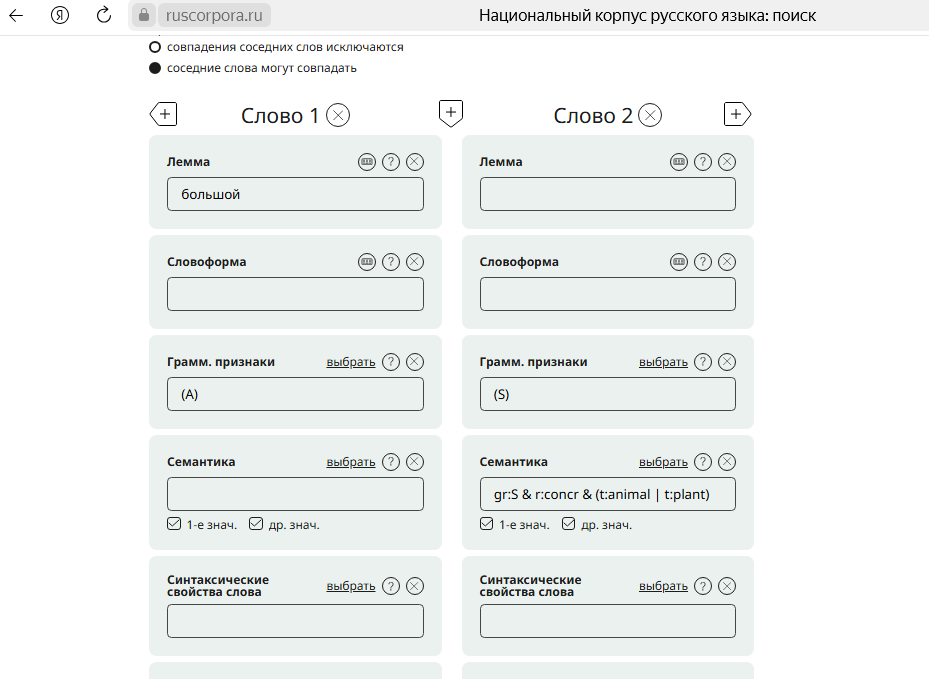
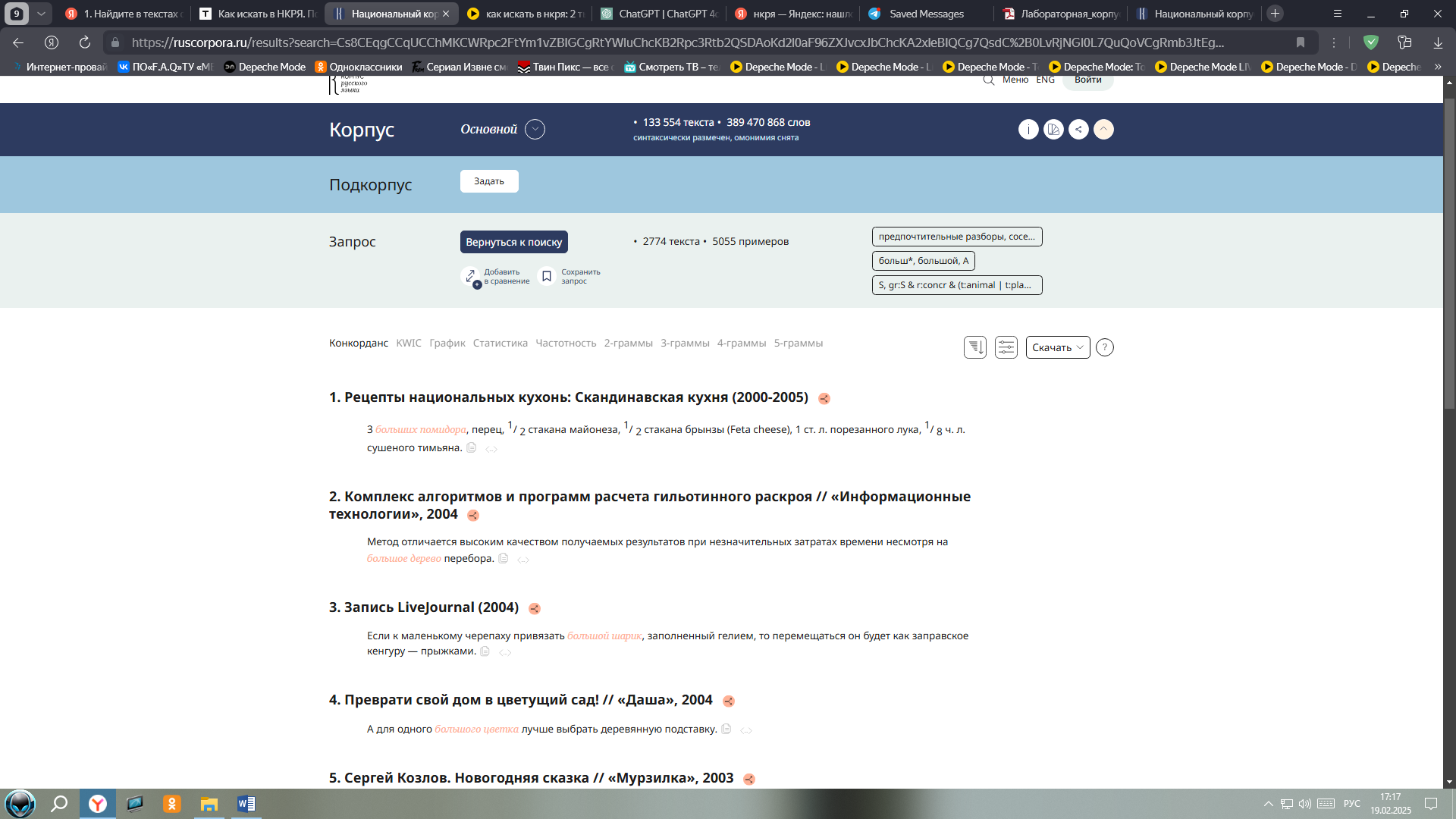
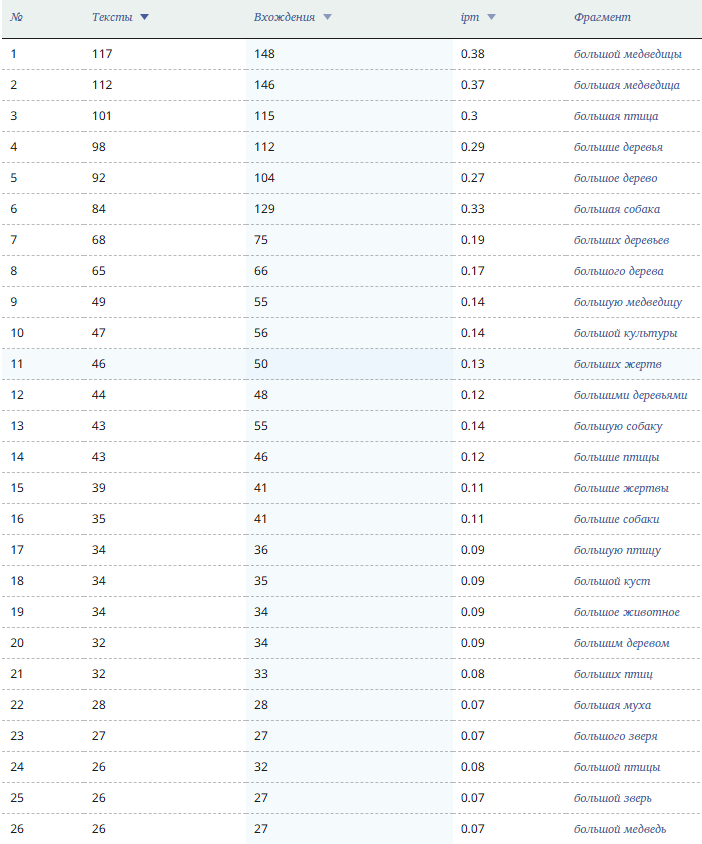
*Воропаева, 2 лабораторная, базовый уровень*

1. часть

1)Найдите в текстах основного корпуса все «большие растения и животные», т.е. первое слово должно быть прилагательным, обозначающим размер (большой), а второе слово – существительным, обозначающим растение или животное. Сколько таких вхождений?



