

Модуль I. Лекция 6.  
Корпусная лингвистика

Pascual Pérez-Paredes



**Corpus Linguistics  
for Education**

A Guide for Research

ROUTLEDGE CORPUS  
LINGUISTICS GUIDES

# План

---

1. Корпусная лингвистика как раздел прикладной лингвистики
2. Понятие корпуса разметки
3. Требования к корпусам
4. Виды корпусов



# Корпусная лингвистика - часть прикладной лингвистики

## 1 Что такое корпуса разметки?

Корпуса разметки - это коллекции языковых данных с аннотациями, которые используются для исследования языка.

## 2 Требования к корпусам

Корпусам необходимо иметь большой объем данных, разнообразные тексты и точные аннотации.

## 3 Виды корпусов

Существуют разные типы корпусов, включая текстовые, параллельные, многоязыковые и специализированные.

# Роль корпусной лингвистики в изучении английского языка

1

Что такое корпусная лингвистика?

Корпусная лингвистика - это научная дисциплина, которая изучает язык на основе корпусов разметки.

2

Преимущества использования корпусной лингвистики

Корпусная лингвистика позволяет более точно анализировать употребление слов и выражений в реальных контекстах.

3

Примеры применения корпусных данных

Корпусные данные помогают лингвистам изучать языковые структуры, лексику, грамматику и даже культурные аспекты языка.



# Как использовать корпусные данные для изучения английского языка

1

Построение частотного словаря

Частотный словарь помогает понять, какие слова наиболее употребляются в разных контекстах.

2

Анализ коллокаций

Корпусы помогают выявить наиболее типичные сочетания слов и помогают избегать неправильных коллокаций.

3

Изучение грамматики и стиля

Анализ корпусов позволяет обнаруживать особенности грамматической структуры и стилистические нюансы.

# Основные вызовы использования корпусной лингвистики



Обработка больших объемов данных

Анализ корпусов требует эффективных методов обработки и интерпретации большого объема текстовых данных.

(Contd.)

#### 4. Register

- varieties of language according to use
- Stylistic functional varieties of a dialect or language
- narrowly defined by reference to subject matter (field of discourse – jargon of fishing, sports etc.) to medium (mode of discourse e.g. printed material, written, letter, message on tape etc.) or level of formality i.e. style (manner of discourse)

Учет вариативности языка

Язык подвержен постоянным изменениям, и лингвистам необходимо учитывать вариативность при анализе корпусов.



Интерпретация результатов

Точная интерпретация результатов и выводов основного значения требуется для выполнения успешного исследования.

В зависимости от характера собранных в корпусе текстов, от их разметки и некоторых других факторов различают следующие виды корпусов (табл. 2).

Наиболее важным видом корпусов является универсальный национальный корпус, создаваемый для разных национальных языков. Создание и расширение универсальных национальных корпусов представляет собой одну из важнейших задач корпусной лингвистики.

Таблица 2

Классификация корпусов [18, 13]

№	Признак	Виды корпусов
1	Форма хранения	звуковые письменные смешанные
2	Язык текстов	русский английский и т.д.
3	«Параллельность»	одноязычные двухязычные многоязычные
4	Стиль	литературные диалектные разговорные публицистические терминологические смешанные
5	Способ доступа	свободно доступные коммерческие закрытые
6	Разметка	размеченные неразмеченные
7	Характер разметки	морфологические синтаксические семантические просодические и т.д.

Универсальный национальный корпус — это собрание текстов конкретного естественного языка, представительное по отношению ко всему языку, которое может служить для исследования самых разнообразных явлений этого языка.

Общепризнанный образец универсального национального корпуса — Британский национальный корпус (BNC) ([www.natcorp.ox.ac.uk](http://www.natcorp.ox.ac.uk)). Для русского языка таким представительным корпусом является Национальный корпус русского языка (НКРЯ) ([www.ruscorpora.ru](http://www.ruscorpora.ru)). Среди корпусов славянских языков выделяется Чешский национальный корпус (<http://ucnk.ff.cuni.cz>), созданный в Карловом университете Праги. Национальные корпуса существуют также для немецкого, китайского, финского и других языков.

Одним из первых известных корпусов является Брауновский корпус (Brown Corpus), созданный в 1963 г. в Брауновском университете (США) для построения частотного словаря американского варианта английского языка. Его объем составлял 1 млн слов. Создатели корпуса (У.Френсис и Г.Кучера) разработали строгую процедуру отбора текстов: в корпус вошли 500 фрагментов прозаических текстов, созданных американскими авторами и напечатанных в 1961 г., по 2000 словоупотреблений каждый. Тексты представляли 15 наиболее распространенных жанров информативной и художественной прозы.

Поиск в корпусе в соответствии с запросом пользователя обеспечивается с помощью специальных программ — корпусных менеджеров.

Корпусный менеджер (англ. corpus manager) — это специализированная поисковая система, включающая программные средства для поиска данных в корпусе, получения статистической информации и предоставления результатов пользователю в удобной форме [6]. Результаты поиска обычно выдаются в виде конкорданса (поэтому корпусные менеджеры еще называют конкордансерами), где искомая единица представлена в ее контекстном окружении с представлением частотных характеристик отдельных языковых единиц, граммом и т.п.

Таким образом, корпус, представляющий собой размеченное собрание текстов с объемом слов не менее 100 млн, дает широкие возможности как для прикладных (работа над принципами автоматической разметки), так и для исследовательских целей.



# Заключение и выводы

## Важность корпусной лингвистики

Корпусная лингвистика имеет ключевое значение для глубокого понимания английского языка и его использования.

## Бесценный инструмент

Корпусы разметки являются бесценными инструментами для лингвистических исследований и обучения английскому языку.

## Непрерывное развитие

Развитие корпусной лингвистики продолжается, и она остается актуальной и востребованной областью изучения языка.

# Вопросы для обсуждения

1. Что такое Корпусная лингвистика?
2. Какими предпосылками определяется целесообразность и смысл использование корпусов?
3. Что понимается под значением «корпусный менеджер»?
4. Что такое «конкорданс»?
5. Расскажите про связь корпусной лингвистики и исторической лингвистики.
6. Расскажите про связь корпусной лингвистики и социолингвистики.
7. В чем заключается задача авторов корпуса?
8. Что такое репрезентативность?
9. Какие можно выделить способы деления корпусов на классы?
10. Какие типы корпусов можно выделить?
11. Расскажите подробнее про деление корпусов по критерию «параллельность».

Спасибо

за

**ВНИМАНИЕ!!!**