Российский государственный гуманитарный университет»  
(ФГБОУ ВО «РГГУ»)

*ИНСТИТУТ ЛИНГВИСТИКИ*  
*УНЦ компьютерной лингвистики*

*Технологии корпусной лингвистики*

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ  
направление **45.03.03 Фундаментальная и прикладная лингвистика**  
профиль **Фундаментальная и прикладная лингвистика**  
Уровень квалификации выпускника (*бакалавр*)

Форма обучения (*очная*)

РПД адаптирована для лиц  
с ограниченными возможностями  
здоровья и инвалидов

Москва 2019

*Технологии корпусной лингвистики*

Рабочая программа дисциплины

Составитель:

к. филол. н., доцент УНЦ компьютерной лингвистики А.Ч. Пиперски

Ответственный редактор:

д. филол. н., проф. Я.Г. Тестелец

УТВЕРЖДЕНО

Протокол заседания УНЦ компьютерной лингвистики

№ 1 от 28.08.2019

## **ОГЛАВЛЕНИЕ**

### **1. Пояснительная записка**

1.1 Цель и задачи дисциплины

1.2. Формируемые компетенции, соотнесённые с планируемыми результатами обучения по дисциплине

1.3. Место дисциплины в структуре образовательной программы

### **2. Структура дисциплины**

### **3. Содержание дисциплины**

### **4. Образовательные технологии**

### **5. Оценка планируемых результатов обучения**

5.1. Система оценивания

5.2. Критерии выставления оценки по дисциплине

5.3. Типовые задания, используемые для текущего контроля успеваемости и промежуточной аттестации обучающихся по дисциплине

5.3.1. Примеры заданий к семинарам

5.3.2. Примеры статей для докладов

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

6.1. Список источников и литературы

6.2. Перечень электронных ресурсов

### **7. Материально-техническое обеспечение дисциплины**

### **8. Обеспечение образовательного процесса для лиц с ограниченными возможностями здоровья и инвалидов**

### **9. Методические материалы**

9.1. Планы семинарских занятий

## **Приложения**

Приложение 1. Аннотация дисциплины

Приложение 2. Лист изменений

## 1. Пояснительная записка

### 1.1. Цель и задачи дисциплины

**Цель дисциплины** – освоение студентами базовых понятий корпусной лингвистики.

#### **Задачи дисциплины:**

Курс нацелен на **формирование** у студентов следующих профессиональных **компетенций**:

владением основными понятиями и категориями современной лингвистики (ОПК-1);

владением основными методами фонологического, морфологического, синтаксического, дискурсивного и семантического анализа с учетом языковых и экстралингвистических факторов (ПК-1);

способностью пользоваться лингвистически ориентированными программными продуктами (ПК-9);

владение 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);

способность проводить квалифицированное тестирование лингвистически ориентированных программных продуктов, электронных ресурсов, лингвистически ориентированных систем и лингвистических компонентов интеллектуальных и информационных электронных систем (ПК-12).

### 1.2. Формируемые компетенции, соотнесённые с планируемыми результатами обучения по дисциплине.

В результате освоения дисциплины обучающийся должен:

#### **знать:**

- основные понятия и методы корпусной лингвистики
- устройство корпусов

#### **уметь:**

- анализировать различные уровни языковой структуры
- решать лингвистические задачи с помощью методов корпусной лингвистики

#### **владеть:**

- современной терминологией корпусной лингвистики
- методами решения лингвистических задач

Коды компетенции	Содержание компетенций	Перечень планируемых результатов обучения по дисциплине
ОПК-1	владение основными понятиями и категориями современной лингвистики	Уметь: решать лингвистические задачи с помощью методов корпусной лингвистики Владеть: методами решения лингвистических задач
ПК-1	владение основными методами	

ПК-9	фонологического, морфологического, синтаксического, дискурсивного и семантического анализа с учетом языковых и экстралингвистических факторов способностью пользоваться лингвистически ориентированными программными продуктами	Знать: устройство корпусов Владеть: современной терминологией корпусной лингвистики  Знать: основные понятия и методы корпусной лингвистики Уметь: анализировать различные уровни языковой структуры
ПК-10	владение принципами создания электронных языковых ресурсов (текстовых, речевых и мультимодальных корпусов; словарей, тезаурусов, онтологии; фонетических, лексических, грамматических и иных баз данных и баз знаний) и умением пользоваться такими ресурсами	
ПК-12	способность проводить квалифицированное тестирование лингвистически ориентированных программных продуктов, электронных ресурсов, лингвистически ориентированных систем и лингвистических компонентов интеллектуальных и информационных электронных систем	

### 1.3. Место дисциплины в структуре образовательной программы

Курс «Технологии корпусной лингвистики» относится к обязательной части блока дисциплин учебного плана по направлению подготовки 45.03.03 «Фундаментальная и прикладная лингвистика». Дисциплина адресована бакалаврам, обучающимся по направлению «Фундаментальная и прикладная лингвистика» (Б1.О.24) без направленности (профиля). Курс читается на 3-м курсе в 2-м семестре УНЦ компьютерной лингвистики ИЛ РГГУ, форма промежуточного контроля – зачёт.