*   
  2774 текста
* 5055 примеров

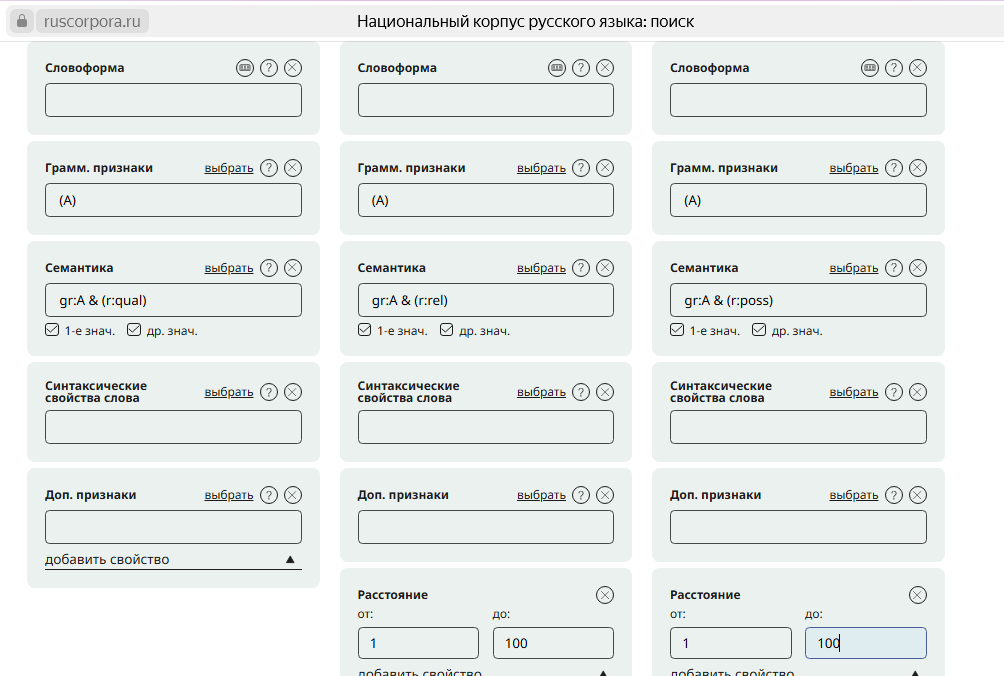








2) В скольких предложениях есть три прилагательных, относящихся одно – к разряду качественных, другое – относительных, а третье – притяжательных?

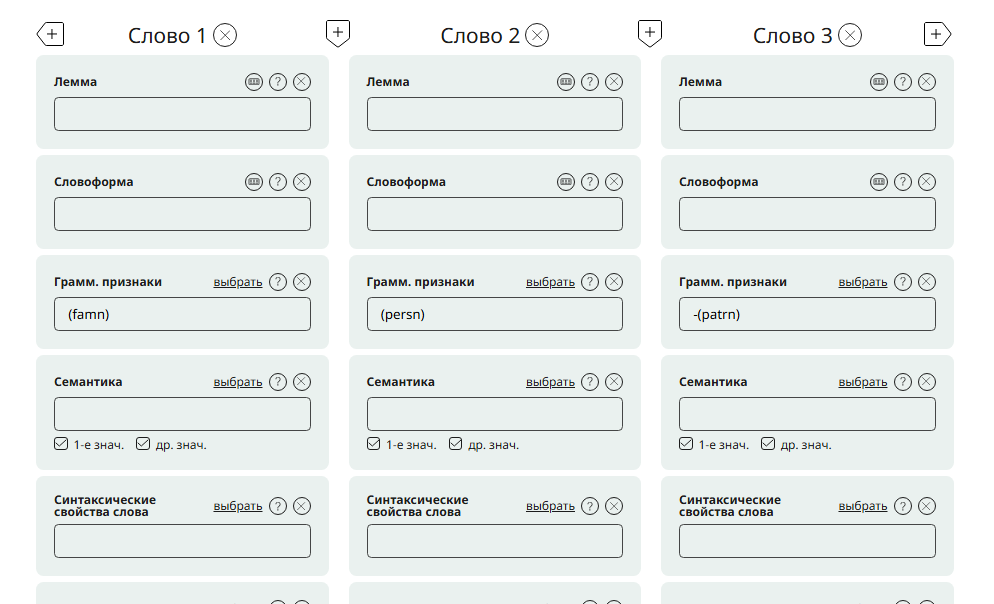


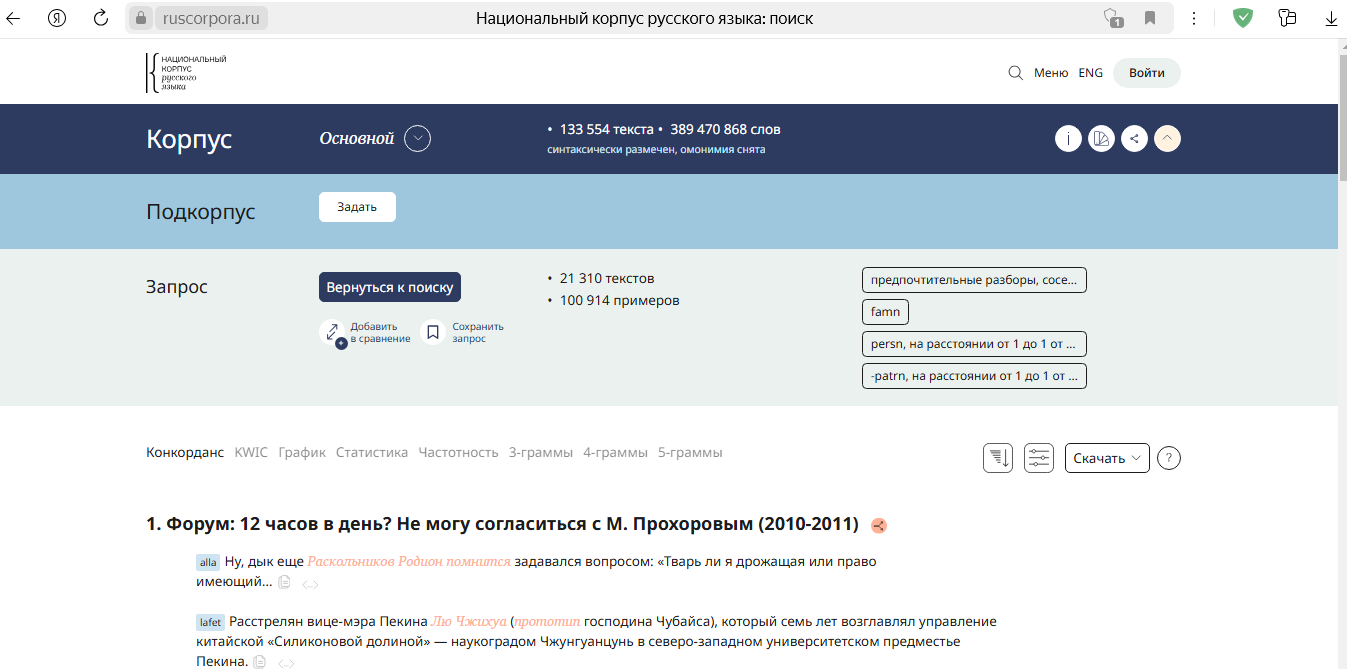


* 1056 текстов
* 2931 пример

3)В скольких предложениях после фамилии и имени нет отчества?

