

Отчёт по проделанной работе по теме: «Использование параллельного корпуса для количественного изучения лингвоспецифичной лексики».

Для выполнения данного задания мне требовалось выбрать два слова: одно лингвоспецифичное слово, а другое неспецифичное. Я выбрала два слова: «злобный» и «дыня». Прежде чем проанализировать их через сайт НКРЯ, я выдвинула гипотезу, что слово «злобный» является лингвоспецифичным, так как скорее всего его можно перевести огромным количеством слов и выражений на английский язык, в отличие от слова «дыня». Национальный корпус русского языка и формулы в самом задании помогли в подтверждении моей гипотезы.

1) Начну со слова «злобный». После того, как я вбила его на сайте, мне открылась страница, из которой видно, что: слово имеет 72 вхождения и огромное количество различных примеров. Я провела подсчёт, и оказалось, что данное слово имеет 37 различных моделей перевода, самая частотная повторяется 12 раз («malicious»), а вторая по частоте – 11 раз («evil»). Уже можно заметить, что слово имеет огромное количество моделей, что является одной из черт лингвоспецифичного слова.

The screenshot shows the search results for the word "злобный" in the National Russian Language Corpus (НКРЯ). The header includes the logo and name of the corpus. Below the search bar, there are links for navigation and settings. The main content area displays the search results for "злобный", showing 42 documents and 72 occurrences. It also provides a list of example sentences with their corresponding English translations, highlighting the word "злобный" in the Russian text. The examples include sentences from "The Da Vinci Code" and "The Carnivorous Carnival".

НАЦИОНАЛЬНЫЙ КОРПУС
РУССКОГО
ЯЗЫКА

[перейти на страницу поиска](#) [сбросить подкорпус](#) [выбрать подкорпус](#) [версия без ударений](#) [настройки](#) [формат KWIC](#)

Результаты поиска в параллельном корпусе

Объем всего корпуса: 1 953 документа, 5 844 884 предложения, 76 759 952 слова.

Поиск ведётся по пользовательскому подкорпусу объемом 1 608 376 предложений, 24 681 277 слов.

злобный

Найдено 42 документа, 72 вхождения.

Поискать в других корпусах: [основном](#), [акцентологическом](#), [газетном](#), [диалектном](#), [мультимедийном](#), [обучающем](#), [поэтическом](#), [синтаксическом](#), [устном](#).

Страницы: 1 2 3 4 5 [следующая страница](#)

1. [Dan Brown. The Da Vinci Code \(2003\)](#) [омонимия не снята] [Все примеры \(1\)](#)

en Teabing's eyes were menacing even across the hangar. "You can expect a call from my lawyers. [Dan Brown. The Da Vinci Code (2003)] [омонимия не снята] ←...→

ru Тибинг метнул в его сторону **злобный** взгляд: — Ждите звонка от моих адвокатов. [Дэн Браун. Код Да Винчи (Н. Рейн, 2004)] [омонимия не снята] ←...→

2. [Lemony Snicket. The Carnivorous Carnival \(2002\)](#) [омонимия не снята] [Все примеры \(4\)](#)

en The driver of the automobile was a man named Count Olaf, a wicked person with one eyebrow instead of two and a greedy desire for money instead of respect for other people. [Lemony Snicket. The Carnivorous Carnival (2002)] [омонимия не снята] ←...→


Чтобы до конца в этом убедиться, я использовала две формулы (меры разброса моделей перевода): «отношение абсолютной частоты самой частотной модели перевода к количеству различных моделей» и «отношение абсолютной частоты самой частотной модели перевода к частоте второй». Оценивать конечные данные после подсчёта некорректно, так как статистика в НКРЯ для данных слов неполная, а вот проанализировать саму формулу и соотношение будет логичнее.

- Первая формула рассчитывалась так: 12 к 37. Данное соотношение показывает, что на каждую модель перевода в среднем приходится сравнительно немного контекстов, а самая частотная из всех моделей занимает небольшой процент (примерно 32%) от общего числа возможных соответствий, ее частота практически не отличается от частот остальных.
- Вторая мера разброса рассчитывалась по-другому: 12 к 11. Можно увидеть, что разница между частотой самой частотной модели перевода и второй модели едва заметна, отличается буквально на один пример. Это присуще лингвоспецифичному слову, так как у него первые несколько моделей перевода должны употребляться практически одинаковое количество раз.