Для успешного освоения материала студент должен опираться на знания, умения и навыки, полученные в рамках курсов «Введение в теорию языка», «Общая фонетика» «Общая морфология», «Общий синтаксис», «Общая семантика», «Понятийный аппарат математики», «Введение в компьютерную лингвистику», «Основы программирования» а также на навыки изучения научной литературы, сформированные при подготовке к другим теоретическим курсам.

Курс направлен на углубление знаний усовершенствование умений и навыков студентов в сфере общей лингвистики и созданию у студентов представления об основных методах и задачах корпусной лингвистики.

## 2. Структура дисциплины

Общая трудоёмкость дисциплины составляет 3 з.е., 108 ч., в том числе контактная работа обучающихся с преподавателем 42 ч., самостоятельная работа обучающихся 66 ч.

### Тематический календарный план курса

№ № раз дел а	Раздел курса	Семестр	Виды учебной деятельности и трудоёмкость (в часах)			Форма контроля успеваемости
			лекции	семинары	СРС	
1	Основные методы лингвистического исследования и место корпусной лингвистики среди них	1	2		2	Контроль посещаемости студентов.
2	Лингвистические корпуса: определение и примеры применения в лингвистических исследованиях	1	2	2	4	Контроль посещаемости студентов. Обсуждение прочитанной научной литературы.
3	Корпуса русского и английского языков (обзор)	1	4	8	16	Контроль посещаемости студентов. Обсуждение прочитанной научной литературы
4	Типы разметки в корпусах. Стандарты морфологической разметки для русского и английского языка. Омонимия и её разрешение	1	2	4	8	Контроль посещаемости студентов. Обсуждение прочитанной научной литературы.
5	Количественные исследования на корпусном материале. Базовые методы статистики в корпусных исследованиях.	1		2	4	
6	Нормирование частотности языковых единиц	1		2	6	

	в корпусах различного объёма. Частотные словари. Закон Ципфа.					
7	Исследование сочетаемости слов при помощи корпусов. Коллокации и меры их оценки. Лексические функции и их корпусное исследование.	1		2	4	
8	Проблема отбора текстов в корпус, репрезентативность и сбалансированность корпуса	1		6	10	
9	Создание пользовательских корпусов. Применение корпусных методов в различных областях лингвистики	1		6	10	
10	Зачёт	1			2	Итоговая контрольная работа.
	<b>Итого:</b>		<b>10</b>	<b>32</b>	<b>66</b>	

### 3. Содержание дисциплины

№	Наименование раздела дисциплины	Содержание
1	Основные методы лингвистического исследования и место корпусной лингвистики среди них	Интроспекция, эксперимент и наблюдение над реальностью. Место корпусной лингвистики в этом противопоставлении. Критика корпусной лингвистики со стороны Н. Хомского.
2	Лингвистические корпуса: определение и примеры применения в лингвистических исследованиях	Корпус как совокупность текстов, разметки и поиска. Применение корпусных методов для исследования морфологии, синтаксиса и семантики. Исследование
3	Корпуса русского и английского языков (обзор)	Основные корпуса русского и английского языков: Национальный корпус русского языка, ГИКРЯ, ruTenTen, Araneum Russicum; Brown Corpus, BNC, english-corpora.org
4	Типы разметки в корпусах. Стандарты морфологической разметки для русского и английского языка. Омонимия и её разрешение	Морфологическая разметка. Синтаксическая разметка. Прочие виды лингвистической разметки. Метаразметка
5	Количественные исследования на корпусном материале. Базовые методы статистики в корпусных исследованиях.	Таблица сопряжённости. Критерий хи-квадрат. Критерий Уилкоксона — Манна — Уитни.
6	Нормирование частотности языковых единиц в корпусах различного объёма. Частотные словари. Закон Ципфа.	Частотность на миллион. Зависимость частотности от ранга слова в частотном списке.
7	Исследование сочетаемости слов при помощи корпусов. Коллокации и меры их оценки. Лексические функции и их корпусное исследование.	Фразеологические сращения, единства и сочетания. Меры связанности коллокаций. Ожидаемая и наблюдаемая частота в корпусе. MI, z-score, t-score и др.
8	Проблема отбора текстов в корпус, репрезентативности и сбалансированности корпуса	Понятие сбалансированности и репрезентативности корпуса. Масштабируемость корпусных исследований.
9	Создание пользовательских корпусов. Применение корпусных методов в различных областях лингвистики	Оффлайн- и онлайн-конкордансеры. Создание специальных корпусов для различных исследовательских задач.