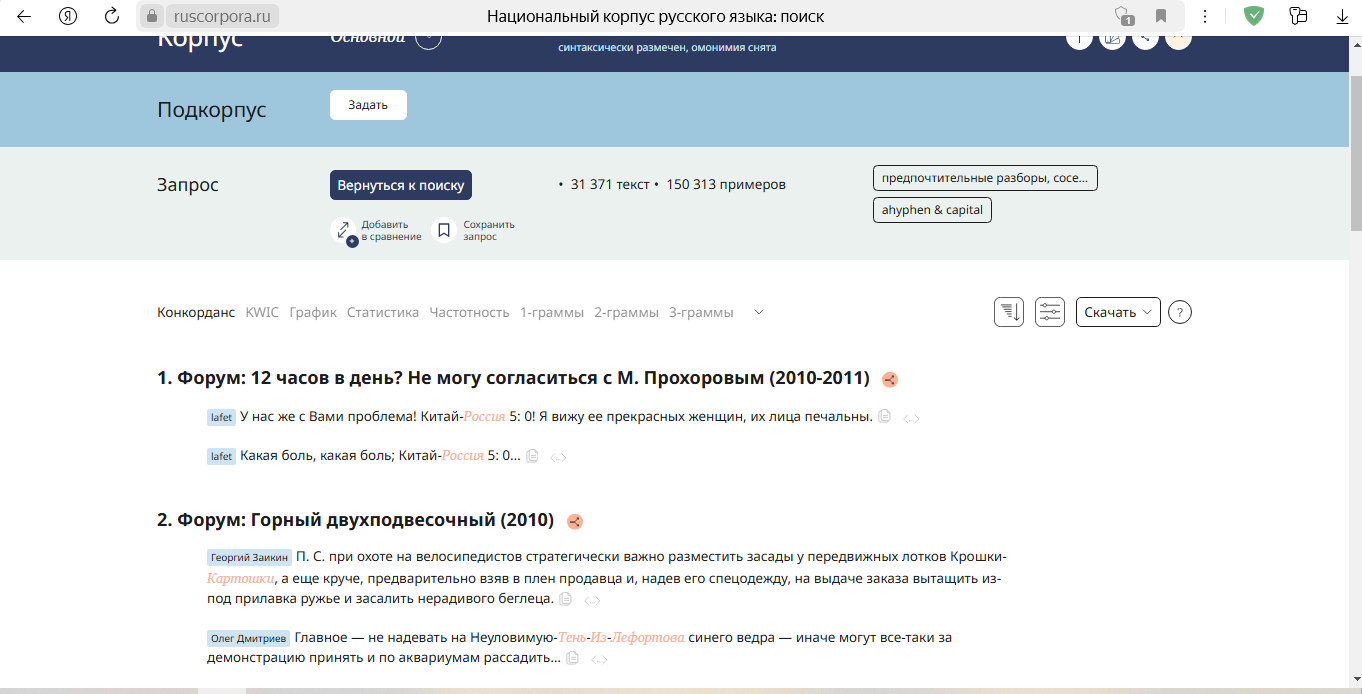
* 21 310 текстов
* 100 914 примеров



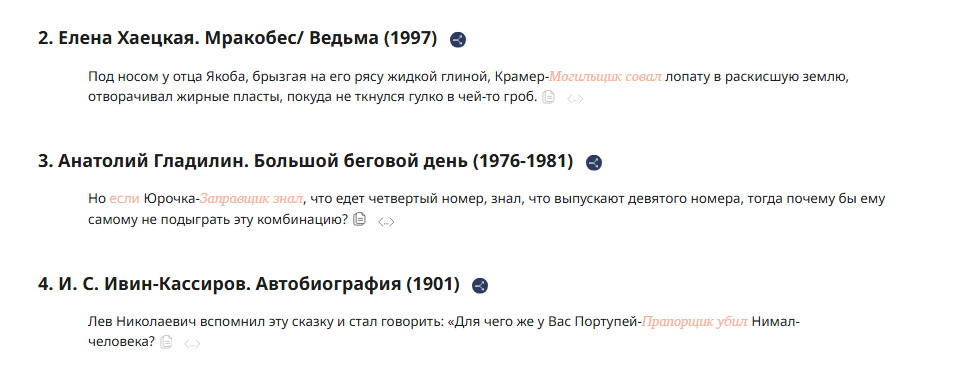


4) В скольких предложениях слово с заглавной буквы стоит после дефиса?

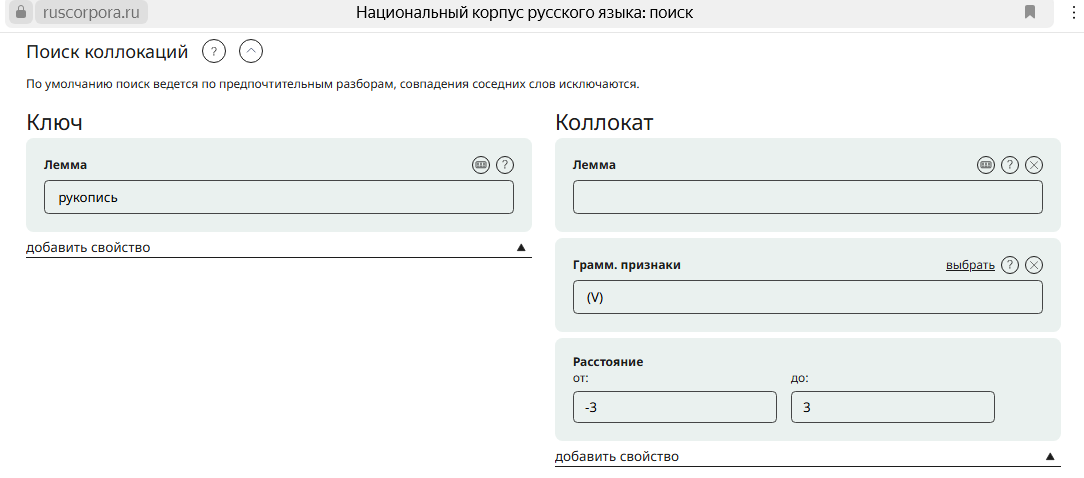
* 31 371 текст
* 150 313 примеров



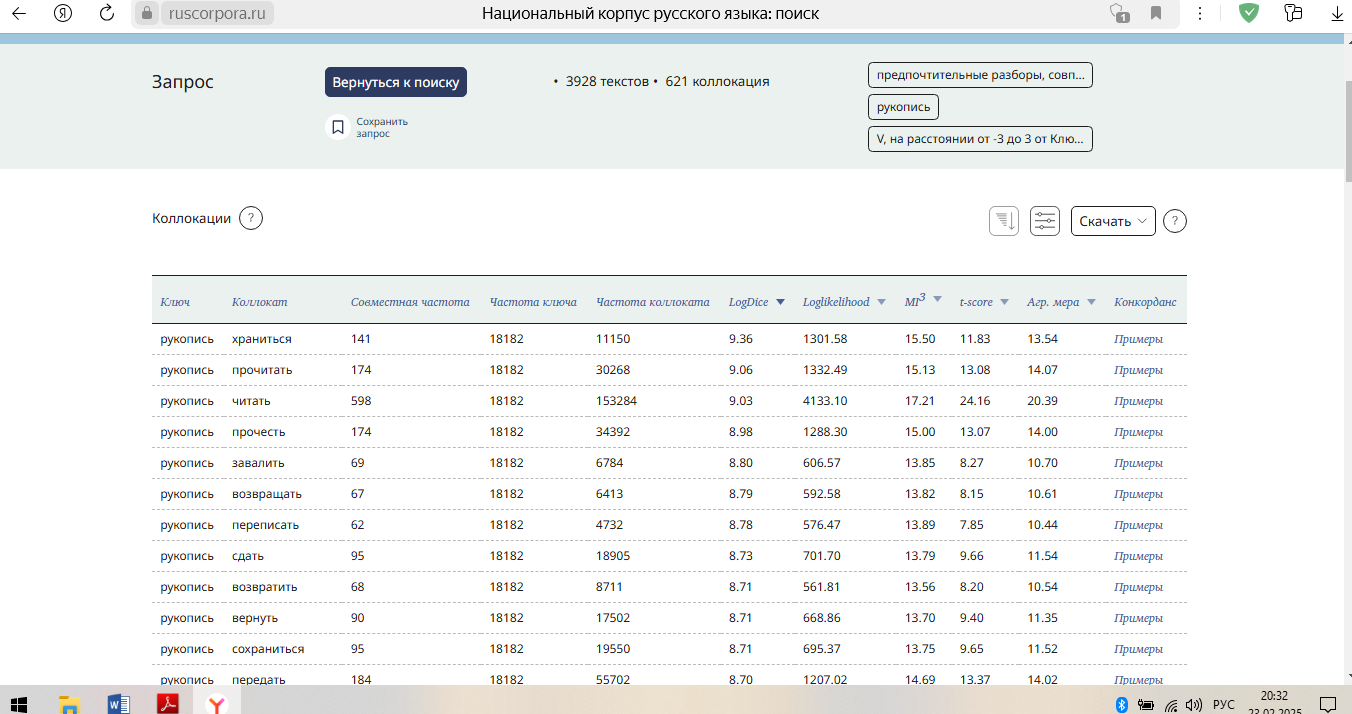
5) Сколько предложений с существительным, оканчивающимся на суффикс *-щик-*, за которым следует переходный глагол?