Слово	Количество вхождений слова	Количество моделей перевода	Самая частотная модель перевода	Вторая модель по частоте перевода
Слово "злобный"	72	37	12	11
Слово "дыня"	4	2	4	1

Формулы вычисления	Слово "злобный"	Слово "дыня"
$(F(M_{\max})) / (NumM)$	12/37	4/2
$(F(M_{\max}) / F(M_{\text{sec}}))$	12/11	4/1

2) Далее рассмотрим второе слово: «дыня». На сайте показано, что слово имеет всего 4 вхождения и маленькое количество примеров. После подсчётов оказалось, что моделей перевода всего 2, самая частотная из них употребляется 4 раза («melon»), а вторая по частоте лишь 1 раз («cantaloupe»). В этом случае также можно заметить, что наблюдается маленькое число моделей перевода, а это черта неспецифичного слова.



НАЦИОНАЛЬНЫЙ КОРПУС
РУССКОГО
ЯЗЫКА

[перейти на страницу поиска](#)
[сбросить подкорпус](#)
[выбрать подкорпус](#)
[версия без ударений](#)
[настройки](#)
[формат KWIC](#)

Результаты поиска в параллельном корпусе

Объем всего корпуса: 1 953 документа, 5 844 884 предложения, 76 759 952 слова.

Поиск ведётся по пользовательскому подкорпусу объемом 1 608 376 предложений, 24 681 277 слов.

дыня

Найдено 4 документа, 4 вхождения.

Поискать в других корпусах: [основном](#), [акцентологическом](#), [газетном](#), [диалектном](#), [мультимедийном](#), [обучающем](#), [поэтическом](#), [синтаксическом](#), [устном](#).

Страницы: 1

1. Ernest Hemingway. For Whom The Bell Tolls (1940) [\[омонимия не снята\]](#) [Все примеры \(1\)](#)

еп

All the time drinking a white wine, cold, light and good at thirty centimos the bottle. And for an end, melon. That is the home of the melon. [Ernest Hemingway. For Whom The Bell Tolls (1940)]

[омонимия не снята]

ru

И пили мы при этом белое вино, холодное, легкое и очень вкусное, по тридцать сантимто бутылка. А на закуску — **дыня**. Там ведь родина дынь. [Эрнест Хемингуэй. По ком звонит колокол (Н. Волгина, Е. Калашникова, 1968)]

[омонимия не снята]

2. John Steinbeck. The Grapes of Wrath (1939) [\[омонимия не снята\]](#) [Все примеры \(1\)](#)

- Что же касается мер разброса, то по первой формуле мы видим соотношение: 4/2. То есть самая частотная из всех моделей занимает огромный процент от общего числа возможных соответствий, чего никак не может быть у лингвоспецифичного слова.
- Вторая мера разброса также доказывает, что это слово неспецифичное. В этой формуле мы видим соотношение: 4/1. Из этого следует, что частота самой частотной модели перевода намного отличается от частоты второй модели (аж на три примера, а не как у слова «злобный»). Такого не может наблюдаться у лингвоспецифичного слова.

Выводы:

- Выдвинутые предположения и гипотезы действительно оправдались. В этом помог не только сайт НКРЯ (своими моделями перевода, примера, контекстами, частотами), но и меры разброса моделей перевода, которые давались в самом ДЗ7.
- Слово «злбный» действительно является лингвоспецифичным, это выяснилось благодаря огромному количеству моделей перевода на английском языке, а также по соотношениям, полученным по формулам (мерам разброса).
- Слово «дыня» оказалось неспецифичным, как, впрочем, и предполагалось. Маленькое количество моделей перевода, абсолютно противоположные показатели по формулам это с точностью доказали.
- НКРЯ оказался очень полезным сайтом, который способен не только выдать огромное количество различных примеров со словами на многих языках и показать большой список моделей перевода, но также разобраться в лингвоспецифичности слов, что безусловно полезно для лингвистов.