#### 4. Образовательные технологии

№ п/п	Наименование раздела	Виды учебных занятий	Образовательные технологии
1	Основные методы лингвистического исследования и место корпусной лингвистики среди них	Лекция 1.	Вводная лекция с использованием презентации
2	Лингвистические корпуса: определение и примеры применения в лингвистических исследованиях	Лекция 2	Лекция с использованием презентации
		Семинар 1	Развёрнутое обсуждение прочитанной научной литературы. Совместное выполнение заданий с помощью компьютера с доступом к сети «Интернет» и электронным ресурсам, в т.ч. текстовым корпусам.
3	Корпуса русского и английского языков (обзор)	Лекция 3	Лекция с использованием презентации
		Лекция 4	Лекция с использованием презентации
		Семинар 2	Совместное выполнение заданий с помощью компьютера с доступом к сети «Интернет» и электронным ресурсам, в т.ч. текстовым корпусам.
		Семинар 3	Совместное выполнение заданий с помощью компьютера с доступом к сети «Интернет» и электронным ресурсам, в т.ч. текстовым корпусам.
		Семинар 4	Совместное выполнение заданий с помощью компьютера с доступом к сети «Интернет» и электронным ресурсам, в т.ч. текстовым корпусам.
		Семинар 5	Совместное выполнение заданий с помощью компьютера с доступом к сети «Интернет» и электронным ресурсам, в т.ч. текстовым корпусам.
4	Типы разметки в корпусах. Стандарты морфологической разметки для русского и английского языка. Омонимия и её разрешение	Лекция 5	Лекция с использованием презентации
		Семинар 6	Индивидуальное выполнение заданий с помощью компьютера с доступом к сети «Интернет» и электронным ресурсам, в т.ч. текстовым корпусам, и совместное обсуждение после.
		Семинар 7	Индивидуальное выполнение заданий с помощью компьютера с доступом к сети «Интернет» и электронным ресурсам, в т.ч. текстовым корпусам, и

			совместное обсуждение после.
5	Количественные исследования на корпусном материале. Базовые методы статистики в корпусных исследованиях.	Семинар 8	Развёрнутое обсуждение прочитанной научной литературы. Индивидуальное выполнение заданий с помощью компьютера с доступом к сети «Интернет» и электронным ресурсам, в т.ч. текстовым корпусам, и совместное обсуждение после.
6	Нормирование частотности языковых единиц в корпусах различного объёма. Частотные словари. Закон Ципфа.	Семинар 9	Развёрнутое обсуждение прочитанной научной литературы. Индивидуальное выполнение заданий с помощью компьютера с доступом к сети «Интернет» и электронным ресурсам, в т.ч. текстовым корпусам, и совместное обсуждение после.
7	Исследование сочетаемости слов при помощи корпусов. Коллокации и меры их оценки. Лексические функции и их корпусное исследование.	Семинар 10	Развёрнутое обсуждение прочитанной научной литературы. Индивидуальное выполнение заданий с помощью компьютера с доступом к сети «Интернет» и электронным ресурсам, в т.ч. текстовым корпусам, и совместное обсуждение после.
8	Проблема отбора текстов в корпус, репрезентативности и сбалансированности корпуса	Семинар 11	Развёрнутое обсуждение прочитанной научной литературы. Индивидуальное выполнение заданий с помощью компьютера с доступом к сети «Интернет» и электронным ресурсам, в т.ч. текстовым корпусам, и совместное обсуждение после.
		Семинар 12	Развёрнутое обсуждение прочитанной научной литературы. Индивидуальное выполнение заданий с помощью компьютера с доступом к сети «Интернет» и электронным ресурсам, в т.ч. текстовым корпусам, и совместное обсуждение после.
		Семинар 13	Развёрнутое обсуждение прочитанной научной литературы. Индивидуальное

			выполнение заданий с помощью компьютера с доступом к сети «Интернет» и электронным ресурсам, в т.ч. текстовым корпусам, и совместное обсуждение после.
9	Создание пользовательских корпусов. Применение корпусных методов в различных областях лингвистики	Семинар 14	Развёрнутое обсуждение прочитанной научной литературы. Индивидуальное выполнение заданий с помощью компьютера с доступом к сети «Интернет» и электронным ресурсам, в т.ч. текстовым корпусам, и совместное обсуждение после.
		Семинар 15	Развёрнутое обсуждение прочитанной научной литературы. Индивидуальное выполнение заданий с помощью компьютера с доступом к сети «Интернет» и электронным ресурсам, в т.ч. текстовым корпусам, и совместное обсуждение после.
		Семинар 16	Развёрнутое обсуждение прочитанной научной литературы. Индивидуальное выполнение заданий с помощью компьютера с доступом к сети «Интернет» и электронным ресурсам, в т.ч. текстовым корпусам, и совместное обсуждение после.
	Зачёт		Защита и обсуждение докладов на ранее заданные темы.

