

**МИНОБРНАУКИ РОССИИ**

**Федеральное государственное автономное образовательное  
учреждение высшего образования "Пермский  
государственный национальный исследовательский  
университет"**

**Кафедра теоретического и прикладного языкознания**

**Авторы-составители: Худякова Екатерина Сергеевна**

**Рабочая программа дисциплины  
КОРПУСНАЯ ЛИНГВИСТИКА I  
Код УМК 100397**

**Утверждено  
Протокол №9  
от «30» мая 2023 г.**

**Пермь, 2023**

## **1. Наименование дисциплины**

Корпусная лингвистика I

## **2. Место дисциплины в структуре образовательной программы**

Дисциплина входит в вариативную часть Блока « М.1 » образовательной программы по направлениям подготовки (специальностям):

Направление подготовки: **44.04.01** Педагогическое образование  
направленность Современное филологическое образование

### **3. Планируемые результаты обучения по дисциплине**

В результате освоения дисциплины **Корпусная лингвистика I** у обучающегося должны быть сформированы следующие компетенции:

**44.04.01** Педагогическое образование (направленность : Современное филологическое образование)

**ПК.2** Способен преподавать по программам бакалавриата и ДПП, ориентированным на соответствующий уровень квалификации

#### **Индикаторы**

**ПК.2.2** Осуществляет теоретико-методологическую рефлексия актуальных проблем в своей предметной области

#### 4. Объем и содержание дисциплины

|                                                                   |                                                                                              |
|-------------------------------------------------------------------|----------------------------------------------------------------------------------------------|
| <b>Направление подготовки</b>                                     | 44.04.01 Педагогическое образование (направленность: Современное филологическое образование) |
| <b>форма обучения</b>                                             | очная                                                                                        |
| <b>№№ триместров, выделенных для изучения дисциплины</b>          | 5                                                                                            |
| <b>Объем дисциплины (з.е.)</b>                                    | 3                                                                                            |
| <b>Объем дисциплины (ак.час.)</b>                                 | 108                                                                                          |
| <b>Контактная работа с преподавателем (ак.час.), в том числе:</b> | 36                                                                                           |
| <b>Проведение лекционных занятий</b>                              | 12                                                                                           |
| <b>Проведение практических занятий, семинаров</b>                 | 24                                                                                           |
| <b>Самостоятельная работа (ак.час.)</b>                           | 72                                                                                           |
| <b>Формы текущего контроля</b>                                    | Защищаемое контрольное мероприятие (2)<br>Итоговое контрольное мероприятие (1)               |
| <b>Формы промежуточной аттестации</b>                             | Зачет (5 триместр)                                                                           |

## **5. Аннотированное описание содержания разделов и тем дисциплины**

### **Корпусная лингвистика I**

Курс является пропедевтикой в корпусную лингвистику. Основное внимание уделено базовым понятиям корпусной лингвистики и принципам компьютерной реализации корпуса. В результате освоения дисциплины студент должен быть способен реализовать проект создания лингвистического корпуса.

#### **Корпусная лингвистика: проблематика, цели, применение.**

Эмпиризм современной лингвистики. Пересечение с компьютерной лингвистикой: применение корпусов в приложениях NLP. Корпусы в лингвистических исследованиях. Краткая история корпусной лингвистики.

#### **Понятие лингвистического корпуса.**

Определение корпуса. Корпусный поворот в лингвистике. Корпусы и близкие понятия. Квантитативные единицы корпуса: токен и токенизация, n-граммы, коллокации. Статистические методы при использовании корпусов. Дистрибуция текста. Практикум по очистке текста и токенизации с применением Python. Библиотеки NLP для подготовки массивов текстов.

#### **Типы и виды корпусов.**

Типология корпусов: по количеству языков; по идиомам языка; по типу данных; по жанрам; по назначению, по синхронности / диахронности, по объему; по организации. Специализированные корпуса: прикладные без пользовательского интерфейса, мультимедийные, для специальных целей (корпуса диалектной речи, корпуса детской речи, корпуса билингвальной речи и др.).

#### **Типы корпусной разметки.**

Понятие разметки корпуса. Типы разметки: метаразметка и ее теги; структурная разметка (токенизация по словам, по предложениям, по структурным элементам текста); собственно лингвистическая разметка (морфологическая, синтаксическая, семантическая и др.). Стандарты разметки EAGLES (European Advisory Group on Language Engineering Standards), TEI (Text Encoding for Interchange), XCES (XML Corpus Encoding Standard). Представление аннотации в xml: атрибут -- значение.