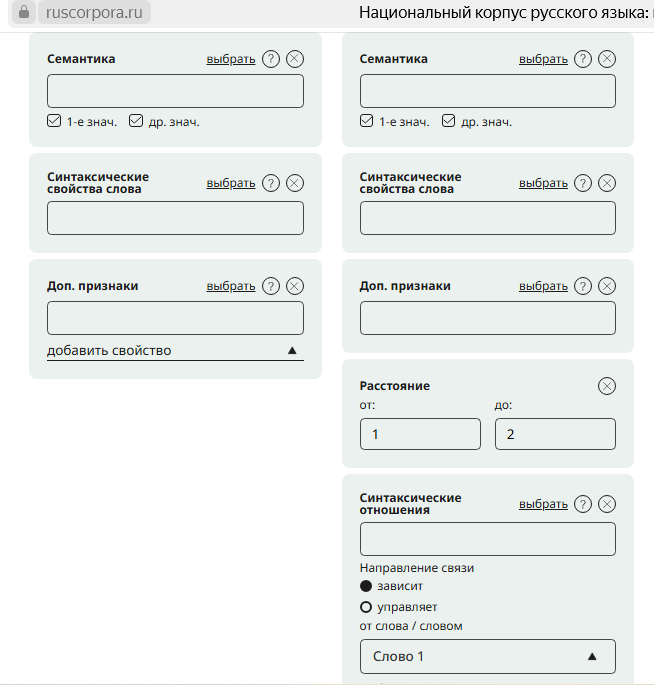
6) Какая самая частотная коллокация с существительным *рукопись*, если второе слово – глагол?

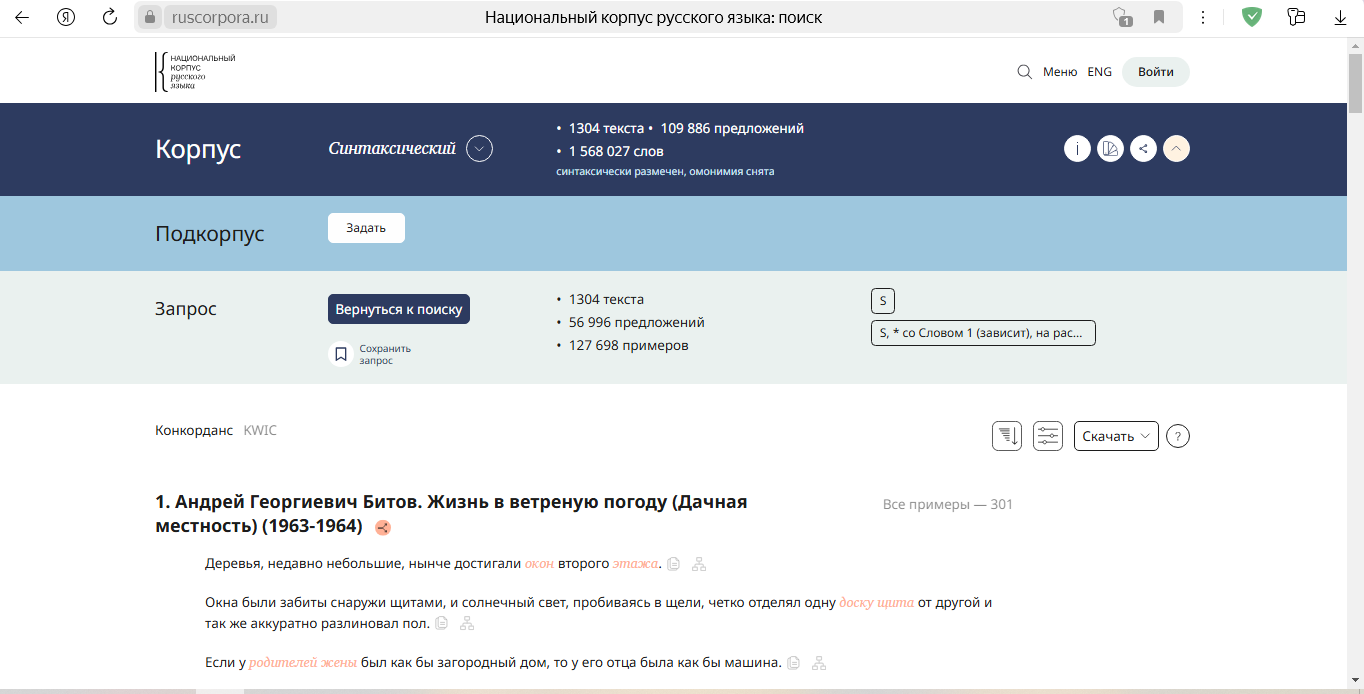


* 3928 текстов
* 621 коллокация
* самая частотная коллокация — **“храниться”**. Это означает, что глагол «храниться» наиболее тесно связан с существительным «рукопись» в корпусе.



7) Сколько в синтаксическом корпусе предложений, в которых оба слова – существительные, связанные квазиагентивным синтаксическим отношением и расстояние между словами от 1 до 2?