## 5. Оценка планируемых результатов обучения

### 5.1. Система оценивания

Форма контроля	Макс. количество баллов	
	За одну работу	Всего
Текущий контроль:		
<ul style="list-style-type: none"> <li>Практические упражнения и задания, выполняемые на занятиях и в качестве домашней работы.</li> </ul>	5 баллов	40 баллов
<ul style="list-style-type: none"> <li>Доклады студентов по статьям из сборников [НКРЯ 2005, Плунгян и др. (ред.). 2009, Biber &amp; Reppen (eds.) 2011, O’Keeffe &amp; McCarthy</li> </ul>	5 баллов	30 баллов

(eds.) 2010]		
Промежуточная аттестация (зачёт)	30 баллов	30 баллов
<b>Итого за дисциплину</b>		100 баллов

Полученный совокупный результат конвертируется в традиционную шкалу оценок и в шкалу оценок Европейской системы переноса и накопления кредитов (European Credit Transfer System; далее – ECTS) в соответствии с таблицей:

100-балльная шкала	Традиционная шкала		Шкала ECTS
95 – 100	отлично	зачтено	A
83 – 94			B
68 – 82	хорошо		C
56 – 67	удовлетворительно		D
50 – 55			E
20 – 49	неудовлетворительно	не зачтено	FX
0 – 19			F

## 5.2. Критерии выставления оценки по дисциплине

<b>Баллы/ Шкала ECTS</b>	<b>Оценка по дисциплине</b>	<b>Критерии оценки результатов обучения по дисциплине</b>
100-83/ А,В	«отлично»/ «зачтено (отлично)»/ «зачтено»	<p>Выставляется обучающемуся, если он глубоко и прочно усвоил теоретический и практический материал, может продемонстрировать это на занятиях и в ходе промежуточной аттестации.</p> <p>Обучающийся исчерпывающе и логически стройно излагает учебный материал, умеет увязывать теорию с практикой, справляется с решением задач профессиональной направленности высокого уровня сложности, правильно обосновывает принятые решения.</p> <p>Свободно ориентируется в учебной и профессиональной литературе.</p> <p>Оценка по дисциплине выставляется обучающемуся с учётом результатов текущей и промежуточной аттестации.</p> <p>Компетенции, закреплённые за дисциплиной, сформированы на уровне – «высокий».</p>
82-68/ С	«хорошо»/ «зачтено (хорошо)»/ «зачтено»	<p>Выставляется обучающемуся, если он знает теоретический и практический материал, грамотно и по существу излагает его на занятиях и в ходе промежуточной аттестации, не допуская существенных неточностей.</p> <p>Обучающийся правильно применяет теоретические положения при решении практических задач профессиональной направленности разного уровня сложности, владеет необходимыми для этого навыками и приёмами.</p> <p>Достаточно хорошо ориентируется в учебной и профессиональной литературе.</p> <p>Оценка по дисциплине выставляется обучающемуся с учётом результатов текущей и промежуточной аттестации.</p> <p>Компетенции, закреплённые за дисциплиной, сформированы на уровне – «хороший».</p>
67-50/ D,E	«удовлетвори- тельно»/ «зачтено (удовлетвори- тельно)»/ «зачтено»	<p>Выставляется обучающемуся, если он знает на базовом уровне теоретический и практический материал, допускает отдельные ошибки при его изложении на занятиях и в ходе промежуточной аттестации.</p> <p>Обучающийся испытывает определённые затруднения</p>

Баллы/ Шкала ECTS	Оценка по дисциплине	Критерии оценки результатов обучения по дисциплине
	«зачтено»	<p>в применении теоретических положений при решении практических задач профессиональной направленности стандартного уровня сложности, владеет необходимыми для этого базовыми навыками и приёмами.</p> <p>Демонстрирует достаточный уровень знания учебной литературы по дисциплине.</p> <p>Оценка по дисциплине выставляются обучающемуся с учётом результатов текущей и промежуточной аттестации.</p> <p>Компетенции, закреплённые за дисциплиной, сформированы на уровне – «достаточный».</p>
49-0/ F,FX	«неудовлетворительно»/ не зачтено	<p>Выставляется обучающемуся, если он не знает на базовом уровне теоретический и практический материал, допускает грубые ошибки при его изложении на занятиях и в ходе промежуточной аттестации.</p> <p>Обучающийся испытывает серьёзные затруднения в применении теоретических положений при решении практических задач профессиональной направленности стандартного уровня сложности, не владеет необходимыми для этого навыками и приёмами.</p> <p>Демонстрирует фрагментарные знания учебной литературы по дисциплине.</p> <p>Оценка по дисциплине выставляются обучающемуся с учётом результатов текущей и промежуточной аттестации.</p> <p>Компетенции на уровне «достаточный», закреплённые за дисциплиной, не сформированы.</p>

### 5.3. Типовые задания, используемые для текущего контроля успеваемости и промежуточной аттестации обучающихся по дисциплине.

#### 5.3.1. Примеры заданий к семинарам

1. Суммарная Average Reduced Frequency (ARF) для всех слов в корпусе ..., чем суммарная обычная частотность всех слов в корпусе.

(А) меньше; (Б) не меньше; (В) больше; (Г) не больше

2. Какую из этих мер рекомендуется применять только с установлением нижнего порога частотности для коллокации?