#### **Понятие корпусного менеджера.**

Общее представление о корпусном менеджере. Элементы интерфейса. Инструмент разметки и инструмент поиска. Знакомство с UAM corpustool, Elan, Exmaralda. Практика по работе с корпусными менеджерами.

#### **Аннотирование корпуса. Проект по созданию корпуса.**

Кейс: создание авторского корпуса. Практическая работа предполагает: определение цели и типа корпуса; подбор текстов и их препроцессинг; обоснование типов разметки; осуществление аннотирования корпуса в корпусном менеджере.

## **6. Методические указания для обучающихся по освоению дисциплины**

Освоение дисциплины требует систематического изучения всех тем в той последовательности, в какой они указаны в рабочей программе.

Основными видами учебной работы являются аудиторные занятия. Их цель - расширить базовые знания обучающихся по осваиваемой дисциплине и систему теоретических ориентиров для последующего более глубокого освоения программного материала в ходе самостоятельной работы. Обучающемуся важно помнить, что контактная работа с преподавателем эффективно помогает ему овладеть программным материалом благодаря расстановке необходимых акцентов и удержанию внимания интонационными модуляциями голоса, а также подключением аудио-визуального механизма восприятия информации.

Самостоятельная работа преследует следующие цели:

- закрепление и совершенствование теоретических знаний, полученных на лекционных занятиях;
- формирование навыков подготовки текстовой составляющей информации учебного и научного назначения для размещения в различных информационных системах;
- совершенствование навыков поиска научных публикаций и образовательных ресурсов, размещенных в сети Интернет;
- самоконтроль освоения программного материала.

Обучающемуся необходимо помнить, что результаты самостоятельной работы контролируются преподавателем во время проведения мероприятий текущего контроля и учитываются при промежуточной аттестации.

Обучающимся с ОВЗ и инвалидов предоставляется возможность выбора форм проведения мероприятий текущего контроля, альтернативных формам, предусмотренным рабочей программой дисциплины. Предусматривается возможность увеличения в пределах 1 академического часа времени, отводимого на выполнение контрольных мероприятий.

Процедура оценивания результатов обучения инвалидов и лиц с ограниченными возможностями здоровья по дисциплине предусматривает предоставление информации в формах, адаптированных к ограничениям их здоровья и восприятия информации.

При проведении текущего контроля применяются оценочные средства, обеспечивающие передачу информации, от обучающегося к преподавателю, с учетом психофизиологических особенностей здоровья обучающихся.

## **7. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине**

При самостоятельной работе обучающимся следует использовать:

- конспекты лекций;
- литературу из перечня основной и дополнительной учебной литературы, необходимой для освоения дисциплины (модуля);
- текст лекций на электронных носителях;
- ресурсы информационно-телекоммуникационной сети "Интернет", необходимые для освоения дисциплины;
- лицензионное и свободно распространяемое программное обеспечение из перечня информационных технологий, используемых при осуществлении образовательного процесса по дисциплине;
- методические указания для обучающихся по освоению дисциплины.

## 8. Перечень основной и дополнительной учебной литературы

### Основная:

1. Галич, Г. Г. Современные направления лингвистики : учебное пособие для студентов магистратуры, обучающихся по направлению 45.04.02 «Лингвистика», профиль подготовки «Практика и лингводидактика профессионально ориентированного перевода» / Г. Г. Галич. — Омск : Издательство Омского государственного университета, 2020. — 184 с. — ISBN 978-5-7779-2464-3. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. <https://www.iprbookshop.ru/108138>
2. Методы когнитивного анализа семантики слова: компьютерно-корпусный подход / В. И. Заботкина, Е. Е. Голубкова, М. А. Кронгауз [и др.] ; под редакцией В. И. Заботкиной. — 2-е изд. — Москва : Издательский Дом ЯСК, 2019. — 342 с. — ISBN 978-5-907117-73-0. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. <http://www.iprbookshop.ru/97629.html>
3. Захаров, В. П. Корпусная лингвистика : учебник для студентов гуманитарных вузов / В. П. Захаров, С. Ю. Богданова. — Иркутск : Иркутский государственный лингвистический университет, 2011. — 161 с. — ISBN 978-5-88267-316-0. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. <http://www.iprbookshop.ru/21088>

