

КОРПУСНАЯ ЛИНГВИСТИКА

ОБЪЕКТ, ПРЕДМЕТ, ЦЕЛЬ, ЗАДАЧИ, ИСТОРИЯ СОЗДАНИЯ,
ПРИМЕРЫ



ПОНЯТИЕ КОРПУСНОЙ ЛИНГВИСТИКИ

Корпусная лингвистика - раздел языкознания, занимающийся разработкой, созданием и использованием текстовых корпусов с применением компьютерных технологий.

Термин впервые введен в употребление в **60-е годы XX века.**

Корпусная лингвистика сформировалась как **отдельный раздел** науки о языке в первой половине **90-х годов XX века.**

ЛИНГВИСТИЧЕСКИЙ КОРПУС ЭТО СОВОКУПНОСТЬ ТЕКСТОВ:

собранных в соответствии с определенными принципами


представленных в электронном виде

унифицированных

размеченных по определенному стандарту

обеспеченных специализированной поисковой системой.

ЦЕЛЕСООБРАЗНОСТЬ СОЗДАНИЯ КОРПУСОВ



**представление
данных в
реальном
контексте**

**достаточно
большая
представитель
ность данных**

**возможность
многократного
использования
единожды
созданного
корпуса для
решения
различных
задач**

ОБЪЕКТ И ПРЕДМЕТ КОРПУСНОЙ ЛИНГВИСТИКИ

Объект корпусной лингвистики – корпус текстов, который, с одной стороны, представляет собой исходный речевой материал для корпусной лингвистики и для других лингвистических дисциплин, с другой стороны, является результатом деятельности корпусной лингвистики.

Предмет корпусной лингвистики - теоретические основы и практические механизмы создания и использования представительных массивов языковых данных, предназначенных для лингвистических исследований в интересах широкого круга пользователей.

