

МИНОБРНАУКИ РОССИИ



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

ИНСТИТУТ ЛИНГВИСТИКИ  
Учебно-научный центр компьютерной лингвистики

**Корпусная лингвистика: построение и использование корпусов**

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

---

45.04.03 Фундаментальная и прикладная лингвистика

*Код и наименование направления подготовки/специальности*

---

Фундаментальная и компьютерная лингвистика

*Наименование направленности (профиля)/ специализации*

Уровень высшего образования: магистратура

Форма обучения: очная

РПД адаптирована для лиц  
с ограниченными возможностями  
здоровья и инвалидов

Москва 2023

**Корпусная лингвистика: построение и использование корпусов**  
Рабочая программа дисциплины

Составитель(и):

к.ф.н., доцент А.Ч.Пиперски

Ответственный редактор:

д.ф.н, профессор Н.А.Коротаев

УТВЕРЖДЕНО

Протокол заседания УНЦ компьютерной лингвистики  
№ 6 от 12 апреля 2023 г.

## ОГЛАВЛЕНИЕ

1.	Пояснительная записка .....	
1.1.	Цель и задачи дисциплины .....	
1.2.	Перечень планируемых результатов обучения по дисциплине, соотнесенных с индикаторами достижения компетенций .....	
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	Оценочные средства (материалы) для текущего контроля успеваемости, промежуточной аттестации обучающихся по дисциплине .....	
6.	Учебно-методическое и информационное обеспечение дисциплины .....	
6.1	Список источников и литературы .....	
6.2	Перечень ресурсов информационно-телекоммуникационной сети «Интернет». ....	
6.3	Профессиональные базы данных и информационно-справочные системы .....	
7.	Материально-техническое обеспечение дисциплины .....	
8.	Обеспечение образовательного процесса для лиц с ограниченными возможностями здоровья и инвалидов .....	
9.	Методические материалы .....	
9.1	Планы семинарских/ практических/ лабораторных занятий .....	
9.2	Методические рекомендации по подготовке письменных работ .....	
9.3	Иные материалы .....	

## 1. Пояснительная записка

### 1.1. Цель и задачи дисциплины

Предметом дисциплины является изучение методов создания корпусов и других электронных, специальных методов сбора материала с использованием корпусов ресурсов для лингвистических исследований грамматики и лексики, методов корпусных исследований.

Курс направлен на решение следующих задач:

- познакомить магистрантов с последними трендами в области корпусной лингвистики, с основными типами лингвистических ресурсов, доступными в Интернете и используемыми в профессиональной деятельности: с наиболее актуальными лингвистическими корпусами (национальными и проблемными), специальными программами обработки текстов, лексикографическими ресурсами и т.п.;
- на примерах показать, какие новые возможности в исследовании грамматики и лексики языка дает использование корпусных, а также применение современных методов обработки этих данных;
- дать представления о проблемах создания корпусов, об основных принципах разработки данных ресурсов и об основных требованиях, предъявляемых к ним;
- познакомить с технологиями и проблемами разметки корпусов;
- научить работать с современными пакетами обработки собственных корпусов;
- научить применять специальные методы лингвистических исследований, использующие данные корпусов, в том числе и статистические методы исследования;
- обучить практическим навыкам по применению корпусных методов в своей исследовательской работе.

### 1.2. Перечень планируемых результатов обучения по дисциплине, соотнесенных с индикаторами достижения компетенций

Компетенция	Индикаторы компетенций	Результаты обучения
ОПК-1 Способен решать профессиональные задачи, применяя основные понятия, категории и положения лингвистических теорий, и актуальные концепции в области лингвистики	ОПК-1.1 Знает основные моменты истории лингвистики; имеет общее представление о разнообразии лингвистических парадигм, о современном состоянии языкознания, о наиболее значимых лингвистических гипотезах нашего времени	<p>Знать:</p> <ul style="list-style-type: none"> <li>– рекомендованные преподавателем труды по изучаемым вопросам;</li> <li>– фонетические, грамматические и словообразовательные явления;</li> <li>– основные моменты истории лингвистики;</li> </ul> <p>Уметь:</p> <ul style="list-style-type: none"> <li>– применять общетеоретические положения лингвистики к конкретным фактам языка;</li> </ul> <p>Владеть:</p> <ul style="list-style-type: none"> <li>– навыками решения профессиональных задач с применением основных</li> </ul>