### Дополнительная:

1. Рассказы о сновидениях. Корпусное исследование устного русского дискурса / О. А. Савельева-Трофимова, В. Л. Цуканова, А. О. Литвиненко [и др.] ; под редакцией А. А. Кибрик, В. И. Подлесская. — Москва : Языки славянских культур, 2009. — 736 с. — ISBN 978-5-9551-0303-7. — Текст : электронный // Цифровой образовательный ресурс IPR SMART : [сайт]. <http://www.iprbookshop.ru/15133>
2. McCarthy M., O'Dell F. English collocations in use. How words work together for fluent and natural English: self-study and classroom use / M. McCarthy, F. O'Dell. — Cambridge: Cambridge University Press, 2006, ISBN 0-521-60378-1. — 190.
3. Fischer R. Lexical change in present-day English: a corpus-based study of the motivation, institutionalization, and productivity of creative neologisms / R. Fischer. — Tübingen: GNV, 1998, ISBN 3-8233-4940-6. — 209.
4. Лесников С. В. Метаязык лингвистики. в 2 т. Т. 1. Проблемы систематизации терминосистемы (гизаурис, классификация, синопсис) / С. В. Лесников. — Санкт-Петербург: Нестор-История, 2021, ISBN 978-5-4469-2007-5. — 512. — Библиогр.: с. 377-478
5. Корпусный анализ русского стиха: [сборник статей] / Рос. акад. наук, Ин-т рус. яз. им. В. В. Виноградова. — Москва: Азбуковник, 2013, ISBN 978-5-91172-073-5. — 2651. — Библиогр. в конце ст.

## 9. Перечень ресурсов сети Интернет, необходимых для освоения дисциплины

<https://ruscorpora.ru/> Национальный корпус русского языка

<https://www.korpus.cz/> Чешский национальный корпус

<https://www.english-corpora.org/bnc/> Британский национальный корпус

<https://transcriber.semograph.com/> Инструмент создания корпуса

## 10. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине

Образовательный процесс по дисциплине **Корпусная лингвистика I** предполагает использование следующего программного обеспечения и информационных справочных систем:

Презентационные материалы (слайды по темам лекционных и практических занятий); доступ в режиме on-line в Электронную библиотечную систему (ЭБС); доступ в электронную информационно-образовательную среду университета.

Образовательный процесс по дисциплине предполагает использование следующего программного обеспечения и информационных систем:

1. Офисный пакет
  2. Приложения для создания корпусов (UAM corpustool, Elan, Exmaralda)
  3. Онлайн приложение для создания корпуса (<https://transcriber.semograph.com/>)
- Программы пп. 1–2 являются свободно распространяемыми (free software).

При освоении материала и выполнения заданий по дисциплине рекомендуется использование материалов, размещенных в Личных кабинетах обучающихся ЕТИС ПГНИУ ([student.psu.ru](http://student.psu.ru)).

При организации дистанционной работы и проведении занятий в режиме онлайн могут использоваться:

система видеоконференцсвязи на основе платформы BigBlueButton (<https://bigbluebutton.org/>).

система LMS Moodle (<http://e-learn.psu.ru/>), которая поддерживает возможность использования текстовых материалов и презентаций, аудио- и видеоконтент, а так же тесты, проверяемые задания, задания для совместной работы.

система тестирования Indigo (<https://indigotech.ru/>).

## 11. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине

Для проведения лекционных занятий и текущего контроля необходима аудитория, оснащенная презентационной техникой (проектор, экран, компьютер/ноутбук) с соответствующим программным обеспечением, меловой (и) или маркерной доской.

Для проведения практических занятий необходим компьютерный класс, оснащенный персональными ЭВМ и соответствующим программным обеспечением. Состав оборудования определен в Паспорте компьютерного класса.

Аудитория для самостоятельной работы, оснащенный компьютерной техникой с возможностью подключения к сети «Интернет», обеспеченный доступом в электронную информационно-образовательную среду университета.

Помещения Научной библиотеки ПГНИУ.

Помещения научной библиотеки ПГНИУ для обеспечения самостоятельной работы обучающихся:

1. Научно-библиографический отдел, корп.1, ауд. 142. Оборудован 3 персональными компьютера с доступом к локальной и глобальной компьютерным сетям.



2. Читальный зал гуманитарной литературы, корп. 2, ауд. 418. Оборудован 7 персональными компьютерами с доступом к локальной и глобальной компьютерным сетям.

3. Читальный зал естественной литературы, корп.6, ауд. 107а. Оборудован 5 персональными компьютерами с доступом к локальной и глобальной компьютерным сетям.

4. Отдел иностранной литературы, корп.2 ауд. 207. Оборудован 1 персональным компьютером с доступом к локальной и глобальной компьютерным сетям.

5. Библиотека юридического факультета, корп.9, ауд. 4. Оборудована 11 персональными компьютерами с доступом к локальной и глобальной компьютерным сетям.