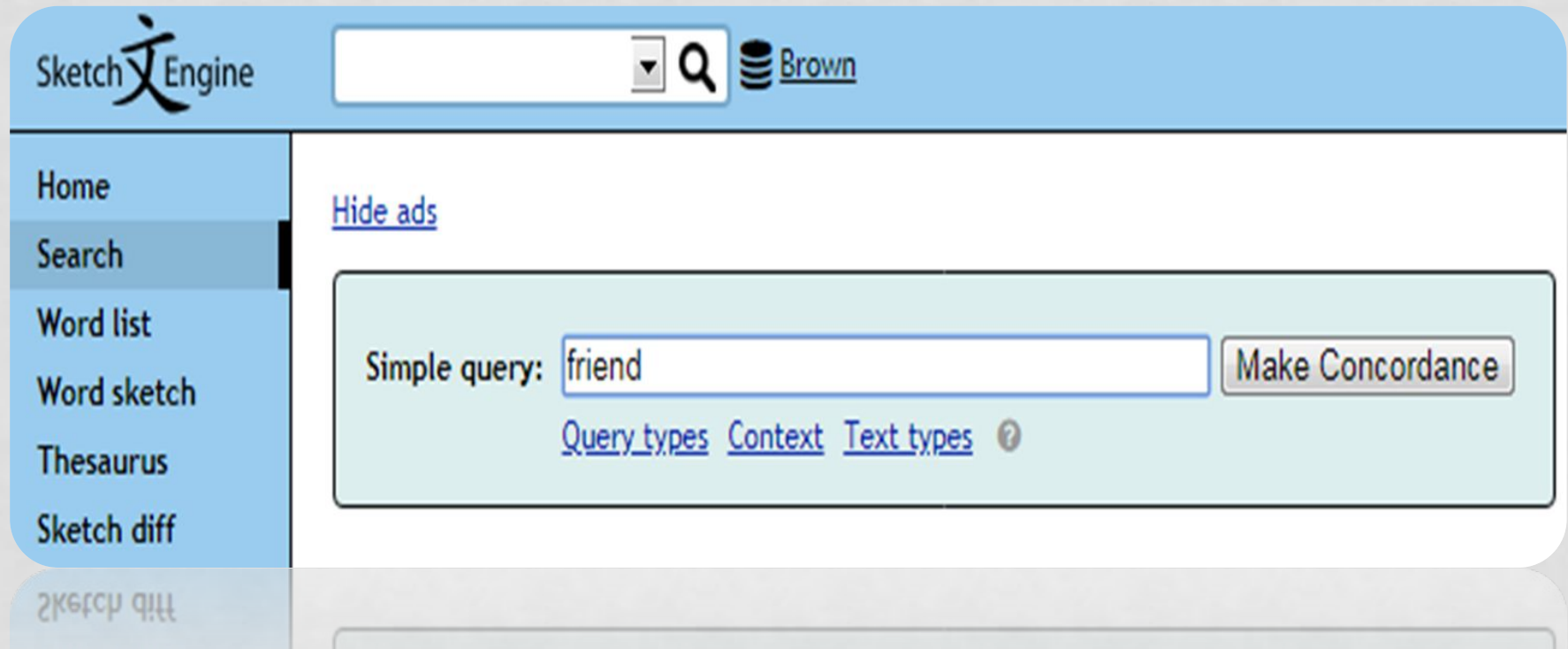
ЦЕЛЬ КОРПУСНОЙ ЛИНГВИСТИКИ - ИССЛЕДОВАНИЕ ЕСТЕСТВЕННОГО ИСПОЛЬЗОВАНИЯ ЯЗЫКА

Задачи:

- Сбор текстов с определенной целью
- Их машинная обработка
- Помощь в создании словарей (лексикографическая поддержка)
- Составление конкордансов (список встречающихся в тексте словоформ со словесным окружением, расположенных в алфавитном порядке)
- Составление частотных словарей
- Создание национальных корпусов
- Исследование использования естественного языка в разных регистрах
- Помощь в создании лингводидактических пособий

ИСТОРИЯ СОЗДАНИЯ ЛИНГВИСТИЧЕСКИХ КОРПУСОВ

- Brown Corpus - создан в 1960-е годы в Университете Брауна, содержал 500 фрагментов текстов по 2 тысячи слов в каждом.



The screenshot displays the Sketch Engine web interface. At the top left is the logo "Sketch Engine" with a stylized character. To its right is a search bar with a dropdown arrow and a magnifying glass icon, followed by the "Brown" corpus logo. A vertical navigation menu on the left includes links for "Home", "Search", "Word list", "Word sketch", "Thesaurus", and "Sketch diff". The "Search" option is currently selected. The main content area features a "Hide ads" link and a search box containing the word "friend". To the right of the search box is a "Make Concordance" button. Below the search box are several blue underlined links: "Query types", "Context", "Text types", and a help icon.

[Hide ads](#)

 Query **friend** 291 (247.52 per million) 

 Page of 15 [Next](#) | [Last](#)

[browndoc#7...](#) which not only keeps his foes guessing but his **friends** as well . Consequently , it is uncertain after
[browndoc#1...](#) probably will sit out most of the series is Bob **Friend** , who was beaten by The Man twice last season on
[browndoc#1...](#) The Man twice last season on dramatic home runs . **Friend** is off to a great start with a 4 <g/> - <g/> 0 record but is <g/> n't
[browndoc#1...](#) to see action here this week . `` We <g/> 're getting **Friend** some runs for a change , and he has been pitching
[browndoc#1...](#) Mrs. Monroe . Only the families and a dozen close **friends** will be present . The bride <g/> 's brother , Walter D.
[browndoc#1...](#) Burkes , who now live in Kankakee , are telling **friends** of the engagement . Miss Burke , a graduate of
[browndoc#1...](#) 's annual bundle party and , as all Thrift Shop **friends** know , that means the admission is a bundle of
[browndoc#1...](#) Governor , also a Somerset countian , have been **friends** since Mr. Simpkins was a child . Now 38 , Mr.
[browndoc#1...](#) , Miss Ethel Gee , 46 , and her newly devoted **friend** , Harry Houghton , 55 , and divorced . They are
[browndoc#2...](#) had the help of hundreds of neighbors and school **friends** . While neighbor women assumed some of the dead
[browndoc#2...](#) . `` I never knew there were such neighbors and **friends** around me and my family . I was <g/> n't sure there were
[browndoc#2...](#) Laotian government decided today to ask its `` **friends** and neighbors " for help in fighting what it
[browndoc#3...](#) looks pretty clean " . Ah , the fair-weather **friends** of yesteryear ! ! When I wheeled about , fanned
[browndoc#3...](#) last spring . He said he contacted a **friend** , Henry Hall Wilson , on the President <g/> 's staff
[browndoc#3...](#) to the Middle East . There are , so my biologist **friends** tell me , mechanisms of adaptation and defense
[browndoc#3...](#) of greeting . He could be lavishly generous with **friends** , cab drivers and bellboys , but with dealers he
[browndoc#4...](#) the hands of Congolese troops supposedly their **friends** . In 18 months , no more grisly incident has been
[browndoc#4...](#) are a nation of joiners , a quality which our **friends** find endearing and sometimes amusing . But it
[browndoc#4...](#) from the craft-industrial rivalry do not win **friends** for labor . Engaged as it is in a battle for world
[browndoc#4...](#) status . An excess of zeal Many of our very best **friends** are reformers . Still we must confess that

