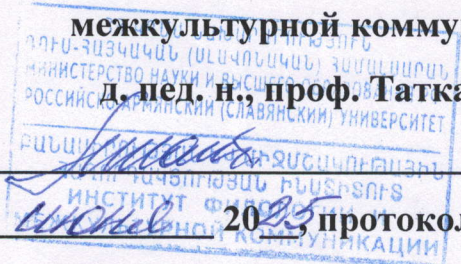


# ГОУ ВПО Российско-Армянский (Славянский) университет

Утверждено

директор Института филологии и  
межкультурной коммуникации  
д. пед. н., проф. Таткало Н. И.

«30» \_\_\_\_\_ 20\_\_ г., протокол № 8



## УЧЕБНО-МЕТОДИЧЕСКИЙ КОМПЛЕКС ДИСЦИПЛИНЫ

Наименование дисциплины: Б1.В.06 Корпусная лингвистика

Авторы:

к.ф.н. Трущелёв Павел Николаевич  
*ученое звание, ученая степень, ФИО*

Направление подготовки: 45.04.01 Филология

Наименование образовательной программы: магистерская программа  
«Преподавание русского языка как иностранного. Лингвокультурология»

# 1. АННОТАЦИЯ

## 1.1. Краткое описание содержания данной дисциплины.

Курс «Корпусная лингвистика» предназначен для филологов широкого профиля. Курс представляет одно из самых передовых направлений в филологии. Курс отражает требования к профессиональным компетенциям современного специалиста с филологическим образованием: компетенции предполагают наличие умений работать с массивами текстовых данных. На занятиях студенты знакомятся с теоретическими основами и практическими методами работы с лингвистическими корпусами – электронными коллекциями текстов, аннотированных с использованием формальных правил. В рамках курса изучаются принципы построения и аннотирования корпусов, методы извлечения и анализа языковых данных, а также прикладные задачи, решаемые с использованием корпусов: частотный анализ, коллокации, изучение вариативности, автоматическая обработка текста и др. Особое внимание уделяется работе с национальными и специализированными корпусами русского языка.

## 1.2. Трудоемкость в академических кредитах и часах, формы итогового контроля (экзамен/зачет)

**3 ECTS, 108 академических часа, форма итогового контроля – зачет.**

## 1.3. Взаимосвязь дисциплины с другими дисциплинами учебного плана специальности (направления).

Курс «Корпусная лингвистика» носит ярко выраженный комплексный характер. Для изучения курса необходимы знания, умения и компетенции, полученные обучающимися при изучении филологических дисциплин, составляющих фундамент лингвистического образования: общее языкознание, лексикология, стилистика, грамматика и др.

Ввиду специфики объекта изучения – аннотированной коллекции текстов, курс «Корпусная лингвистика» имеет теснейшую связь со всеми филологическими дисциплинами. В первую очередь существует связь с дисциплинами, направленными на изучение структурных и функциональных характеристик системы языка и речи, поскольку такие характеристики выступают предметом аннотирования.

## 1.4. Результаты освоения программы дисциплины:

<b>Код компетенции</b> <i>(в соответствии рабочим с учебным планом)</i>	<b>Наименование компетенции</b> <i>(в соответствии рабочим с учебным планом)</i>	<b>Код индикатора достижения компетенций</b> <i>(в соответствии рабочим с учебным планом)</i>	<b>Наименование индикатора достижений компетенций</b> <i>(в соответствии рабочим с учебным планом)</i>
--	---	--	---