6. Читальный зал географического факультета, корп.8, ауд. 419. Оборудован 6 персональными компьютерами с доступом к локальной и глобальной компьютерным сетям.

Все компьютеры, установленные в помещениях научной библиотеки, оснащены следующим программным обеспечением:

Операционная система ALT Linux;

Офисный пакет Libreoffice.

Справочно-правовая система «КонсультантПлюс»

**Фонды оценочных средств для аттестации по дисциплине  
Корпусная лингвистика I**

**Планируемые результаты обучения по дисциплине для формирования компетенции.  
Индикаторы и критерии их оценивания**

**ПК.2**

**Способен преподавать по программам бакалавриата и ДПП, ориентированным на соответствующий уровень квалификации**

| <b>Индикатор</b>                                                                                                 | <b>Планируемые результаты обучения</b>                                                                    | <b>Критерии оценивания результатов обучения</b>                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                                           |
|------------------------------------------------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| <b>ПК.2.2</b><br>Осуществляет теоретико-методологическую рефлексию актуальных проблем в своей предметной области | Знает технологии создания корпуса, владеет инструментарием создания корпуса, умеет создать учебный корпус | <p style="text-align: center;"><b>Неудовлетворител</b></p> <p>Не знает технологий создания корпуса, не владеет инструментарием создания корпуса, не умеет создавать учебный корпус</p> <p style="text-align: center;"><b>Удовлетворительн</b></p> <p>Знает некоторые технологии создания корпуса, владеет некоторыми инструментами создания корпуса, умеет создавать учебный корпус, но допускает ошибки в разметке корпуса</p> <p style="text-align: center;"><b>Хорошо</b></p> <p>Знает основные технологии создания корпуса, владеет инструментарием создания корпуса, в целом умеет создавать учебный корпус</p> <p style="text-align: center;"><b>Отлично</b></p> <p>Знает технологии создания корпуса, владеет инструментарием создания корпуса, умеет создавать учебный корпус</p> |

## Оценочные средства текущего контроля и промежуточной аттестации

Схема доставки : Базовая

**Вид мероприятия промежуточной аттестации : Зачет**

**Способ проведения мероприятия промежуточной аттестации :** Оценка по дисциплине в рамках промежуточной аттестации определяется на основе баллов, набранных обучающимся на контрольных мероприятиях, проводимых в течение учебного периода.

**Максимальное количество баллов : 100**

### Конвертация баллов в отметки

«отлично» - от 81 до 100

«хорошо» - от 61 до 80

«удовлетворительно» - от 43 до 60

«неудовлетворительно» / «незачтено» менее 43 балла

| Компетенция<br>(индикатор)                                                                                       | Мероприятие<br>текущего контроля                                                                 | Контролируемые элементы<br>результатов обучения                                                               |
|------------------------------------------------------------------------------------------------------------------|--------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------|
| <b>ПК.2.2</b><br>Осуществляет теоретико-методологическую рефлексия актуальных проблем в своей предметной области | Типы и виды корпусов.<br><b>Защищаемое контрольное мероприятие</b>                               | Знание типологии корпусов, умение определять тип корпуса для исследовательских задач, владение интерфейсом    |
| <b>ПК.2.2</b><br>Осуществляет теоретико-методологическую рефлексия актуальных проблем в своей предметной области | Типы корпусной разметки.<br><b>Защищаемое контрольное мероприятие</b>                            | Знание типологии корпусной аннотации; умение определять тип разметки и тип тега, владение интерфейсом корпуса |
| <b>ПК.2.2</b><br>Осуществляет теоретико-методологическую рефлексия актуальных проблем в своей предметной области | Аннотирование корпуса.<br>Проект по созданию корпуса.<br><b>Итоговое контрольное мероприятие</b> | Знание технологий создания корпуса, владение инструментарием создания корпуса, умение создать учебный корпус  |

### Спецификация мероприятий текущего контроля

#### Типы и виды корпусов.

Продолжительность проведения мероприятия промежуточной аттестации: **2 часа**

Условия проведения мероприятия: **в часы аудиторной работы**

Максимальный балл, выставляемый за мероприятие промежуточной аттестации: **30**

Проходной балл: **13**

| Показатели оценивания                                                                                                                                                                                   | Баллы |
|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-------|
| Знает типологии корпусов, умеет определять тип корпуса для исследовательских задач, владеет интерфейсом. Представлено 10 корпусов, за полную и верную квалификацию каждого - макс. 3 балла. Ошибок нет. | 30    |