		теоретических знаний лингвистики.
ОПК-6 Способен осуществлять эффективное управление разработкой программных средств и информационных проектов в сфере своей профессиональной деятельности	ОПК-6.1 Знает основы командной и управленческой деятельности; принципы организации эффективного взаимодействия при разработке программных средств и компьютерных ресурсов с лингвистической составляющей	<p>Знать:</p> <ul style="list-style-type: none"> <li>– основы командной и управленческой деятельности;</li> <li>– принципы организации эффективного взаимодействия при разработке программных средств и компьютерных ресурсов с лингвистической составляющей;</li> </ul> <p>Уметь:</p> <ul style="list-style-type: none"> <li>– осуществлять управление разработкой программных средств и информационных проектов в сфере лингвистики;</li> </ul> <p>Владеть:</p> <ul style="list-style-type: none"> <li>– навыками командной разработки программных средств и информационных проектов.</li> </ul>

### 1.3. Место дисциплины в структуре образовательной программы

Дисциплина «Корпусная лингвистика: построение и использование корпусов» относится к обязательной части блока дисциплин учебного плана.

В результате освоения дисциплины формируются знания, умения и владения, необходимые для изучения следующих дисциплин и прохождения практик: Русская корпусная грамматика.

### 2. Структура дисциплины

Общая трудоёмкость дисциплины составляет 3 з.е., 108 академических часов.

#### Структура дисциплины для очной формы обучения

Объем дисциплины в форме контактной работы обучающихся с педагогическими работниками и (или) лицами, привлекаемыми к реализации образовательной программы на иных условиях, при проведении учебных занятий:

Семестр	Тип учебных занятий	Количество часов
1	Лекции	6
1	Практические занятия	24
Всего:		30

Объем дисциплины (модуля) в форме самостоятельной работы обучающихся составляет 78 академических часов.

### 3. Содержание дисциплины

№ п/п	Наименование раздела дисциплины	Содержание
1.	Введение. Общее представление о корпусах и корпусной лингвистике. Стандарты разметки. Типы разметки корпусов.	Краткая история корпусной лингвистики. Преимущества современных корпусных исследований. Возможность объединения формального и эмпирического подхода в современной корпусной лингвистике. Компьютерные ресурсы, необходимые лингвистам для решения различных задач. Задачи, решаемые с помощью компьютерных ресурсов. ДЗ1. Упр. по теме “Особенности поиска и управления выдачи в НКРЯ”.
2.	Проблемные корпуса (параллельные, диалектные, мультимедийные и др.)	Корпус. Национальный корпус. Проблемный корпус. Основные единицы. Основные требования, предъявляемые к корпусу. ДЗ2. Практическая работа. Разметка мультимедийного / аудиокорпуса
3.	Поиск в корпусе. Использование языка SQP для поиска в корпусе. Составление сложных запросов к корпусу.	Поиск в корпусе. Основные требования и параметры поиска. ДЗ3. Запросы для поиска сложных конструкций с использованием языка регулярных выражений: письменный отчет
4.	Особенности различных типов разметки. Морфологическая разметка	Стандарты и типы разметки. ДЗ4. Практическая работа. Тестирование морфологической или синтаксической разметки: письменный отчет
5.	Особенности различных типов разметки. Синтаксическая разметка	Стандарты и типы разметки. ДЗ5. Практическая работа. Тестирование морфологической или синтаксической разметки: письменный отчет
6.	Особенности разметки: другие типы разметки	Стандарты и типы разметки. ДЗ6. Практическая работа. Семантическая разметка / анафорическая разметка
7.	Методы корпусных исследований. Анализ примеров корпусных исследований	Основные принципы и методы разметки корпусов. Современные технологии разметки корпусов. ДЗ7. Реферирование статьи из рекомендованного списка. Составление краткого ТЗ к собственному исследовательскому проекту.
8.	Инструменты разметки собственного исследовательского корпуса	Типы программ обработки текста, методы работы с программами обработки текста. ДЗ8. Разработка параметров и схемы разметки исследовательского корпуса, разметка корпуса с помощью специальной программы