УК-4	Способен применять современные коммуникативные технологии, в том числе на иностранном(ых) языке(ах), для академического и профессионального взаимодействия	УК-4.1	Владеет навыками работы с электронными почтовыми системами, видео-конференц-связью, мессенджерами и социальными сетями для эффективного обмена информацией
		УК-4.2	Способен создавать и проводить мультимедийные презентации на русском и иностранных языках
		УК-4.3	Умеет использовать специализированные платформы для проведения научных мероприятий, взаимодействовать с участниками
ПК-1	Владеет навыками самостоятельного проведения научных исследований в области системы языка в синхроническом и диахроническом аспектах, в сфере устной, письменной и виртуальной коммуникации	ПК-1.1	Может формулировать гипотезы, определять цели и задачи исследования, выбирать соответствующие методы и подходы, а также собирать и анализировать лингвистические данные
		ПК-1.2	Владеет методами исследования текстов
		ПК-1.3	Может самостоятельно разрабатывать и писать научные статьи
ПК-2	Владеет навыками квалифицированного анализа, оценки, реферирования, оформления, редактирования и продвижения результатов собственной научной деятельности	ПК-2.1	Умеет выявлять ключевые аспекты и тенденции, использовать подходящие методы анализа и интерпретации результатов, а также делать обоснованные выводы на основе полученных данных
		ПК-2.2	Умеет критически оценивать свои работы и работы других, предлагать решения для их улучшения и дальнейшего развития
		ПК-2.3	Владеет навыками составления аннотаций, рефератов и обзоров, передающих ключевые аспекты научных исследований, а также представления результатов перед аудиторией
ПК-5	Владеет навыками самостоятельного приобретения, в том числе с помощью информационных технологий, и использования в практической деятельности новых знаний и умений в областях, непосредственно связанных со сферой деятельности	ПК-5.1	Знает современные информационные технологии и ресурсы, которые могут быть использованы для обучения и повышения квалификации
		ПК-5.2	Умеет эффективно искать, отбирать и критически оценивать информацию из

			различных источников, включая интернет, научные базы данных, профессиональные сообщества и онлайн-курсы
		ПК-5.3	Владеет навыками использования информационных технологий и цифровых инструментов для организации обучения, а также в исследовательских целях

## 2. УЧЕБНАЯ ПРОГРАММА

### 2.1. Цели и задачи дисциплины.

Целью освоения курса «Корпусная лингвистика» является формирование у студентов знаний в области создания и применения лингвистических корпусов, приобретение практических умений работы с корпусами и связанными с ними компьютерными инструментами в сфере цифровой филологии.

Задачи освоения дисциплины:

- освоить базовый понятийный аппарат;
- получить навыки работы с корпусом;
- получить навыки построения и аннотирования лингвистического корпуса (морфологическая, синтаксическая, прагматическая разметка и т. п.);
- получить навыки интерпретации частотной и статистической информации, полученной из корпусов;
- познакомиться с принципами применения корпусных методов в различных областях лингвистики.

### 2.2. Трудоемкость дисциплины и виды учебной работы (в академических часах и зачетных единицах).

Виды учебной работы	Всего, в акад. часах	Распределение по семестрам					
		1 сем	2 сем	3 сем	4 сем.	5 сем	6 сем.
1	2	3	4	5	6	7	8
1. Общая трудоемкость изучения дисциплины по семестрам, в т. ч.:	<b>108</b>		<b>108</b>				
1.1. Аудиторные занятия, в т. ч.:	<b>32</b>		<b>32</b>				
1.1.1. Лекции	16		16				
1.1.2. Практические занятия	16		16				
1.2. Самостоятельная работа, в т. ч.:	<b>58</b>		<b>58</b>				
1.2.1. Подготовка к экзамену	8		8				
1.2.2. Другие виды самостоятельной работы	50		50				
Итоговый контроль (экзамен)	<b>18</b>		<b>18</b>				

## 2.3. Содержание дисциплины

### 2.3.1. Тематический план и трудоемкость аудиторных занятий (модули, разделы дисциплины и виды занятий) по рабочему учебному плану

Разделы и темы дисциплины	Всего (ак. часов)	Лекции (ак. часов)	Практ. занятия (ак. часов)
1	2=3+4	3	4
Корпусы и корпусная лингвистика в гуманитарной науке и практике	4	2	2
Использование лингвистических корпусов	8	4	4
Создание лингвистических корпусов	10	5	5
Программы для работы с корпусом	10	5	5

### 2.3.2. Краткое содержание разделов дисциплины в виде тематического плана

#### Тема 1. Корпусы и корпусная лингвистика в гуманитарной науке и практике

Лингвистический корпус: сущность и основные характеристики. Актуальные сферы применения корпусов. Классификации лингвистических корпусов. Особые типы корпусов.