|                                                                                                                                                                                                                                                              |    |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| Частично знает типологии корпусов, хорошо умеет определять тип корпуса для исследовательских задач, хорошо владеет интерфейсом. Представлено 10 корпусов, за полную и верную квалификацию каждого -макс. 3 балла. За ошибки баллы не начисляются             | 24 |
| Частично знает типологии корпусов, частично умеет определять тип корпуса для исследовательских задач, частично владеет интерфейсом. Представлено 10 корпусов, за полную и верную квалификацию каждого -макс. 3 балла. За ошибки баллы не начисляются         | 18 |
| Не знает типологии корпусов, не умеет определять тип корпуса для исследовательских задач, не владеет интерфейсом. Представлено 10 корпусов, за полную и верную квалификацию каждого -макс. 3 балла. За ошибки баллы не начисляются. Набрано менее 13 баллов. | 12 |

### **Типы корпусной разметки.**

Продолжительность проведения мероприятия промежуточной аттестации: **2 часа**

Условия проведения мероприятия: **в часы аудиторной работы**

Максимальный балл, выставляемый за мероприятие промежуточной аттестации: **30**

Проходной балл: **13**

| <b>Показатели оценивания</b>                                                                                                                                                                                                                                            | <b>Баллы</b> |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------|
| Полностью знает типологии разметки, умеет определять тип аннотации для исследовательских задач, отлично владеет интерфейсом. Представлено 10 корпусов, за полную и верную квалификацию каждого -макс. 3 балла. Ошибок в определении типа разметки нет.                  | 30           |
| Частично знает типологии разметки, хорошо умеет определять тип аннотации для исследовательских задач, хорошо владеет интерфейсом. Представлено 10 корпусов, за полную и верную квалификацию каждого -макс. 3 балла. За ошибки баллы не начисляются                      | 24           |
| Частично знает типологии разметки, частично умеет определять тип аннотации для исследовательских задач, частично владеет интерфейсом. Представлено 10 корпусов, за полную и верную квалификацию каждого -макс. 3 балла. За ошибки баллы не начисляются.                 | 18           |
| Не знает типологии разметки, не умеет определять тип аннотации для исследовательских задач, не владеет интерфейсом. Представлено 10 корпусов, за полную и верную квалификацию каждого -макс. 3 балла. За ошибки баллы не начисляются. Набрано суммарно менее 13 баллов. | 12           |

### **Аннотирование корпуса. Проект по созданию корпуса.**

Продолжительность проведения мероприятия промежуточной аттестации: **2 часа**

Условия проведения мероприятия: **в часы аудиторной работы**

Максимальный балл, выставляемый за мероприятие промежуточной аттестации: **40**

Проходной балл: **17**

| <b>Показатели оценивания</b>                                                                                                                                                                                                                                                                                                                                                       | <b>Баллы</b> |
|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------------|
| Отлично знает технологии создания корпуса, полностью владеет инструментарием создания корпуса, умеет создать учебный корпус. Корпус представителен (10 баллов), корпус предобработан и размечен (10 баллов), аннотации соответствуют цели проекта (5 баллов); поиск по корпусу осуществляется (5 баллов); единицы аннотированы верно (5 баллов), нет технических сбоев (5 баллов). | 40           |

|                                                                                                                                                                                                                                                                                                                                                                                                                                     |    |
|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|----|
| Хорошо знает технологии создания корпуса, хорошо владеет инструментарием создания корпуса, в целом умеет создать учебный корпус. Критерии оценки: корпус представлен (10 баллов), корпус предобработан и размечен (10 баллов), аннотации соответствуют цели проекта (5 баллов); поиск по корпусу осуществляется (5 баллов); единицы аннотированы верно (5 баллов), нет технических сбоев (5 баллов). Суммарно набрано 32 балла.     | 32 |
| Частично знает технологии создания корпуса, частично владеет инструментарием создания корпуса, в целом умеет создать учебный корпус. Критерии оценки: корпус представлен (10 баллов), корпус предобработан и размечен (10 баллов), аннотации соответствуют цели проекта (5 баллов); поиск по корпусу осуществляется (5 баллов); единицы аннотированы верно (5 баллов), нет технических сбоев (5 баллов). Суммарно набрано 24 балла. | 24 |
| Не знает технологий создания корпуса, не владеет инструментарием создания корпуса, не умеет создавать учебный корпус. Критерии оценки: корпус представлен (10 баллов), корпус предобработан и размечен (10 баллов), аннотации соответствуют цели проекта (5 баллов); поиск по корпусу осуществляется (5 баллов); единицы аннотированы верно (5 баллов), нет технических сбоев (5 баллов). Суммарно набрано менее 16 балла.          | 15 |