9.	Составление конкордансов, частотных списков, списков коллокаций с использованием специальных программ	Программы разметки собственных исследовательских корпусов Д39. Обработка данных корпуса с использованием специальных программ
10.	Статистические методы исследований	Области использования корпусных данных
11.	Основные статистические характеристики данных. Меры средней тенденции и изменчивости, коэффициент вариации и др.	Методы сбора и статистической обработки корпусных данных. Д310. Выполнение упр. по общему статистическому описанию данных. Первичный статистический анализ данных по проекту
12.	Статистические критерии зависимости	Общие статистические характеристики: меры средней тенденции и изменчивости. Проверка статистических критериев, исследование зависимостей. Д311. Выполнение упр. по использованию статистических критериев. Проверка статистической гипотезы в рамках исследовательского проекта
13.	Методы корреляционного анализа	Корреляционный анализ. Д312. Упр. по теме «Корреляционный анализ»
14.	Методы корпусной лексикографии.	Примеры использования корпусов в обучении и научных исследованиях: методология создания дидактических материалов с использованием корпусов; методология создания исследовательского корпуса с использованием корпусов общего назначения. Использование корпусов в лексикографической работе. Статистические методы в лексикографии Д313. Практическое задание по теме «Статистические корпусные методы в лексикографии»
15.	Сравнение двух корпусов. Стилеметрия	Примеры корпусных диалектологических, диахронических, социолингвистических и гендерных исследований, исследований стиля. Сравнение корпусов. Стилеметрия. Д314. Анализ частотных словарей, сравнение словарей двух корпусов
16.	Корпусные исследования в грамматике. Пример. Корпусное исследование конструкции	Примеры применения корпусного анализа в грамматических исследованиях. Специальные лексикографические и грамматические ресурсы, созданные на основе корпусов, включая FrameNet, PropBank и т.д. Д315. Реферирование статьи из рекомендованного списка.

#### 4. Образовательные технологии

Для проведения учебных занятий по дисциплине используются различные образовательные технологии. Для организации учебного процесса может быть использовано электронное обучение и (или) дистанционные образовательные технологии.

## 5. Оценка планируемых результатов обучения

### 5.1 Система оценивания

Форма контроля	Макс. количество баллов	
	За одну работу	Всего
Текущий контроль:		
- домашние задания	5 баллов	30 баллов
- выполнение заданий на семинаре	5 баллов	10 баллов
- участие в соревновании	20 баллов	20 баллов
Промежуточная аттестация – зачет		40 баллов
<b>Итого за семестр</b>		<b>100 баллов</b>

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			C
56 – 67	удовлетворительно	зачтено	D
50 – 55			E
20 – 49	неудовлетворительно	не зачтено	FX
0 – 19			F

### 5.2 Критерии выставления оценки по дисциплине

Баллы/ Шкала ECTS	Оценка по дисциплине	Критерии оценки результатов обучения по дисциплине
100-83/ A,B	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>

Баллы/ Шкала ECTS	Оценка по дисциплине	Критерии оценки результатов обучения по дисциплине
82-68/ С	хорошо/ зачтено	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. Обучающийся правильно применяет теоретические положения при решении практических задач профессиональной направленности разного уровня сложности, владеет необходимыми для этого навыками и приёмами. Достаточно хорошо ориентируется в учебной и профессиональной литературе. Оценка по дисциплине выставляется обучающемуся с учётом результатов текущей и промежуточной аттестации. Компетенции, закреплённые за дисциплиной, сформированы на уровне – «хороший».
67-50/ D,E	удовлетво- рительно/ зачтено	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. Обучающийся испытывает определённые затруднения в применении теоретических положений при решении практических задач профессиональной направленности стандартного уровня сложности, владеет необходимыми для этого базовыми навыками и приёмами. Демонстрирует достаточный уровень знания учебной литературы по дисциплине. Оценка по дисциплине выставляется обучающемуся с учётом результатов текущей и промежуточной аттестации. Компетенции, закреплённые за дисциплиной, сформированы на уровне – «достаточный».
49-0/ F,FX	неудовлет- ворительно/ не зачтено	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. 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. Демонстрирует фрагментарные знания учебной литературы по дисциплине. Оценка по дисциплине выставляется обучающемуся с учётом результатов текущей и промежуточной аттестации. Компетенции на уровне «достаточный», закреплённые за дисциплиной, не сформированы.