#### Тема 2. Использование лингвистических корпусов

Зарубежные национальные корпусы. Современные корпусы русского языка. Лексикографические практики на базе корпусов. Грамматические исследования, основанные на корпусах. Лингвистические корпусы в изучении дискурса. Использование корпуса в обучении языкам.

#### Тема 3. Создание лингвистических корпусов

Технологии проектирования и разработки корпусов. Отбор источников и процедуры обработки естественного языка: токенизация, лемматизация, стемминг, парсинг. Разметка: виды и средства. Лингвистическая и экстралингвистическая разметка. Стандартизация в корпусной лингвистике. Международные стандарты и проекты.

#### Тема 4. Программы для работы с корпусом

Конкордансер и корпусный менеджер как поисковая система. Языки запросов корпусных менеджеров. Корпусные менеджеры на базе веба. Компьютерные инструменты использования корпусов. Программы анализа, лингвистической обработки и преобразования текста.

### 2.3.3. Краткое содержание практических занятий

**Формат:** очные занятия в аудиториях университета.



итоговых оценках промежуточных контролей								
Веса оценок промежуточных контролей в итоговых оценках промежуточных контролей								
Вес итоговой оценки 1-го промежуточног о контроля в результатирующе й оценке промежуточных контролей								
Вес итоговой оценки 2-го промежуточног о контроля в результатирующе й оценке промежуточных контролей								
Вес результатирующе й оценки промежуточных контролей в результатирующе й оценке итогового контроля								
<b>Вес итогового контроля (Экзамен/зачет ) в результатирующе й оценке итогового контроля</b>								
	$\Sigma = 1$	$\Sigma = 1$	$\Sigma = 1$	$\Sigma = 1$	$\Sigma = 1$	$\Sigma = 1$	$\Sigma = 1$	$\Sigma = 1$

Примечание: М – учебный модуль.

### 3. ТЕОРЕТИЧЕСКИЙ БЛОК

*(указываются материалы, необходимые для освоения учебной программы дисциплины)*

#### 3.1. Материалы по теоретической части курса

##### 3.1.1. учебники:

Захаров В. П., Богданова С. Ю. Корпусная лингвистика; 3-е изд., перераб. СПб.: изд-во С.-Петербур. ун-та, 2020. 234 с.

### 3.1.2. учебные пособия:

Беляева Л. Н., Пиотровская К. Р. Сетевой инструментарий филолога. СПб.: Книжный дом, 2014. 48 с.

Копотев М. Введение в корпусную лингвистику. Praha: Animedia Company, 2014.

### 3.1.3. электронные материалы (электронные учебники, учебные пособия, курсы и краткие конспекты лекций, презентации РРТ и т.п.)

База данных «Дети-инофоны Вологодской области». Режим доступа: <https://inophone35.ru/>, свободный.

Детский корпус. Режим доступа: <http://www.web-corpora.net/children/>, свободный.

Корпус русских учебных текстов (КРУТ). Режим доступа: [http://web-corpora.net/learner\\_corpus](http://web-corpora.net/learner_corpus), свободный.

Корпус «Один речевой день». Режим доступа: <https://ord.spbu.ru/>, свободный.

Лингвистические корпуса и сервисы. Режим доступа: <http://www.web-corpora.net/>, свободный.

Ляшевская О. Н., Шаров С. А. Частотный словарь современного русского языка (на материалах Национального корпуса русского языка). 2009. Режим доступа: <http://dict.ruslang.ru/freq.php>, свободный.

Мультимедийный словарь прагматических маркеров. Режим доступа: <https://www.ord-multimedia-dict.com/>, свободный.

Национальный корпус русского языка. Режим доступа: <https://ruscorpora.ru/>, свободный.

Рассказы о сновидениях и другие корпуса звучащей речи. Режим доступа: <https://spokencorpora.ru/>, свободный.

Русская корпусная грамматика. Режим доступа: <http://rusgram.ru/>, свободный.

Русский учебный корпус. Режим доступа: <http://www.web-corpora.net/RLC>, свободный.

Сравнительная поэтика и сравнительное литературоведение. Режим доступа: <https://cpcl.info/>, свободный.