(А) MI; (Б) t-мера; (В) z-мера; (Г) simple-II

3. Мера Average Reduced Frequency (ARF) теоретически может быть равна обычной частоте для слова, которое встречается в корпусе с частотой ...  
(А) 1; (Б) 2; (В) 100; (Г) с любой частотой
4. Сколько степеней свободы имеет таблица 3×3 с фиксированными суммами по строкам и по столбцам?  
(А) 0; (Б) 1; (В) 4; (Г) 9
5. Какое сочетание слов в русском языке будет предсказано униграммной моделью как наиболее частотное?  
(А) и и; (Б) и не; (В) а не; (Г) не а
6. Какое из этих значений не является стандартным порогом значимости, применяемым в статистических исследованиях?  
(А) 0,2; (Б) 0,05; (В) 0,01; (Г) 0,001
7. Корпуса с известной степенью сходства (Known-Similarity Corpora) используются для ...  
(А) выявления ключевых слов; (Б) оценки лексического разнообразия; (В) оценки мер сравнения корпусов; (Г) машинного перевода
8. Если сравнивать частоту двух слов в разных корпусах, какие результаты даёт критерий  $\chi^2$ ?  
(А) для высокочастотных слов результат почти всегда значим, а для низкочастотных почти всегда незначим; (Б) для высокочастотных слов результат обычно менее значим, чем для низкочастотных; (В) для высокочастотных слов результат почти всегда значим, и только для низкочастотных слов результат действительно зависит от наличия различий между корпусами; (Г) результат по критерию  $\chi^2$  не зависит от частоты слова
9. Если увеличить все значения в выборке в два раза, что произойдёт со стандартным отклонением?  
(А) не изменится; (Б) уменьшится; (В) увеличится; (Г) невозможно определить
10. Какое минимальное значение R в частотном словаре Ляшевской и Шарова может иметь слово, которое встретилось в корпусе, легшем в основу словаря, 50 раз?  
(А) 0; (Б) 1; (В) 50; (Г) 100
11. Слово w встретилось в корпусе объёмом N = 100000 токенов на 17-й, 548-й, 12319-й и 83500-й позиции. Вычислите ARF этого слова.
12. Каково минимальное и максимальное значение R для слов, которые в частотном словаре Ляшевской и Шарова имеют D=54?
13. С помощью критерия  $\chi^2$  оцените различие по частотности для слова «волна» в газетном и в поэтическом подкорпусе НКРЯ.

14. Используя корпус Araneum Russicum Minus ([http://unesco.uniba.sk/guest/run.cgi/first\\_form](http://unesco.uniba.sk/guest/run.cgi/first_form)), перечислите пять наиболее сильных по t-мере коллокатов, встречающихся непосредственно после слова «продать» (на уровне лемм).
15. Слово w встретилось в корпусе, разбитом на 10 частей, в общей сложности 40 раз. Приведите любое возможное распределение частотности этого слова по 10 сегментам корпуса, при котором D Жуйяна будет лежать в интервале [50;60].
16. При каких значениях n в формуле для вычисления %DIFF Адама Килгаррифа слово «system» окажется более характерным для British Academic Written English Corpus в сравнении с British Academic Spoken English Corpus, чем слово «strength»? (оба этих корпуса есть в SketchEngine)
17. В файле HP1and3.xls содержится частотный словарь для первой и третьей книги о Гарри Поттере. Приблизительно оцените, при каких значениях n в формуле для вычисления %DIFF Адама Килгаррифа в топ-10 слов, наиболее характерных для первой книги на фоне третьей, попадут не меньше четырёх нарицательных существительных.

### **5.3.2 Примеры статей для докладов**

Из сборников:

Киселёва, Ксения Л., Владимир А. Плунгян, Екатерина В. Рахилина, Сергей Г. Татевосов (ред.). 2009. Корпусные исследования по русской грамматике. М: Пробел–2000.

Biber, Douglas & Randi Reppen (eds.). 2011. *Corpus linguistics*. 4 vols. London: Sage.

McEnery, Tony & Andrew Hardie. 2012. *Corpus linguistics: Method, theory and practice*. Cambridge & New York: Cambridge University Press.

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

### **6.1. Список источников и литературы**

#### **Основной учебник**

Николаев И. С., Митренина О. В., Ландо Т. М. (ред.). 2016. Прикладная и компьютерная лингвистика. М.: URSS.

### **6.2. Перечень электронных ресурсов**

[ruscorpora.ru](http://ruscorpora.ru)

[English-corpora.org](http://English-corpora.org)

## **7. Материально-техническое обеспечение дисциплины**

Лекционные занятия по курсу «Технологии корпусной лингвистики» проводятся с использованием компьютерных презентаций, поэтому в аудитории необходимы



компьютер и проектор, а также соответствующее освещение. В ходе семинарских занятий студенты должны быть обеспечены компьютерами с лицензионным программным обеспечением и выходом в Интернет, т.е. занятия должны проходить в компьютерных классах. Для эффективной работы также необходима доска, размер и расположение которой должны позволять всем слушателям видеть написанное.