### 5.3 Оценочные средства (материалы) для текущего контроля успеваемости, промежуточной аттестации обучающихся по дисциплине

В качестве домашних заданий предлагаются задания следующих типов

Д31. Упр. по теме “Особенности поиска и управления выдачи в НКРЯ”.

Д32. Практическая работа. Разметка мультимедийного / аудиокорпуса.

Д33. Запросы для поиска сложных конструкций с использованием языка регулярных выражений: письменный отчет.

Д34. Практическая работа. Тестирование морфологической или синтаксической разметки: письменный отчет.

Д35. Практическая работа. Тестирование морфологической или синтаксической разметки: письменный отчет.

Д36. Практическая работа. Семантическая разметка / анафорическая разметка.

Контрольная работа по теме: «Корпуса и лингвистические ресурсы».

Д37. Реферирование статьи из рекомендованного списка. Составление краткого ТЗ к собственному исследовательскому проекту.

Д38. Разработка параметров и схемы разметки исследовательского корпуса, разметка корпуса с помощью специальной программы.

- Д39. Обработка данных корпуса с использованием специальных программ.  
 Д310. Выполнение упр. по общему статистическому описанию данных. Первичный статистический анализ данных по проекту.  
 Д311. Выполнение упр. по использованию статистических критериев. Проверка статистической гипотезы в рамках исследовательского проекта.  
 Д312. Упр. по теме «Корреляционный анализ».  
 Д313. Практическое задание по теме «Статистические корпусные методы в лексикографии».  
 Д314. Анализ частотных словарей, сравнение словарей двух корпусов.  
 Д315. Реферирование статьи из рекомендованного списка.

Зачет ориентирован на следующие контрольные вопросы

- Определение корпуса. Задачи, решаемые с использованием корпусов.
- Параметры поиска в корпусе.
- Требования к корпусу, предъявляемые пользователем.
- Характеристики основных корпусов.
- Типы проблемных корпусов.
- Характеристика лингвистических ресурсов одного из типов: словари, тезаурусы, специальные лексикографические ресурсы: фреймбанки и т.п., лингвистические базы данных.
- Особенности проблемных корпусов.
- Виды разметки в корпусе.
- Корпуса в психо- и социолингвистических исследованиях

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

**6.1 Список источников и литературы**

Основная литература

1. *Баранов А.Н.* Корпусная лингвистика // Баранов А.Н. Введение в прикладную лингвистику. М., 2003. С. 112–137.
2. *Ермолаев О.Ю.* Математическая статистика для психологов. 2-е изд., испр. - М.: МПСИ, Флинта, 2003 - 336 с. С. 10-64, с. 101-175, с. 202-222
3. *Захаров В.П.* Корпусная лингвистика: Учебно-метод. пособие. – СПб., 2005. – 48 с. Доступно для скачивания по адресу: <http://vp-zakharov.narod.ru/publications.html>
4. *Захаров В.П., Хохлова М.В.* Анализ эффективности статистических методов выявления коллокаций в текстах на русском языке // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог» 2010. Выпуск 9 (16). <http://www.dialog-21.ru/dialog2010/materials/pdf/22.pdf>
5. *Кустова Г. И., Ляшевская О. Н., Падучева Е. В., Рахилина Е. В.* Семантическая разметка лексики в Национальном корпусе русского языка: принципы, проблемы, перспективы. // Национальный корпус русского языка: 2003-2005. Результаты и перспективы. - М., 2005. С. 155-174.
6. *Ляшевская О. Н., Плунгян В. А., Сичинава Д. В.* О морфологическом стандарте Национального корпуса русского языка. // Национальный корпус русского языка: 2003-2005. Результаты и перспективы. - М., 2005. С. 111-135.
7. *Ляшевская О.Н., Кузнецова Ю.Л.* Русский фреймнет: к задаче создания корпусного словаря конструкций // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог 2009» (Бекасово, 27-31 мая 2009 г.). Вып. 8 (15). М.: РГГУ, 2009. С. 306-312. [pdf](#)