 Page of 15 [Next](#) | [Last](#)

Search

Word list


Word sketch

Thesaurus

Sketch diff

Corpus info

My jobs

User guide 

Save

Change options

Cluster

Sort by freq

Hide gramrels

More data

Less data

Sketch grammar

Translate

- Arabic

- Bulgarian

- Czech

- Dutch

- French

- German

- Italian

- Polish

- Russian

- Spanish

Menu position

friend (noun)

Brown freq = 285 (242.41 per million)

object of

30.18

meet	<u>2</u>	6.96
ask	<u>2</u>	6.79
make	<u>7</u>	6.19
become	<u>2</u>	5.85
have	<u>13</u>	5.47
see	<u>2</u>	5.41
be	<u>27</u>	4.43

subject of

18.60

continue	<u>2</u>	7.66
hold	<u>2</u>	6.49
make	<u>2</u>	4.40
be	<u>18</u>	3.85
friends are		
have	<u>3</u>	3.36

modifier

39.65

close	<u>7</u>	9.45
fair-weather	<u>2</u>	9.15
peer	<u>2</u>	8.96
intimate	<u>2</u>	8.81
neighborhood	<u>3</u>	8.75
old	<u>11</u>	8.57
old friends		

liberal	<u>2</u>	8.55
good	<u>14</u>	8.22
English	<u>2</u>	8.05
boy	<u>4</u>	7.45
few	<u>3</u>	6.86
many	<u>4</u>	6.69

and/or

24.91

relative	<u>3</u>	9.74
neighbor	<u>5</u>	9.72
anyone	<u>2</u>	7.94
school	<u>2</u>	5.86

pp of-i

10.18

mine	<u>7</u>	10.75
Heard	<u>2</u>	10.44

pp obj of-i

6.67

Many	<u>2</u>	9.23
lot	<u>3</u>	8.37

pp obj to-i

4.56

write	<u>2</u>	6.69
-------	----------	------

predicate of

1.40

stranger	<u>2</u>	9.69
----------	----------	------

pp at-i

1.40

court	<u>2</u>	7.86
-------	----------	------

ИСТОРИЯ СОЗДАНИЯ ЛИНГВИСТИЧЕСКИХ КОРПУСОВ

- **1970 годы** - частотный словарь русского языка Л. Н. Засориной, 1 миллион слов
- **1980-е годы**- русский корпус, созданный в Университете Уппсалы, Швеция, 1-1,5 миллиона слов
- **1980-е годы** - Bank of English, Бирмингемский Университет, на данный момент - [Word Banks Online](#), 500-550 миллионов слов
- **1985 год** - Машинный фонд русского языка, создававшийся по инициативе А. П. Ершова
- **1991-1994 годы** - British National Corpus, BNC, 100 миллионов слов

**THANK YOU FOR
ATTENTION!**