## **8. Обеспечение образовательного процесса для лиц с ограниченными возможностями здоровья и инвалидов**

В ходе реализации дисциплины используются следующие дополнительные методы обучения, текущего контроля успеваемости и промежуточной аттестации обучающихся в зависимости от их индивидуальных особенностей:

- для слепых и слабовидящих:
  - лекции оформляются в виде электронного документа, доступного с помощью компьютера со специализированным программным обеспечением;
  - письменные задания выполняются на компьютере со специализированным программным обеспечением, или могут быть заменены устным ответом;
  - обеспечивается индивидуальное равномерное освещение не менее 300 люкс;
  - для выполнения задания при необходимости предоставляется увеличивающее устройство; возможно также использование собственных увеличивающих устройств;
  - письменные задания оформляются увеличенным шрифтом;
  - экзамен и зачёт проводятся в устной форме или выполняются в письменной форме на компьютере.
- для глухих и слабослышащих:
  - лекции оформляются в виде электронного документа, либо предоставляется звукоусиливающая аппаратура индивидуального пользования;
  - письменные задания выполняются на компьютере в письменной форме;
  - экзамен и зачёт проводятся в письменной форме на компьютере; возможно проведение в форме тестирования.
- для лиц с нарушениями опорно-двигательного аппарата:
  - лекции оформляются в виде электронного документа, доступного с помощью компьютера со специализированным программным обеспечением;
  - письменные задания выполняются на компьютере со специализированным программным обеспечением;
  - экзамен и зачёт проводятся в устной форме или выполняются в письменной форме на компьютере.

При необходимости предусматривается увеличение времени для подготовки ответа.

Процедура проведения промежуточной аттестации для обучающихся устанавливается с учётом их индивидуальных психофизических особенностей. Промежуточная аттестация может проводиться в несколько этапов.

При проведении процедуры оценивания результатов обучения предусматривается использование технических средств, необходимых в связи с индивидуальными особенностями обучающихся. Эти средства могут быть предоставлены университетом, или могут использоваться собственные технические средства.

Проведение процедуры оценивания результатов обучения допускается с использованием дистанционных образовательных технологий.

Обеспечивается доступ к информационным и библиографическим ресурсам в сети Интернет для каждого обучающегося в формах, адаптированных к ограничениям их здоровья и восприятия информации:

- для слепых и слабовидящих:
  - в печатной форме увеличенным шрифтом;
  - в форме электронного документа;
  - в форме аудиофайла.
- для глухих и слабослышащих:
  - в печатной форме;
  - в форме электронного документа.
- для обучающихся с нарушениями опорно-двигательного аппарата:
  - в печатной форме;
  - в форме электронного документа;
  - в форме аудиофайла.

Учебные аудитории для всех видов контактной и самостоятельной работы, научная библиотека и иные помещения для обучения оснащены специальным оборудованием и учебными местами с техническими средствами обучения:

- для слепых и слабовидящих:
  - устройством для сканирования и чтения с камерой SARA CE;
  - дисплеем Брайля PAC Mate 20;
  - принтером Брайля EmBraille ViewPlus;
- для глухих и слабослышащих:
  - автоматизированным рабочим местом для людей с нарушением слуха и слабослышащих;
  - акустический усилитель и колонки;
- для обучающихся с нарушениями опорно-двигательного аппарата:
  - передвижными, регулируемые эргономическими партами СИ-1;
  - компьютерной техникой со специальным программным обеспечением.

## **9. Методические материалы**

### **9.1. Планы семинарских занятий**

#### **Семинар 1. Лингвистические корпуса: определение и примеры применения в лингвистических исследованиях (2 ч.).**

##### Вопросы для обсуждения:

Основные методы лингвистического исследования: интроспекция, эксперимент и наблюдение над реальностью. Место корпусной лингвистики в этом противопоставлении. Лингвистические корпуса: определение и примеры применения в лингвистических исследованиях.

##### Список литературы:

- Николаев И. С., Митренина О. В., Ландо Т.М. (ред.). 2016. *Прикладная и компьютерная лингвистика*. М.: URSS.
- Cheng, Winnie. 2012. *Exploring corpus linguistics: Language in action*. London & New York: Routledge.

#### Материально-техническое обеспечение занятия:

В ходе лабораторных работ студенты должны быть обеспечены компьютерами с лицензионным программным обеспечением и выходом в Интернет.

#### **Семинары 2–5. Корпуса русского и английского языков (обзор) (8 ч.).**

##### Вопросы для обсуждения:

Особенности устройства различных корпусов русского и английского языков. Практическая работа с корпусами.

1. Национальный корпус русского языка (НКРЯ)
2. ruWac
3. ruTenTen
4. Хельсинкский аннотированный корпус (ХАНКО)
5. Интегрум
6. Открытый корпус (OpenCorpora)
7. Генеральный Интернет-корпус русского языка (ГИКРЯ)
8. British National Corpus (BNC)
9. Corpus of Contemporary American English (COCA)
10. Corpus of Global Web-Based English (GloWbe)
11. Brown Corpus
12. Google Books: Google Ngrams Viewer и поисковый интерфейс на сайте Brigham Young University

##### Список литературы:

- Николаев И. С., Митренина О. В., Ландо Т.М. (ред.). 2016. *Прикладная и компьютерная лингвистика*. М.: URSS.
- McEnery, Tony & Andrew Hardie. 2012. *Corpus linguistics: Method, theory and practice*. Cambridge & New York: Cambridge University Press.