8. *Ляшевская, О. Н., Шаров С. А.* Частотный словарь современного русского языка : на материалах Национального корпуса русского языка. М. : Азбуковник, 2009. XV, 1074 с. (1)
9. *Резникова Т. И.* Обзор общедоступных корпусов русского языка в Интернете. // *НТИ*, 2005, серия 2. №3. С. 2-9. (доступна с [http://catalog.viniti.ru/search\\_extra.asp](http://catalog.viniti.ru/search_extra.asp))
10. *Савчук С.О.* Метатекстовая разметка в Национальном корпусе русского языка: базовые принципы и основные функции // Национальный корпус русского языка: 2003-2005. Результаты и перспективы. - М., 2005. С. 62-88
11. *Chris Manning* and *Hinrich Schütze*, Chapter 5. *Collocations*. Foundations of Statistical Natural Language Processing, MIT Press. Cambridge, MA: May 1999. Доступна с сайта [http://reslib.com/book/Foundations\\_of\\_Statistical\\_Natural\\_Language\\_Processing](http://reslib.com/book/Foundations_of_Statistical_Natural_Language_Processing) или <http://nlp.stanford.edu/fsnlp/promo/colloc.pdf>

#### Рекомендованная литература

1. *Андрющенко В.М.* Концепция и архитектура машинного фонда русского языка // Отв. ред.
2. *Баранов А.Н.* Корпусная лингвистика // Баранов А.Н. Введение в прикладную лингвистику. М., 2003. С. 112–137.
3. *Гмурман В. Е.* *Руководство к решению задач по теории вероятностей и математической статистике.* 9-е изд., стер.—М.: *Высшая школа*, 2004.— 404 с.
4. *Добрушина Н. Р.* [Как использовать Национальный корпус русского языка в образовании?](http://ruscorpora.ru/sbornik2005/18dobrushina.pdf) // Национальный корпус русского языка: 2003—2005. М.: Индрик, 2005. <http://ruscorpora.ru/sbornik2005/18dobrushina.pdf>
5. Доклады научной конференции «Корпусная лингвистика и лингвистические базы данных» / Под ред. А.С. Герда. СПб., 2002.
6. *Ермолаев О.Ю.* Математическая статистика для психологов. 2-е изд., испр. - М.: МПСИ, Флинта, 2003 - 336 с. С. 10-64, с. 101-175, с. 202-222
7. *Захаров В.П.* Корпусная лингвистика. Слайды к курсу лекций. (доступны по адресам <http://company.yandex.ru/academic/class/courses/zakharov.xml>, <http://teormin.ifmo.ru/education/corpus-linguistics/>)
8. *Захаров В.П.* Корпусная лингвистика: Учебно-метод. пособие. – СПб., 2005. – 48 с. Доступно для скачивания по адресу: <http://vp-zakharov.narod.ru/publications.htm>
9. *Захаров В.П., Хохлова М.В.* Анализ эффективности статистических методов выявления коллокаций в текстах на русском языке // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог» 2010. Выпуск 9 (16). <http://www.dialog-21.ru/dialog2010/materials/pdf/22.pdf>
10. *Зубов А.В.* Информационные технологии в лингвистике: Учебное пособие. - М.: Издательский центр «Академия», 2004. - 208 с.
11. Инструментарий русистики: корпусные подходы. Slavica Helsingiensia Series. Вып. 34. Helsinki 2008.
12. *Коваль С.А.* Роль корпуса в создании реалистичных моделей словоизменительной морфологии. URL: [http://skowal.narod.ru/research/corpora2006/Koval\\_Corpora.2006.htm](http://skowal.narod.ru/research/corpora2006/Koval_Corpora.2006.htm)
13. *Кустова Г. И., Ляшевская О. Н., Падучева Е. В., Рахилина Е. В.* Семантическая разметка лексики в Национальном корпусе русского языка: принципы, проблемы, перспективы. // Национальный корпус русского языка: 2003-2005. Результаты и перспективы. - М., 2005. С. 155-174.
14. *Ляшевская О. Н., Плунгян В. А., Сичинава Д. В.* О морфологическом стандарте Национального корпуса русского языка. // Национальный корпус русского языка: 2003-2005. Результаты и перспективы. - М., 2005. С. 111-135.