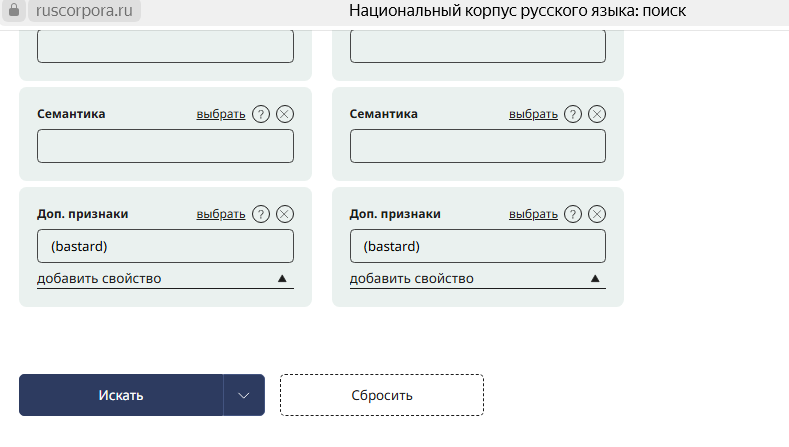


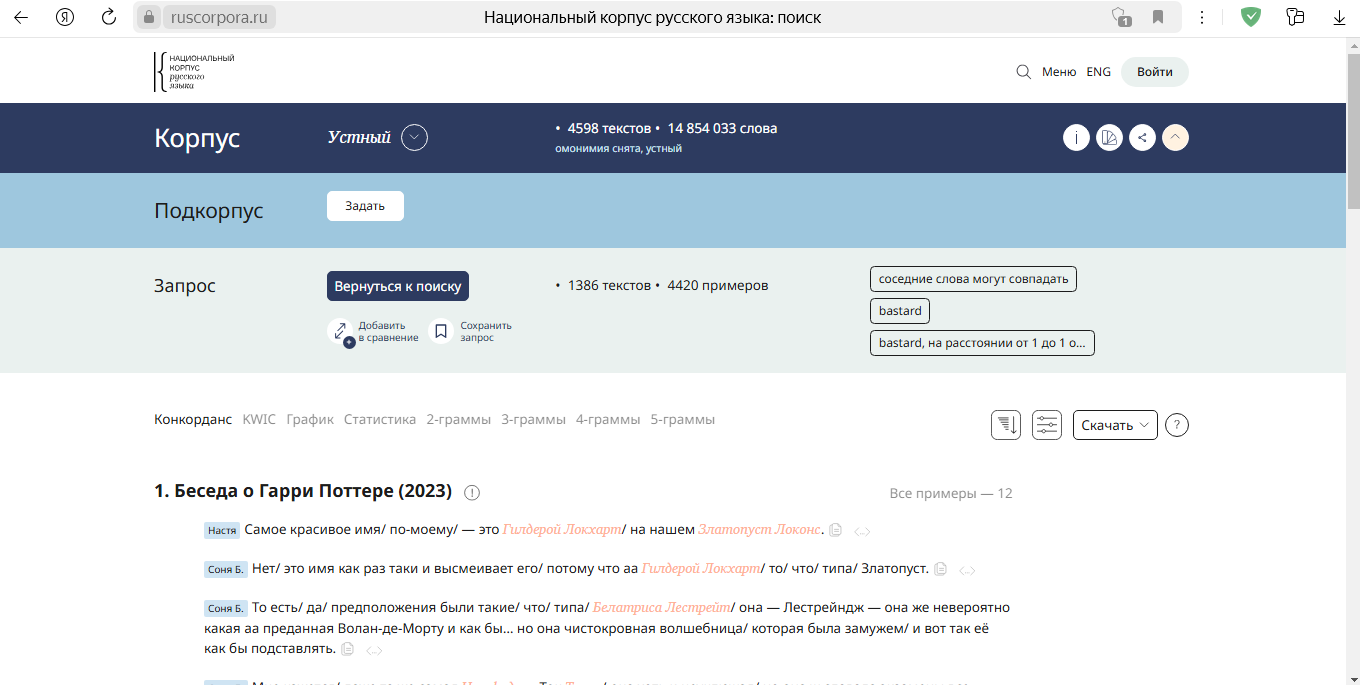


**Суть квазиагентивных отношений:** Квазиагентивные отношения (или «квазисубъектные отношения») — это такой тип синтаксической связи, при котором одно существительное (зависимое) обозначает инструмент, средство, причину или обстоятельство, с помощью которого/которой/чего совершается действие, выраженное другим существительным (главным словом).

8) Сколько в Устном корпусе предложений, в которых встречаются два идущих подряд несловарных слова?

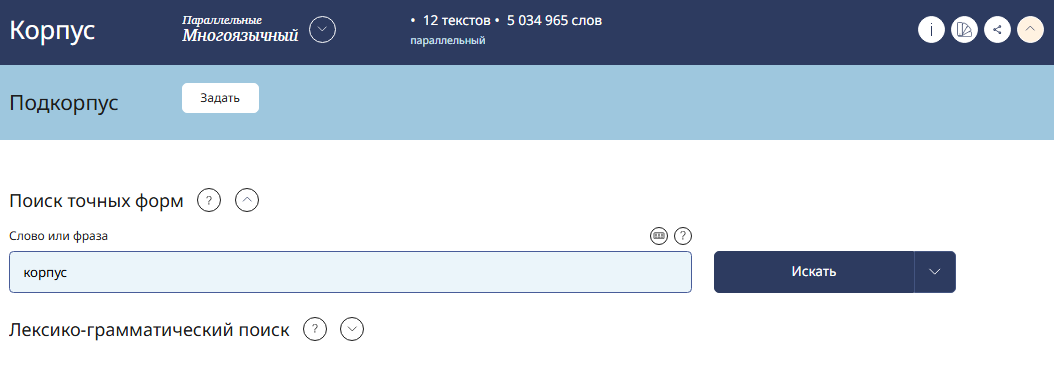
* 1386 текстов
* 4420 примеров

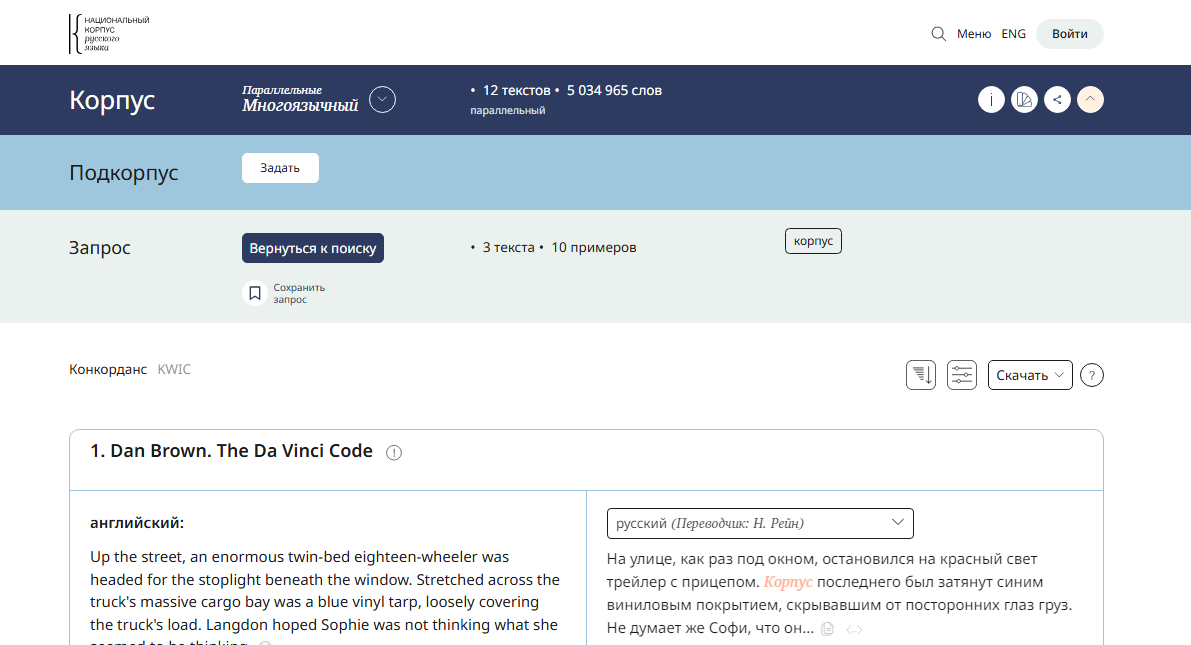




9) В скольких предложениях в Многоязычном корпусе встречается слово *корпус*?

