



Британские корпуса и словари

Какие из корпусов
авторитетны?

История и настоящее

- Какие корпуса считаются исторически важными для развития корпусной лингвистики
- Британский и Международный корпуса
- Британские корпуса и учебные словари
- Современные проблемы, поднимаемые в корпусной лингвистике на уровне международных конференций – конференция Евроколл в Кракове – Eurocall 2005 in August

Создание корпусов

- The Survey of English Usage 1960-ые
Р.Кирк 0,5 млн – 1970-ые London-Lund
Corpus of Spoken English 0,5 млн
- The Brown University Corpus of American
English – **первый
компьютеризированный корпус** – Г.
Кучера и В.Н.Франсис

Еще 6 корпусов, о которых необходимо знать

- The *Lancaster-Oslo/Bergen* (LOB) 1970-ые к Брауновскому корпусу
 - The *Collins-Birmingham International Language Database* (COBUILD) 1980-ые
- Д. Синклер в конечном счете достиг 20 млн.
- *Bank of English* 450млн. к началу 2002

Самые важные для дальнейшего развития

- *The Longman/Lancaster English Language Corpus* 30 млн. слов письм. текстов
- *The British National Corpus **BNC*** –90 млн. слов письм. текста и 10 млн. слов устного текста
- *The International Corpus of English* с. Гринбаум 1980. К 1991 году участвуют 20 стран – данные о региональном и евро английском.

British National Corpus

- Longman, Oxford University Press, Chambers Harrap, the Oxford University Computing Service, the University of Lancaster, and the British Library
- Наиболее авторитетен среди российских специалистов

Британский национальный корпус 90 млн.слов (письм.)

- информ70% 1975-1990-ые
книги (50%), периодика (20-30%)
международные дела – 15%, социально-общественные дела 15%
- творч.30% 1960-1974 (25%) и 1975-1990-ые (75%)
- 5 млн. слов письменных и неопубл. текстов

Британский национальный корпус 5-10 млн. устных текстов

- образование 25%
- бизнес 25%
- публицистика 25%
- досуг 25%

монолог 40% и диалог 60 %

Международный корпус английского языка (Британский компонент)

- сконструирован в 1991 С.Гринбаумом
- Инструмент, программа по менеджменту и анализу корпуса называется

ICE CUP – International Corpus of English.
Corpus Utility Programme

Международный корпус английского языка (Британский компонент)

- **устные тексты**
- **рукописные тексты**
- **печатные тексты**

**данные 20 стран, где английский явл.
первым или официальным вторым
языком страны**

20 стран-участниц:

- Австралия, Камерун, Канада, Карибы, Фиджи, Гана, Гон-Конг, Индия, Ирландия, Кения, Малаи, Новая Зеландия, Нигерия, Филиппины, Сьерра Леоне, Сингапур, Ю.Африка, Танзания, Соединенное королевство, США.

Британский компонент МКА

- 500 текстов из 2000 слов

Устные тексты МКА (данные 1991) 300 текстов

- Диалог 180

из них – частных 100 и 80 общедоступн.

- Монолог 120

Из них расшифров. 70 и нерасшифр. 50

тексты МКА

(данные 1991) 200 текстов

- Непечатные 50
- Печатные 150 из них академ.инфо 40, популярная инфо 40, инфо репортаж 20, инструкции 20, «убеждающ.» 10 (колонка редактора), творческие 20

Корпуса и британские учебные словари -

- OALD -Оксфордский словарь использует БНК и Оксфордскую корпусную коллекцию
- MED - Макмиллановский словарь – Мировой английский корпус
- LDCE – лонгмановский словарь Лонгман-корпусная сеть
- CIDE – кэмбриджский международный корпус

Какие международные семинары по корпусной лингвистике наиболее известны?

- Это семинары-школы Д.Синклера в его ТСЦ в Италии
- Это семинары-школы сестры Д.Синклера, автора идеи БНК-Британского национального корпуса, гл. консультанта Американского корпуса – Сью Аткинс (см. <http://www.lexmasterclass.com/>)
- В Европейской ассоциации компьютерной лингводидактики имеется специальная группа по корпусной лингвистике (см. <http://www.eurocall-languages.org/>)



Корпусная лингвистика – раздел деятельности европейской ассоциации EUROCall

- Самостоятельность обучения и корпусные данные
- Перевод и корпусные данные
- Применение корпусных данных при оценке уровня языка
- Развитие корпусных данных в педагогических целях