15. *Ляшевская О.Н., Кузнецова Ю.Л.* Русский фреймнет: к задаче создания корпусного словаря конструкций // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог 2009» (Бекасово, 27-31 мая 2009 г.). Вып. 8 (15). М.: РГГУ, 2009. С. 306-312. [pdf](#)
16. *Ляшевская, О. Н., Шаров С. А.* Частотный словарь современного русского языка: на материалах Национального корпуса русского языка. М.: Азбуковник, 2009. XV, 1074 с. (1)
17. [Научно-техническая информация. Сер. 2. 2003. № 6. http://www.viniti.ru/cgi-bin/nti/nti.pl?action=show&year=2\\_2003&issue=6](http://www.viniti.ru/cgi-bin/nti/nti.pl?action=show&year=2_2003&issue=6)
18. Научно-техническая информация. Сер. 2. 2005. № 3.
19. Национальный корпус русского языка: 2003–2005. М.: Индрик, 2005.
20. Национальный корпус русского языка: 2006–2008. Новые результаты и перспективы. С.-Петербург 2009
21. *Носенко И.А.* Начала статистики для лингвистов. М., 1981.
22. *Овчинникова, И. Г., Уланова И. А.* Компьютерное моделирование вербальной коммуникации: учеб.-метод. пособие. М.: Флинта: Наука, 2009. 134с. (2)
23. *Резникова Т. И.* Обзор общедоступных корпусов русского языка в Интернете. // *НТИ*, 2005, серия 2. №3. С. 2-9. (доступна с [http://catalog.viniti.ru/search\\_extra.asp](http://catalog.viniti.ru/search_extra.asp))
24. *Савчук С.О.* Метатекстовая разметка в Национальном корпусе русского языка: базовые принципы и основные функции // Национальный корпус русского языка: 2003-2005. Результаты и перспективы. - М., 2005. С. 62-88
25. [Труды международного семинара по компьютерной лингвистике и ее приложениям «Диалог 2000», «Диалог 2001», «Диалог 2002», «Диалог 2003», «Диалог 2004», «Диалог 2005».](http://www.dialog-21.ru/materials/) (<http://www.dialog-21.ru/materials/>)
26. Хэндауты к занятиям по теме «Статистическим методам в лингвистических исследованиях»
27. *Шаров С.А.* Представительный *корпус* русского языка в контексте мирового опыта. *НТИ*, 2003, серия 2. (<http://corpus.leeds.ac.uk/serge/publications/nti03-draft.pdf>)
28. A (very) brief description of the query syntax. <http://www.ims.uni-stuttgart.de/projekte/CorpusWorkbench/CQPSyntax.html>
29. *Biber D., S Conrad, R Reppen.* Corpus Linguistics: Investigating Language Structure and Use. - Cambridge University Press, 1998. URL: <http://books.google.com/books?id=2h5F7TXa6psC>
30. *Chris Manning and Hinrich Schütze*, Chapter 5. *Collocations*. Foundations of Statistical Natural Language Processing, MIT Press. Cambridge, MA: May 1999. Доступна с сайта [http://reslib.com/book/Foundations\\_of\\_Statistical\\_Natural\\_Language\\_Processing](http://reslib.com/book/Foundations_of_Statistical_Natural_Language_Processing) или <http://nlp.stanford.edu/fsnlp/promo/colloc.pdf>
31. *Chris Manning and Hinrich Schütze*, Chapter 7. Word sense disambiguation. Chapter 8. Lexical acquisition. Foundations of Statistical Natural Language Processing, MIT Press. Cambridge, MA: May 1999. Доступна с сайта [http://reslib.com/book/Foundations\\_of\\_Statistical\\_Natural\\_Language\\_Processing](http://reslib.com/book/Foundations_of_Statistical_Natural_Language_Processing)
32. Developing linguistic corpora: a guide to good practice. Edited by Martin Wynne. URL: <http://www.ahds.ac.uk/creating/guides/linguistic-corpora/index.htm>
33. EAGLES: *Recommendations for the morphosyntactic annotation of corpora*, EAGTCWGMAC/ R. 1996. Available from <ftp://ftp.ilc.pi.cnr.it/pub/eagles/corpora/annotate.ps.gz>
34. *Kennedy, Graeme.* An Introduction to Corpus Linguistics / Graeme Kennedy. - London: Longman, 1998. - 315 p
35. *Leech G.* Corpus Annotation Schemes // Literary and Linguistics Computing. 1993. Vol. 8, N. 4. P. 275–281.
36. *McEnery A., Wilson A.* Corpus Linguistics. Edinburgh, 1996. (<http://www.lancs.ac.uk/fss/courses/ling/corpus/>)

37. *Sinclair J.* Corpus, Concordance, Collocation. Oxford University Press, 1991.  
 38. Woods, Anthony; Paul Fletcher & Arthur Hughes. 1986. *Statistics in Language Studies*. Cambridge: Cambridge University Press.