#### Материально-техническое обеспечение занятия:

В ходе лабораторных работ студенты должны быть обеспечены компьютерами с лицензионным программным обеспечением и выходом в Интернет.

#### **Семинары 6–7. Типы разметки в корпусах. Стандарты морфологической разметки для русского и английского языка. Омонимия и её разрешение (обзор) (4 ч.).**

##### Вопросы для обсуждения:

Стандарты морфологической разметки для русского и английского языка. Омонимия и её разрешение. Основные морфоанализаторы и синтаксические парсеры для русского и английского языка.

Список литературы:

- Николаев И. С., Митренина О. В., Ландо Т.М. (ред.). 2016. *Прикладная и компьютерная лингвистика*. М.: URSS.
- McEnery, Tony & Andrew Hardie. 2012. *Corpus linguistics: Method, theory and practice*. Cambridge & New York: Cambridge University Press.

Материально-техническое обеспечение занятия:

В ходе лабораторных работ студенты должны быть обеспечены компьютерами с лицензионным программным обеспечением и выходом в Интернет.

**Семинар 8. Количественные исследования на корпусном материале. Базовые методы статистики в корпусных исследованиях (2 ч.).**

Вопросы для обсуждения:

Противопоставление качественных и количественных исследований. Анализ частотности с помощью корпусов и преимущества данного метода по сравнению с другими способами получения количественных данных о языковых единицах.

Список литературы:

- Николаев И. С., Митренина О. В., Ландо Т.М. (ред.). 2016. *Прикладная и компьютерная лингвистика*. М.: URSS.
- Biber, Douglas & Randi Reppen (eds.). 2011. *Corpus linguistics*. 4 vols. London: Sage.

Материально-техническое обеспечение занятия:

В ходе лабораторных работ студенты должны быть обеспечены компьютерами с лицензионным программным обеспечением и выходом в Интернет.

**Семинар 9. Нормирование частотности языковых единиц в корпусах различного объёма. Частотные словари. Закон Ципфа. (2 ч.).**

Вопросы для обсуждения:

Проценты, промилле и вхождения на миллион. Сопоставление корпусов разных объёмов. Основные частотные словари русского языка. Особенности различных мер частотности, применяемых в частотных словарях. Зависимость частотности языковых единиц от их места в частотном списке. Проблема «длинного хвоста» — многочисленных редко встречающихся слов.

#### Список литературы:

- Николаев И. С., Митренина О. В., Ландо Т.М. (ред.). 2016. *Прикладная и компьютерная лингвистика*. М.: URSS.
- Jurafsky, Dan & James H. Martin. 2017. *Speech and language processing: An introduction to natural language processing, computational linguistics, and speech recognition*. 3rd edition draft.

#### Материально-техническое обеспечение занятия:

В ходе лабораторных работ студенты должны быть обеспечены компьютерами с лицензионным программным обеспечением и выходом в Интернет.

### **Семинар 10. Исследование сочетаемости слов при помощи корпусов. Коллокации и меры их оценки. Лексические функции и их корпусное исследование. (2 ч.)**

#### Вопросы для обсуждения:

Фразеологические сращения, единства и сочетания. Меры связанности коллокаций. Ожидаемая и наблюдаемая частота в корпусе. MI, z-score, t-score, logDice и другие меры. Содержательная интерпретация автоматически извлечённых коллокаций.

#### Список литературы:

- Николаев И. С., Митренина О. В., Ландо Т.М. (ред.). 2016. *Прикладная и компьютерная лингвистика*. М.: URSS.

#### Материально-техническое обеспечение занятия:

В ходе лабораторных работ студенты должны быть обеспечены компьютерами с лицензионным программным обеспечением и выходом в Интернет.

### **Семинары 11–13. Проблема отбора текстов в корпус, репрезентативности и сбалансированности корпуса (6 ч.)**

#### Вопросы для обсуждения:

Понятие сбалансированности и репрезентативности корпуса. Масштабируемость корпусных исследований. Оценка основных корпусов русского и английского языков с точки зрения репрезентативности и сбалансированности.

#### Список литературы:

- Николаев И. С., Митренина О. В., Ландо Т.М. (ред.). 2016. *Прикладная и компьютерная лингвистика*. М.: URSS.
- Копотев, Михаил. 2014. *Введение в корпусную лингвистику*. Praha: Animedia Company.

#### Материально-техническое обеспечение занятия:

В ходе лабораторных работ студенты должны быть обеспечены компьютерами с лицензионным программным обеспечением и выходом в Интернет.

