

*ЛИНГВИСТИЧЕСКИЕ
КОРПУСЫ: НОВОЕ
НАПРАВЛЕНИЕ ПРОЕКТНОЙ
ДЕЯТЕЛЬНОСТИ*

Львова Ольга Владимировна
доцент кафедры информатизации
образования МГПУ

ЧТО

зачем

как

Лингвистический корпус -

большой, представленный в электронном виде, унифицированный, структурированный, размеченный, филологически компетентный массив языковых данных, предназначенный для решения конкретных лингвистических задач, который характеризуется

- представлением лингвистических данных в реальном контексте, что создает возможность их всестороннего и объективного изучения;
- достаточно большой представительностью данных (при большом объеме корпуса), что гарантирует типичность данных и обеспечивает полноту представления всего спектра языковых явлений;
- возможностью многократного использования единожды созданного корпуса для решения различных лингвистических задач.

Возможности по получению из корпуса необходимой информации:

- поиск конкретных словоформ;
- поиск словоформ по леммам;
- поиск группы словоформ в виде разрывной или неразрывной синтагмы;
- поиск словоформ по набору морфологических признаков;
- отображение информации о происхождении, типе текста и т.п.;
- вывод результатов поиска с указанием контекста заданной длины;
- получение различных лексико-грамматических статистических данных;
- сохранение отобранных строк конкорданса в отдельном файле на компьютере пользователя и др.

Пользователей корпусов интересует метатекстовая информация и примеры употребления тех или иных языковых элементов и конструкций.

- частота встречаемости различных языковых элементов;
- устойчивые словосочетания;
- какие слова встречаются вместе регулярно;
- грамматические или лексические функции слов;
- связи с другими лексемами;
- иные данные по лексикографии и грамматике

В преподавания языка лексикографический анализ на базе корпусов помогает:

- раскрыть контекстное употребление тех или иных слов, особенно синонимичных (например, *small/little, big/large*);
- частотную сочетаемость одних слов с другими (например, *chance of + ing vs chance to + infinitive*);
- четко определить семантику слов (например, *made out of, made from, made up of, made of, made with*);
- проследить регулярность появления определенных слов в тех или иных стилях.

Учебные корпуса (*Learner Corpora*)

электронный корпус текстов группы лиц, изучающих иностранных язык. Основной целью организации учебных корпусов является их анализ на предмет выявления способов и эффективности освоения изучаемого языка (*Language Acquisition*).

Такие корпуса могут быть использованы для лингвистического анализа на предмет **выявления лексических или синтаксических ошибок** при освоении иностранного языка, позволяя установить частотность тех или иных **типов языковых ошибок, характерные контексты**, что необходимо для выработки **планов и методических приемов** для дальнейшей коррекции в обучении языку.

Параллельные корпуса в обучении языку и переводу (*Parallel Corpora*)

образцы профессионального перевода **художественной литературы разных жанров и стилей**, а также научно-технических, публицистических и деловых текстов.

Ресурсы, которые могли бы выступать эталонами перевода и оценке перевода в тех или иных "стандартных" условиях.

Практикум по корпусной лингвистике

- Веб как корпус
- Репрезентативность и достоверность
- Наиболее распространенные корпуса
- Навыки пользования корпусами
- Получение семантических характеристик и составление упражнений
- Получение конкорданса для определенных словосочетаний и выводы по их использованию
- Изучение стилевых различий

Сочетаемость слов

chance of + -ing vs. chance to + infinitive

a chance of + -ing

1 trictly on the book, he has little chance of beating Dandy Scot - assuming that
2 her team looked as if they had any chance of becoming championship contenders.
3 ter realises that he has as good a chance of bringing about the hoped-for mirac
4 re quite prepared to take a 50-50 chance of catching cold during the summer shr
5 eady in the dock and there was no chance of consulting him. I was in time to he
6 remained throughout, with never a chance of drying out, until the fantasticall
7 as no point in questioning and no chance of failing at had suited him. When
8 ility than 0.1, he has a moderate chance of getting it through the test but if
9 anywhere where there's the least chance of hastening your recovery." "Oh -
10 er in this respect, he has a fair chance of having it rejected. Summarizing, if
11 ools. We must see they have every chance of playing their part. There are b
12 or what she had missed and saw no chance of reaching in life. She wished she we
13 to his sister, which was that its chance of reaching its destination had seeme
14 general aim, then we stand a good chance of seeing thriving economies growing
15 ing them in the team just for the chance of winning here. We are preparing for

b chance to + infinitive

1 ould be staged. Rob did not have a chance to be alone with Gaby until after
2 d. "He wouldn't have given me a chance to do anything else. He'd have shot
3 stolen; their presence gave me the chance to do what I wanted to do." "Why
4 Grant. He might still give him the chance to escape, and if Grant did escape
5 r this gives the roots a reasonable chance to get themselves established befor
6 ht and -" "Give the poor girl a chance to get a word in!" Jean laughed. "H
7 on the television and she took the chance to go upstairs to unpack. She ha
8 , as he mischievously looked for a chance to hit back. Never strong physically
9 that British audiences will have a chance to judge this powerful creation for
10 ance of economy. It gives him a chance to make the Health Service look as i
11 be with her, because she wanted a chance to observe him with Magda in order t
12 "I know. At last you've got the chance to order other people about and you
13 alize, who ought to be given every chance to read such first-hand accounts of
14 eveloped countries. He now has the chance to recognise (if he can eat his own
15 n marrying her; he won't give her a chance to slip through his fingers again."
16 THE British public has now had the chance to take a close-up look at Dr. Hend
17 , and who have not recently had the chance to vote for a parliamentary candida