## 6.2 Перечень ресурсов информационно-телекоммуникационной сети «Интернет»

### А. Корпуса и порталы по корпусной лингвистике

- Национальный корпус русского языка (<http://www.ruscorpora.ru> )
- Электронное издание Русской грамматики 1980 (<http://www.rusgram.narod.ru> )
- Сервис научной литературы JSTOR (<http://www.jstor.org/> )
- BNC <http://www.natcorp.ox.ac.uk/>
- COCA – Corpus of Contemporary American English (COCA) <http://www.americancorpus.org/>
- Корпуса современного немецкого языка и оболочка для их исследования COSMAS II <http://www.ids-mannheim.de/cosmas2/>
- <http://corpus.byu.edu/>

### Б. Специальные программы обработки текстов

#### Б.1. Конкордансеры

AntConc - скачивается с [http://www.antlab.sci.waseda.ac.jp/antconc\\_index.html](http://www.antlab.sci.waseda.ac.jp/antconc_index.html)

TextStat - <http://neon.niederlandistik.fu-berlin.de/en/textstat/>

SCP – Simple Concordance Program - <http://www.textworld.com/scp/>  
<http://www.concordancesoftware.co.uk/>

#### Б.2. Инструменты для сбора данных по N-граммам

<http://www.textworld.com/>

kfNgram - <http://www.kwicfinder.com/kfNgram/kfNgramHelp.html>

### В. Ресурсы на основе корпусов

<http://dict.ruslang.ru/> - словари на основе Национального корпуса русского языка

<http://www.sketchengine.co.uk/>

FrameBank

### Г. Программы морфологической, синтаксической, семантической разметки

<http://aot.ru/>

<http://www.connexor.eu/technology/machinese/>

<http://beta.visl.sdu.dk/>

### Д. Программы статистической обработки данных

Microsoft Excel, пакет языка R <http://www.r-project.org/>

### Е. Программы разметки собственного исследовательского корпуса

UAMStools – доступна с сайта [Programms\UAMCorpusToolInstaller1\\_3\\_2.exe](http://www.uam-stools.com/Programms/UAMCorpusToolInstaller1_3_2.exe)

### Инструментарий и документация

Комплект необходимых материалов: раздаточного материала к семинарам, текстов заданий и подробных инструкций по их выполнению, необходимых программных средств, презентаций лекций, электронного учебника по корпусной лингвистике В.П.Захарова находится в электронной библиотеке кафедры.

Теоретический материал: презентации лекций ([Lect materials](#)), дополнительный лекционный материал ([Lect materials\Other useful lections](#)), учебник по корпусной лингвистике ([Lect materials\Testbook VZ-corpora.doc](#)), глава из учебника А.Н.Баранова:

[Lect materials\baranov\\_112-137.pdf](#), список основной литературы по курсу можно найти в папке:

[Lect materials](#) в файле [Lect materials\Литература.doc](#). А также архив папки в письме

«Материалы по корпусной лингвистике: теория».

### 6.3 Профессиональные базы данных и информационно-справочные системы

Доступ к профессиональным базам данных: <https://liber.rsuh.ru/ru/bases>

№п /п	Наименование
1	Международные реферативные наукометрические БД, доступные в рамках национальной подписки в 2020 г. Web of Science Scopus
2	Профессиональные полнотекстовые БД, доступные в рамках национальной подписки в 2020 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

### 7. Материально-техническое обеспечение дисциплины

№п /п	Наименование ПО	Производитель	Способ распространения
1	Adobe Master Collection CS4	Adobe	лицензионное
2	Microsoft Office 2010	Microsoft	лицензионное
3	Windows 7 Pro	Microsoft	лицензионное
7	Microsoft Share Point 2010	Microsoft	лицензионное
12	Windows 10 Pro	Microsoft	лицензионное
13	Kaspersky Endpoint Security	Kaspersky	лицензионное
14	Microsoft Office 2016	Microsoft	лицензионное
15	Visual Studio 2019	Microsoft	лицензионное
16	Adobe Creative Cloud	Adobe	лицензионное
17	Zoom	Zoom	лицензионное

### 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 SE; 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/ практических/ лабораторных занятий**

#### **Занятие 1**

Проблемные корпуса (параллельные, диалектные, мультимедийные и др.)