Текстометр: Анализ текста онлайн. Режим доступа: <https://textometr.ru/>, свободный.

AntConc: A freeware corpus analysis toolkit for concordancing and text analysis. Режим доступа: <https://laurenceanthony.net/software.html>, свободный.

ArgNetBank Studio: Платформа для моделирования и анализа аргументации в дискурсе. Режим доступа: <https://uniserv.iis.nsk.su/arg/>, свободный.

English-Corpora.org. Режим доступа: <https://www.english-corpora.org/>, свободный.

Corpus of Russian spoken in Bashkortostan. Режим доступа: <https://lingconlab.ru/BashkirRus/>, свободный.

Corpus of Russian spoken in Chuvashia. Режим доступа: <https://lingconlab.ru/ChuvashRus/>, свободный.

Lancaster Stats Tools: Analyse corpus data with ease. Режим доступа: <https://corpus-stats.lancs.ac.uk/>, свободный.

MyStem: Морфологический анализ текста на русском языке. Режим доступа: <https://yandex.ru/dev/mystem/>, свободный.

Mystem+: онлайн-ресурс, на котором Вы можете протестировать системы морфологического анализа текстов для русского языка. Режим доступа: <http://web-corpora.net/wsgi/mystemplus.wsgi/mystemplus/>, свободный.

Ru.Lingva: Интеллектуальные решения для текстовой аналитики. Режим доступа: <https://rulex.kpfu.ru/>, свободный.

Russian Drama Corpus. Режим доступа: <https://dracor.org/rus>, свободный.

UDPipe. Режим доступа: <https://lindat.mff.cuni.cz/services/udpipe/>, свободный.

VOYANT: see through your text. Режим доступа: <https://voyant-tools.org/>, свободный.

### **3.1.4. др. варианты материалов, необходимых для освоения учебной программы дисциплины:**

Беляева Л. Н., Камшилова О. Н., Пиотровская К. Р. Сетевые лингвистические технологии. СПб.: СПб.: РГПУ им. А. И. Герцена, 2019. 111 с.

Введение в корпусную лингвистику // Открытое образование. Режим доступа: <https://openedu.ru/course/hse/CORPUS/>, свободный.

Зубов В. И., Риехакайнен Е. И. (В)начале было слово: о единицах анализа дискурса на примере корпуса речи учителей // Социо- и психолингвистические исследования. 2023. Вып. 11. С. 14–18.

Курс корпусной лингвистики. Режим доступа: <https://rykov-cl.narod.ru/t.html>, свободный.

Моделирование в корпусной лингвистике: специализированные корпуса русского языка / В. П. Захаров, И. В. Азарова, О. А. Митрофанова, А. М. Попов, М. В. Хохлова. СПб., изд-во С.-Петербур. ун-та, 2019. 207 с.

Плунгян В. Корпусная лингвистика // ПостНаука. 2013. Режим доступа: <https://postnauka.ru/video/7783>, свободный.

Публикации по Национальному корпусу русского языка. Режим доступа: <http://www.ruscorpora.ru/corpora-biblio.html>, свободный.

Рассказы о сновидениях: корпусное исследование устного русского дискурса / под ред. А. А. Кибрика, В. И. Подлесской. М.: Языки славянской культуры, 2009. 736 с.

Сичинава Д. В. К задаче создания корпусов русского языка // Научно-техническая информация. Сер. 2. Информационные процессы и системы. 2002. № 11. С. 25–31.

Baker P., Hardie A., McEnery T. A Glossary of Corpus Linguistics. Edinburgh: Edinburgh University Press, 2006. 188 p.

## **4. Фонды оценочных средств**

*(указываются материалы, необходимые для проверки уровня знаний в соответствии с содержанием учебной программы дисциплины).*

### **4.1. Планы практических и семинарских занятий:**

разметка текстов для создания коллекции текстов;

работа с корпусами текстов: извлечение лингвистических данных, обработка данных, интерпретация данных;

работа с информационными корпусными инструментами: AntConc, Mystem+, RuLingva, Текстометр и др.