Семантика слов

a made out of

1 grey fur. STUCK on big round ears MADE OUT OF car mats. <subh> GAVE the 21-
2 high looks as though it has been MADE OUT OF white plastic; then,
3 ut beware - the later models are MADE OUT OF cheaper materials that only

b made from

1 d in the Southwest. Beverages are MADE FROM seeds the world over.
2 garden design. The paths would be MADE FROM bricks left over from the
3 m seeds the world over. Coffee is MADE FROM the roasted and ground seeds

c made up of

1 ian Huxley proposed to an audience MADE UP OF members of the British Associa
2 Association is governed by a board MADE UP OF representatives from each of
3 gave her. The engagement ring was MADE UP OF 13 diamonds and a dozen sapphi

d made of

1 ne; and since the universe is also MADE OF atoms, it is just a supermachine
2 looking-glass and candlesticks are MADE OF ivory. There are twelve pictures
3 s was doing some work. The bath is MADE OF silver. Normally kept in the

e made with

1 shampoos and conditioners are all MADE WITH natural vegetable, plant and
2 to, poached mushrooms, fruit salad MADE WITH 50g frozen raspberries, apple,
3 enience processed foods, which are MADE WITH added salt, should be avoided.



Exercise

- 1 The hospitals have an Ethical Committee, **made** doctors, nurses and members of the local community, which examines and approves proposals for research in the hospitals.
- 2 Coal used to be an important source of a whole range of organic chemicals such as phenol which is used in the preparation of dyes and drugs. Now most of these products are **made** petroleum.
- 3 Frangipane was originally a creamy filling **made** egg, butter, sugar, milk and almonds, invented in France, so they say, by an Italian called Mr Frangipani.
- 4 He had boots **made** leather.
- 5 There was a very small bunch of flowers at my great-aunt's flat, and it was **made** wire and buttons.
- 6 The vaccine was **made** cells taken from the patients' own bodies.
- 7 In the film Keanu Reeves tried to impress a girl in a club by presenting her with an origami swan **made** toilet paper.
- 8 Mirin is a naturally sweet Japanese seasoning **made** rice, water and salt.
- 9 The brain is **made** a hundred thousand million neurons, arranged in a complicated series of networks.
- 10 All items in the catalogue are **made** 100% cotton.



Comparing *I* in student and published writing

MBA: student writing

3 ed on the news on TV or from magazine, I am in the opinion that service sector i
4 ll require more consumption of fuels. I am almost certain that there will be a
7 r share of world imports composition. I believe services commodities will incre
8 rld imports composition. In the future I believe there will be a new trend in In
9 gain or loosing a million dollars. So I believe services commodities will be mo
19 factures sector. As a result of this, I can predict that there will be a decrea
23 services imports appeared. After 1987, I do not think that there was a big chang
24 of world imports about one third. But I don't think it will grow so much more b
26 duce further. As a result, therefore, I expect that the countries producing sur
27 s portion more than other commodities. I expect service industry will be a major
28 is like "services" and is intangible. I feel that the intangible value like ima
37 e to be developed in the next future. I personally see the above idea mode as a
41 es" in the world. But, before I go on, I should make a point. After these crite
42 ow. For more detail analysis, I think I should take deeper consideration and sh
45 . But from a long term point of view, I suppose the composition of world import
46 to the answer for the second question, I supposed that world imports of machiner
47 here is a declining trend. After 1987, I think that services will take a larger
49 about the same as in 1965. Since 1987, I think that the most important change wo
89 7 the trend has not change very much. I would say therefore that services and m
90 and face the original. In this point I would like to point out that the percen

JOURNALS: published writing

1 esiduals during the estimation period. I also computed Patell's (1976, p. 258) v
4 technology. The question: When should I buy? has one logical answer: NOW.
5 f returns (SVR) metric (in all cases), I choose to present only the results of t
6 th 'over the wall' statements such as 'I don't care how you do it, just do it'.
7 information environment. In addition, I examine several subhypotheses based on
8 and HA3) Size Test. In this section, I first test the hypothesis of differenti
12 a seemingly perennial question: should I invest now or wait for the 'next'
13 Marsh it was a long way from reality: 'I just did not want to be part of a big f
14 -makers keep asking themselves, 'How do I know? What evidence is there?' The othe
15 technology? By information technology I mean the hardware and software, compute
16 h as: 'If I were doing this what would I need?' Another useful heuristic reveale
18 ion required per ASR No. 190. That is, I test the hypothesis that inflation-adju
20 wo or three key questions such as: 'If I were doing this what would I need?' Ano
21 tegies both domestically and globally. I will, therefore, focus more on the impl