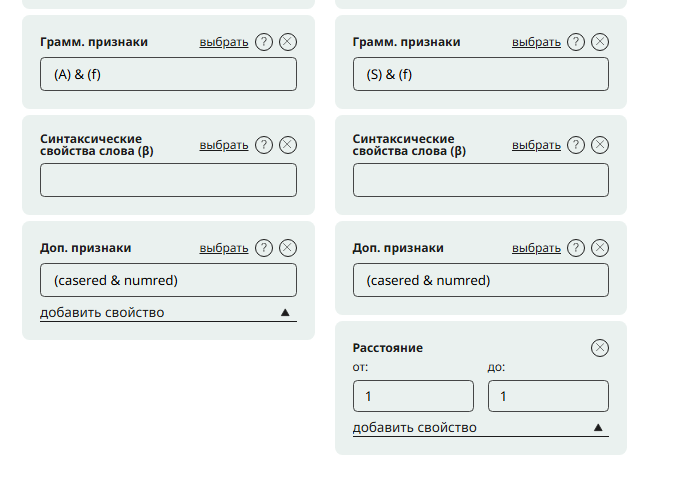
* 3 текста
* 10 примеров

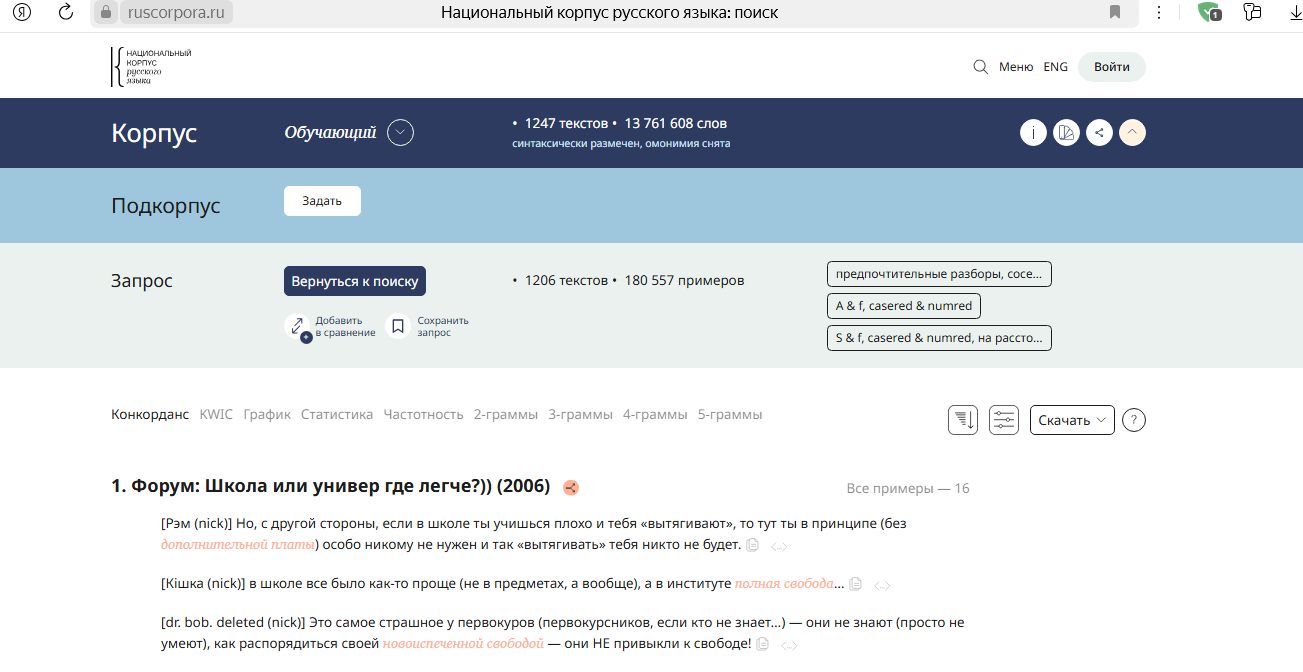




10) В скольких предложениях в Обучающем корпусе за прилагательным женского рода стоит существительное женского рода, причем они должным быть в одном и том же числе и падеже?

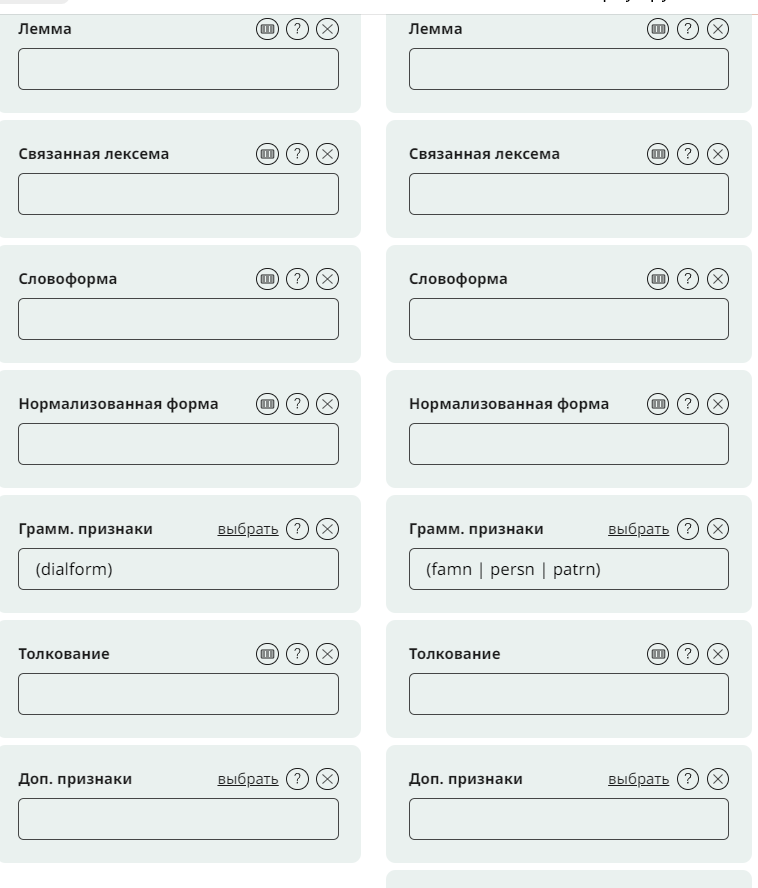
* 1206 текстов
* 180 557 примеров

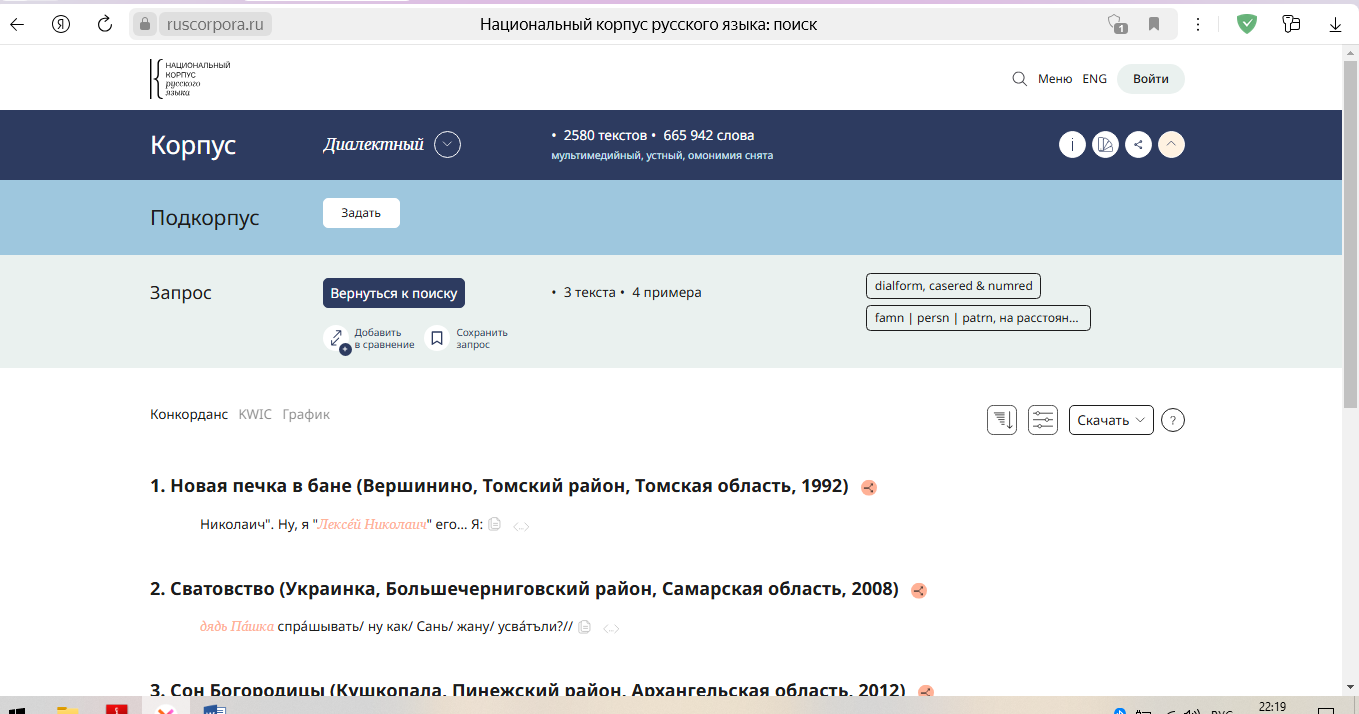




11) Сколько предложений в Диалектном корпусе с диалектными лексемами, за которыми следует антропоним?