### **Семинары 14–16. Создание пользовательских корпусов. Применение корпусных методов в различных областях лингвистики (6 ч.)**

#### Вопросы для обсуждения:

Оффлайн- и онлайн-конкордансеры. Создание специальных корпусов для различных исследовательских задач. Использование базовых навыков программирования при разработке собственных корпусов.

#### Список литературы:

- Николаев И. С., Митренина О. В., Ландо Т.М. (ред.). 2016. *Прикладная и компьютерная лингвистика*. М.: URSS.
- Копотев, Михаил. 2014. *Введение в корпусную лингвистику*. Praha: Animedia Company.

#### Материально-техническое обеспечение занятия:

В ходе лабораторных работ студенты должны быть обеспечены компьютерами с лицензионным программным обеспечением и выходом в Интернет.

## АННОТАЦИЯ ДИСЦИПЛИНЫ

Курс «Технологии корпусной лингвистики» относится к обязательной части блока дисциплин учебного плана по направлению подготовки 45.03.03 «Фундаментальная и прикладная лингвистика». Дисциплина адресована бакалаврам, обучающимся по направлению «Фундаментальная и прикладная лингвистика» (Б1.О.24) без направленности (профиля). Курс читается на 3-м курсе в 1-м семестре УНЦ компьютерной лингвистики ИЛ РГГУ, форма промежуточного контроля – зачёт.

Дисциплина реализуется в учебно-научном центре компьютерной лингвистики.

**Цель дисциплины** – освоение студентами базовых понятий корпусной лингвистики.

### **Задачи дисциплины:**

Курс нацелен на **формирование** у студентов следующих профессиональных **компетенций**:

владением основными понятиями и категориями современной лингвистики (ОПК-1);

владением основными методами фонологического, морфологического, синтаксического, дискурсивного и семантического анализа с учетом языковых и экстралингвистических факторов (ПК-1);

способностью пользоваться лингвистически ориентированными программными продуктами (ПК-9);

владение 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);

способностью проводить квалифицированное тестирование лингвистически ориентированных программных продуктов, электронных ресурсов, лингвистически ориентированных систем и лингвистических компонентов интеллектуальных и информационных электронных систем (ПК-12).

В результате освоения дисциплины обучающийся должен:

#### **знать:**

- основные понятия и методы корпусной лингвистики
- устройство корпусов

#### **уметь:**

- анализировать различные уровни языковой структуры
- решать лингвистические задачи с помощью методов корпусной лингвистики

#### **владеть:**

- современной терминологией корпусной лингвистики
- методами решения лингвистических задач

По дисциплине предусмотрена промежуточная аттестация в форме зачёта.

Общая трудоемкость освоения дисциплины составляет 3 зачетных единицы.



## ЛИСТ ИЗМЕНЕНИЙ

№	Текст актуализации или прилагаемый к РПД документ, содержащий изменения	Дата	№ протокола
1	<i>Обновлена основная и дополнительная литература</i>	28.05.2017	8
2	Приложение №1		
3	<i>Обновлена основная и дополнительная литература</i>	23.06.2018	9
4	Приложение №2		

**Состав программного обеспечения (ПО), современных профессиональных баз данных (БД) и информационно-справочных систем (ИСС) (2019 г.)**

**1. Перечень ПО**

*Таблица 1*

№п/п	Наименование ПО	Производитель	Способ распространения ( <i>лицензионное или свободно распространяемое</i> )
1	Microsoft Office 2010	Microsoft	лицензионное
2	Windows 7 Pro	Microsoft	лицензионное
3	Kaspersky Endpoint Security	Kaspersky	лицензионное

**2. Перечень БД и ИСС**

*Таблица 2*

№п/п	Наименование
	Международные реферативные наукометрические БД, доступные в рамках национальной подписки в 2017 г. Web of Science Scopus
	Профессиональные полнотекстовые БД, доступные в рамках национальной подписки в 2017 г. Журналы Oxford University Press ProQuest Dissertation & Theses Global SAGE Journals Журналы Taylor and Francis
	Профессиональные полнотекстовые БД JSTOR Издания по общественным и гуманитарным наукам
	Компьютерные справочные правовые системы Консультант Плюс, Гарант

**1. Образовательные технологии (к п.4 на 2019 г.)**

В период временного приостановления посещения обучающимися помещений и территории РГГУ. для организации учебного процесса с применением электронного обучения и дистанционных образовательных технологий могут быть использованы следующие образовательные технологии:

– электронные учебники, учебные пособия, научные издания в электронном виде и доступ к иным электронным образовательным ресурсам;

- системы для электронного тестирования;
- консультации с использованием телекоммуникационных средств.

2	Профессиональные полнотекстовые БД, доступные в рамках национальной подписки в 2019 г. Журналы Cambridge University Press ProQuest Dissertation & Theses Global SAGE Journals Журналы Taylor and Francis
3	Профессиональные полнотекстовые БД JSTOR Издания по общественным и гуманитарным наукам Электронная библиотека Grebennikon.ru
4	Компьютерные справочные правовые системы Консультант Плюс, Гарант