## **4.2. Материалы по практической части курса:**

### **4.2.1. учебно-методические пособия:**

Беляева Л. Н., Пиотровская К. Р. Сетевой инструментарий филолога. СПб.: Книжный дом, 2014. 48 с.

Копотев М. Введение в корпусную лингвистику. Praha: Animedia Company, 2014.

### **4.2.2. др. виды материалов:**

АОТ: автоматическая обработка текстов. Режим доступа: <http://www.aot.ru/>, свободный.

База данных «Дети-инофоны Вологодской области». Режим доступа: <https://inophone35.ru/>, свободный.

Детский корпус. Режим доступа: <http://www.web-corpora.net/children/>, свободный.

Инструкция для пользователя национальным корпусом. Режим доступа: <https://studiorum.ruscorpora.ru/manual/>, свободный.

Корпус русских учебных текстов (КРУТ). Режим доступа: [http://web-corpora.net/learner\\_corpus](http://web-corpora.net/learner_corpus), свободный.

Корпус «Один речевой день». Режим доступа: <https://ord.spbu.ru/>, свободный.

Лингвистические корпуса и сервисы. Режим доступа: <http://www.web-corpora.net/>, свободный.

Ляшевская О. Н., Шаров С. А. Частотный словарь современного русского языка (на материалах Национального корпуса русского языка). 2009. Режим доступа: <http://dict.ruslang.ru/freq.php>, свободный.

Мультимедийный словарь прагматических маркеров. Режим доступа: <https://www.ord-multimedia-dict.com/>, свободный.

Национальный корпус русского языка. Режим доступа: <https://ruscorpora.ru/>, свободный.

Рассказы о сновидениях и другие корпуса звучащей речи. Режим доступа: <https://spokencorpora.ru/>, свободный.

Русская корпусная грамматика. Режим доступа: <http://rusgram.ru/>, свободный.

Русский учебный корпус. Режим доступа: <http://www.web-corpora.net/RLC>, свободный.

Сравнительная поэтика и сравнительное литературоведение. Режим доступа: <https://cpcl.info/>, свободный.

Текстометр: Анализ текста онлайн. Режим доступа: <https://textometr.ru/>, свободный.

AntConc: A freeware corpus analysis toolkit for concordancing and text analysis. Режим доступа: <https://laurenceanthony.net/software.html>, свободный.

ArgNetBank Studio: Платформа для моделирования и анализа аргументации в дискурсе. Режим доступа: <https://uniserv.iis.nsk.su/arg/>, свободный.

Baker P., Hardie A., McEnery T. A Glossary of Corpus Linguistics. Edinburgh: Edinburgh University Press, 2006. 188 p.

English-Corpora.org. Режим доступа: <https://www.english-corpora.org/>, свободный.

Corpus of Russian spoken in Bashkortostan. Режим доступа: <https://lingconlab.ru/BashkirRus/>, свободный.

Corpus of Russian spoken in Chuvashia. Режим доступа: <https://lingconlab.ru/ChuvashRus/>, свободный.

MyStem: Морфологический анализ текста на русском языке. Режим доступа: <https://yandex.ru/dev/mystem/>, свободный.

Mystem+: онлайн-ресурс, на котором Вы можете протестировать системы морфологического анализа текстов для русского языка. Режим доступа: <http://web-corpora.net/wsgi/mystemplus.wsgi/mystemplus/>, свободный.

Ru.Lingva: Интеллектуальные решения для текстовой аналитики. Режим доступа: <https://rulex.kpfu.ru/>, свободный.

Russian Drama Corpus. Режим доступа: <https://dracor.org/rus>, свободный.

Russian Learner Translator Corpus (RusLTC). Режим доступа: <https://rus-ltc.org/>, свободный.

UDPipe. Режим доступа: <https://lindat.mff.cuni.cz/services/udpipe/>, свободный.

#### **4.3. Вопросы и задания для самостоятельной работы студентов:**

подготовка корпуса текстов (или фрагментов) по заданным параметрам: стиль, источник, объем, язык и т. д.;

аннотирование текстов для корпуса по заданным параметрам: морфологическая разметка, синтаксическая разметка, прагматические маркеры и т. п.