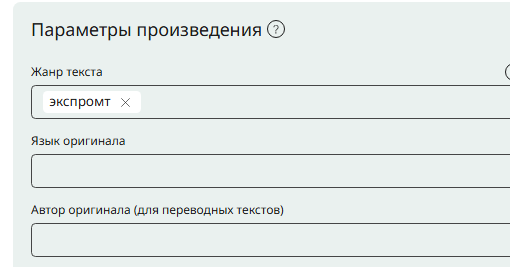
* 3 текста
* 4 примера

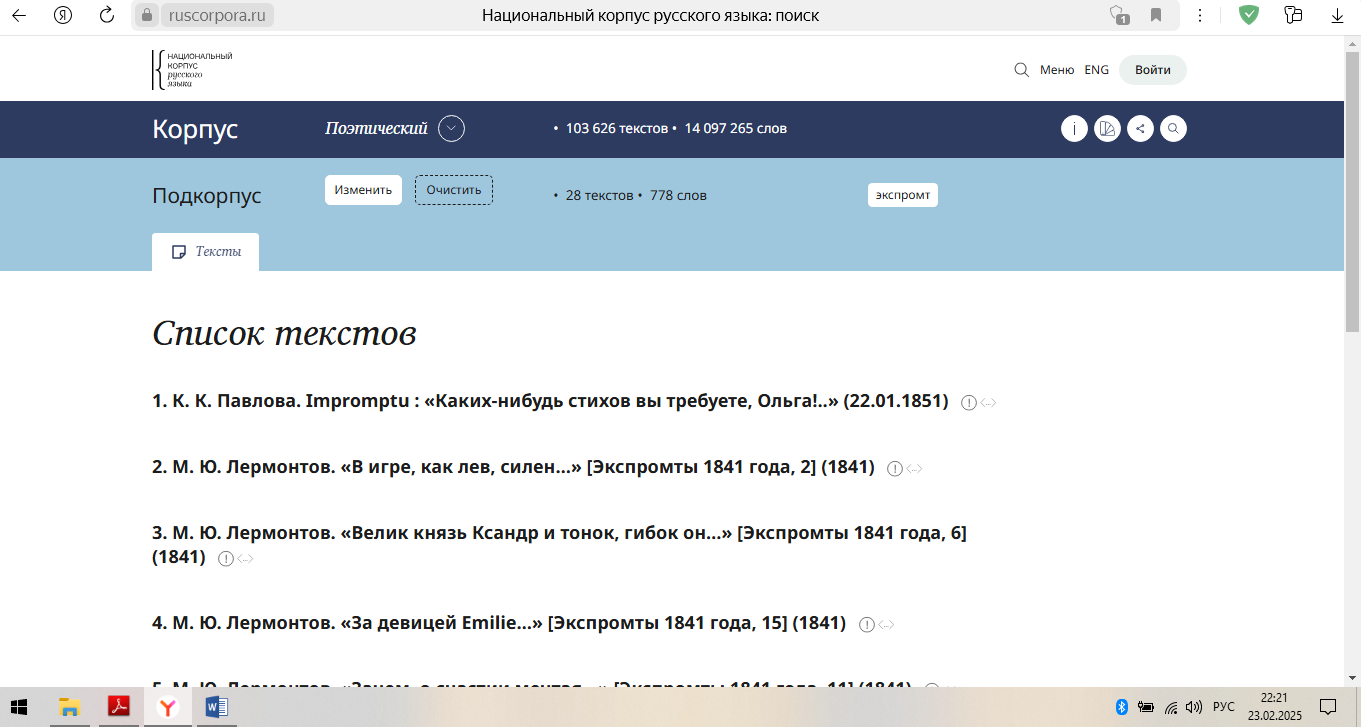




12) Сколько в Поэтическом корпусе экспромтов?

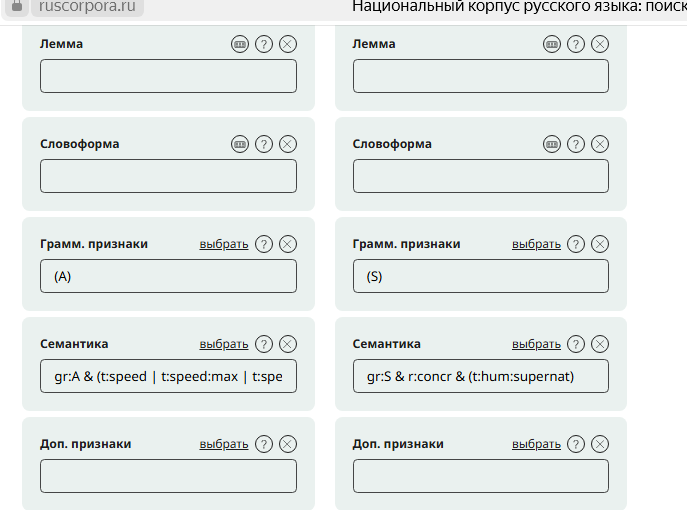
* 28 текстов
* 778 слов

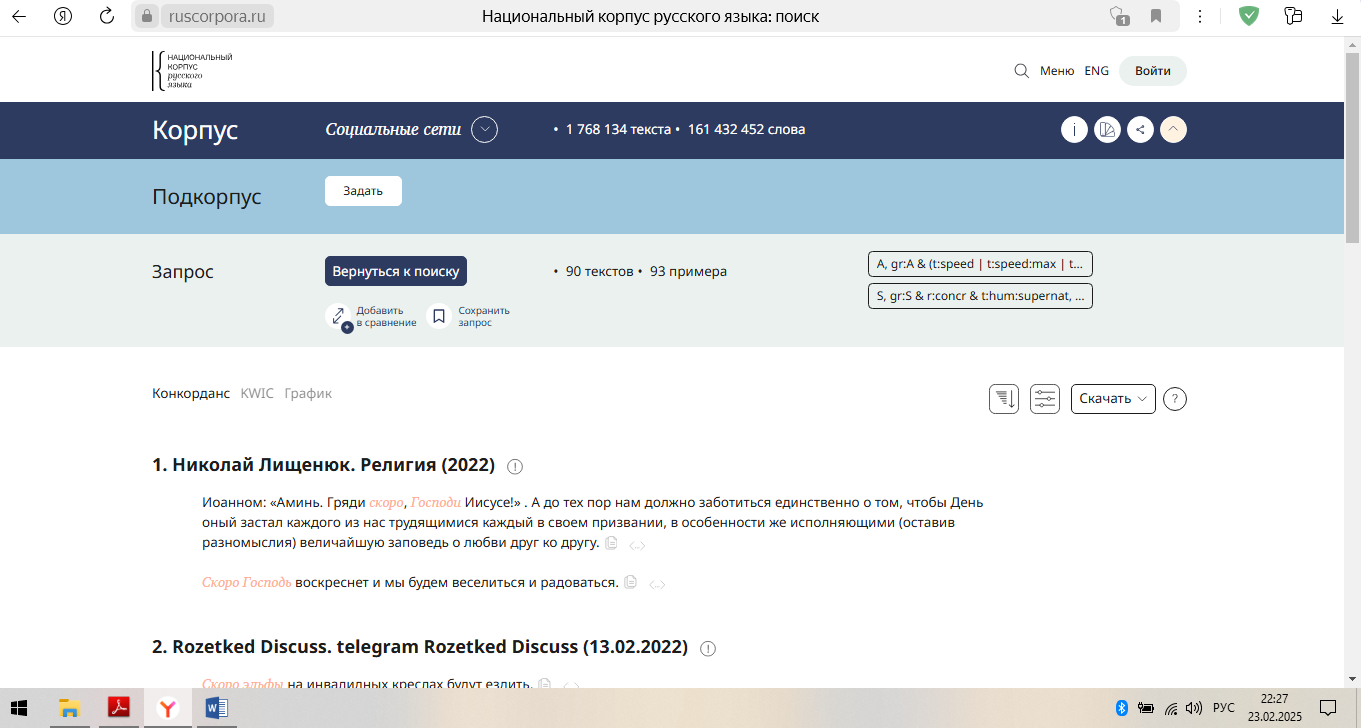




13) В скольких предложениях в корпусе Социальных сетей встречается сочетание прилагательного с семантическим признаком «скорость» и существительного, обозначающего сверхъестественное существо?