**Цель занятия:** познакомить студентов с основными типами проблемных корпусов, особенностями их организации и использования.

#### **Занятие 2**

Поиск в корпусе. Использование языка SQR для поиска в корпусе. Составление сложных запросов к корпусу.

Цель занятия: научить студентов строить запросы к корпусам с использованием специального языка запросов CQP.

### **Занятие 3**

Особенности различных типов разметки. Морфологическая разметка

Цель занятия: познакомить студентов с основными типами морфологической аннотации в корпусе, методами разметки и методами тестирования

### **Занятие 4**

Особенности различных типов разметки. Синтаксическая разметка

Цель занятия: познакомить студентов с организацией синтаксически размеченных корпусов, с методами синтаксической разметки и методами ее тестирования.

### **Занятие 5**

Особенности различных типов разметки. Семантическая и анафорическая разметка

Цель занятия: познакомить студентов с видами семантической разметки, со специализированными лексикографическими ресурсами FrameNet, Фрейс Банк

### **Занятие 6**

Инструменты разметки собственного исследовательского корпуса

Цель занятия: познакомить студентов с технологиями разметки собственного исследовательского корпуса, научить размечать корпус с использованием специализированных программ

### **Занятие 7**

Составление конкордансов, частотных списков, списков коллокаций с использованием специальных программ

Цель занятия: познакомить студентов со специальными программами обработки собственных корпусов: конкордансерами, их возможностями и параметрами, научить работать с программой обработки собственного корпуса – составлению конкордансов, частотных списков, списков n-грамм

### **Занятие 8**

Основные статистические характеристики данных. Меры средней тенденции и изменчивости, коэффициент вариации и др.

Цель занятия: познакомить студентов с основными понятиями описательной статистики, дать представление об основных статистических характеристиках лингвистических данных

### **Занятие 9**

Статистические критерии зависимости. Проверка гипотез. Критерий хи-квадрат. Критерий Стьюдента. Критерий Манн-Уитни. Критерий Фишера.

Цель занятия: познакомить студентов с основными понятиями проверки статистических гипотез, научить применять различные статистические критерии для проверки гипотез при исследовании языковых данных

### **Занятие 10**

Корреляционный анализ

Цель занятия: познакомить студентов с основными понятиями корреляционного анализа, научить применять корреляционный анализ для исследования взаимозависимости различных факторов, влияющих на распределение некоторого языкового явления в корпусе / языке

### **Занятие 11**

Методы корпусной лексикографии

Цель занятия: познакомить студентов с основными методами корпусных методов в лексикографии, научить использовать корпусной материал при создании различных лексикографических ресурсов (словарей, специальных словарей, тезаурусов), познакомить с основными методами статистической обработки данных, используемых при создании лексикографических ресурсов: методами выделения коллокаций, исследования синонимов, разрешения семантической неоднозначности, составления лексикографического портрета лексемы

### **Занятие 12**

Сравнение двух корпусов. Стилеметрия

Цель занятия: познакомить студентов с основными сравнения лексического состава корпусов, а также с параметрами и методами стилеметрии.

### **Занятие 13**

Корпусные исследования в грамматике

Цель занятия: познакомить студентов с примерами использования корпусных методов в грамматических исследованиях, основными положениями грамматики конструкций и корпусными методами, используемыми в рамках данного направления.

### **Занятие 14**

Обсуждение исследовательских проектов

Цель занятия: ознакомить студентов с различными примерами корпусных исследований на основе обсуждений, выполненных в рамках курса проектов, выработать навыки анализа корпусных исследований различных языковых явлений

## **9.2 Другие материалы**

Все необходимые для обучения материалы даются на лекциях и практических занятиях.