#### **4.4. Перечень экзаменационных вопросов:**

понятие корпуса, основные требования к лингвистическому корпусу;

история развития корпусной лингвистики, классификация корпусов;

использование корпусных данных в различных областях лингвистики;

типы лингвистической разметки корпусов;

инструментарий создания корпуса;

инструментарий работы с корпусом;

корпусы русского языка: виды и задачи;

количественные методы в корпусных исследованиях (описательная статистика, ipm, tf/idf, MI, t-score)

#### **4.5. Образцы экзаменационных билетов:**

## БИЛЕТ № 1

1. Понятие корпуса, основные требования к лингвистическому корпусу.
2. Практическое задание.

### **4.6. Образцы экзаменационных практических заданий:**

выполнение задания по работе с корпусом;  
демонстрация навыков аннотирования текста.

## **5. Методический блок**

### **5.1. Методические рекомендации для студентов по подготовке к практическим занятиям, по организации самостоятельной работы студентов при изучении конкретной дисциплины.**

Практическое занятие – форма систематических учебных занятий, с помощью которых студенты вырабатывают навыки проведения корпусного исследования.

Практические занятия проводятся по вычитанному на лекциях материалу и связаны, как правило, с детальным разбором отдельных вопросов лекций. Лекционный материал будет закрепляться как в результате обсуждения и анализа языкового материала, так и с помощью решения проблемных ситуаций, задач.

Приветствуется проактивность студентов в этом вопросе. При выполнении заданий в рамках как аудиторной, так и самостоятельной работы студент всегда может обращаться к авторитетной справочной и научной литературе, которая представляет оригинальные и релевантные методы анализа.

Студенты могут обращаться к лингвистическим моделям типа GPT в ознакомительных целях. Результаты работы с лингвистическими моделями должны быть тщательно проверены студентом. Студенту следует учитывать, что анализ, проведенный лингвистической моделью, является неполным и неповторяемым (что делает его недостаточным).

При подготовке к практическим занятиям следует использовать литературу, а также руководствоваться приведенными указаниями и рекомендациями.

На практических занятиях приветствуется активное участие в обсуждении конкретных ситуаций, поиск на основе полученных знаний наиболее эффективных решений поставленных проблем, поиск полезного дополнительного материала по тематике занятий.

Поощряется участие студентов в проводимых университетом проектах, посвященных русскому языку (напр., «Тотальный диктант», «Дни русского языка», «День славянской письменности» и т. п.)

Студенту рекомендуется следующая схема подготовки к занятию. 1. Проработать конспект лекций. 2. Ознакомиться с темой, используя рекомендованную литературу. 3. Выполнить домашнее задание. 4. Сформулировать вопросы к преподавателю. На занятия следует приходиться подготовленным – иметь необходимые вещи.

Занятия могут проводиться в форме беседы со всеми студентами группы или с отдельными студентами. Этот вид занятия называется коллоквиумом (собеседование). Коллоквиумы проводятся по конкретным вопросам дисциплины. Коллоквиум отличается, в первую очередь тем, что во время этого занятия могут быть опрошены все студенты или значительная часть студентов группы.

За время обучения студенты проводят одну собственную уникальную корпусную работу, которая предполагает или подготовку корпуса текстового материала, или разметку текстов из корпуса.

## **5.2. Критерии формирования оценок:**

### **5.2.1. Текущая оценка формируется следующим образом:**

<b>Критерий</b>	<b>Максимальный балл</b>	<b>Описание</b>
Знание теории	30	Оценивается уровень владения языком дисциплины, знание лекционных тем, знание ресурсов по дисциплине.
Практические навыки	70	Оцениваются умения работать с корпусом и с текстом как частью корпуса.
Итого	100	

### **5.2.2. Оценка за ответ на экзамене формируется следующим образом:**

<b>Критерий</b>	<b>Максимальный балл</b>	<b>Описание</b>
Знание теории	30	Оценивается информация о предмете обсуждения, а также знание терминов, основных ученых, сведений о корпусных ресурсах.
Практические навыки	70	Оцениваются умения работать с корпусом и с текстом как частью корпуса.
Итого	100	