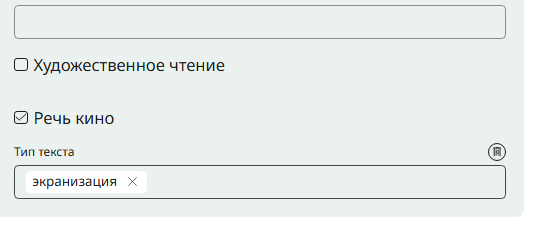
* 90 текстов
* 93 примера





14) Сколько в Акцентологическом корпусе экранизаций?

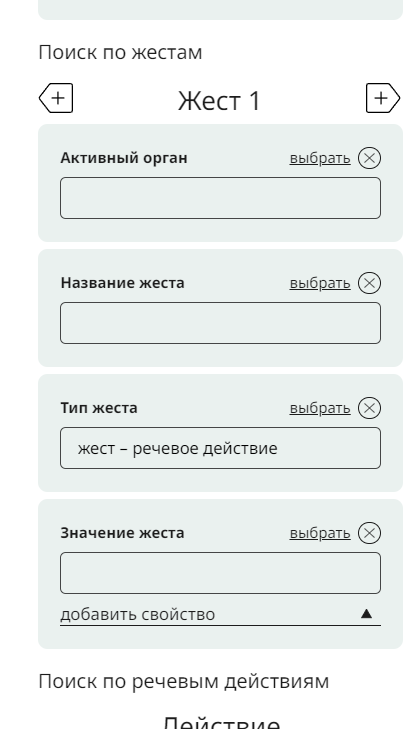
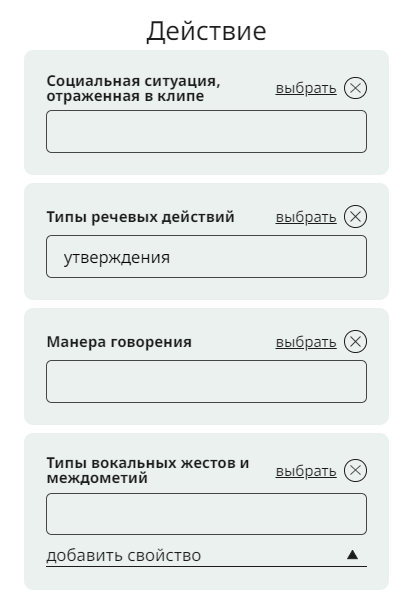
* 7 текстов
* 7 примеров

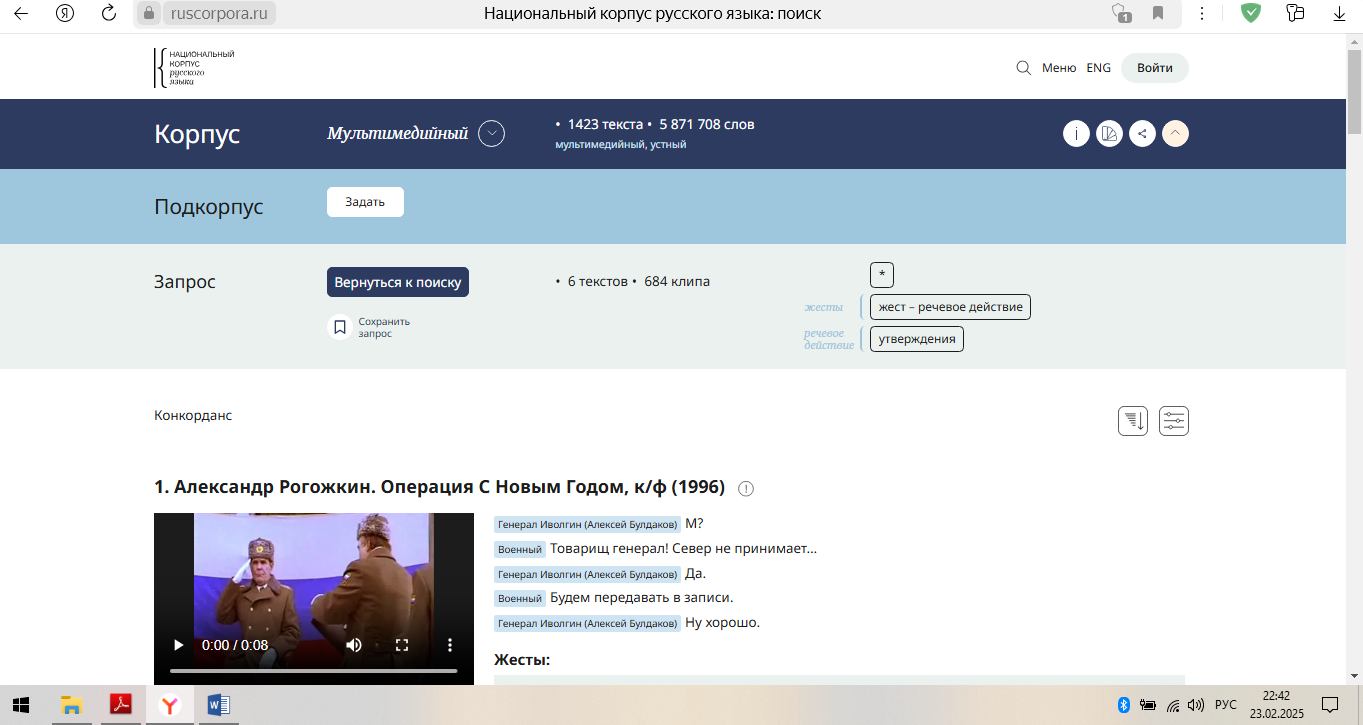




15) В скольких документах Мультимедийного корпуса встречаются сцены с утверждениями, сопровождающиеся жестами-речевыми действиями?

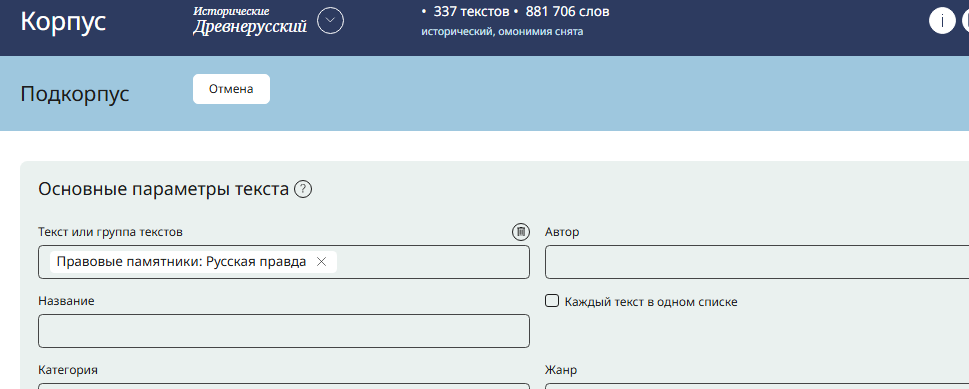
* 6 текстов
* 684 клипа



16) Сколько раз в Историческом корпусе в «Русской Правде» встречается слово *холопъ*?

* 1 текст
* 29 примеров





17) Перед вами отрывок из «Наказа Комиссии о составлении проекта нового Уложения» Екатерины 2.

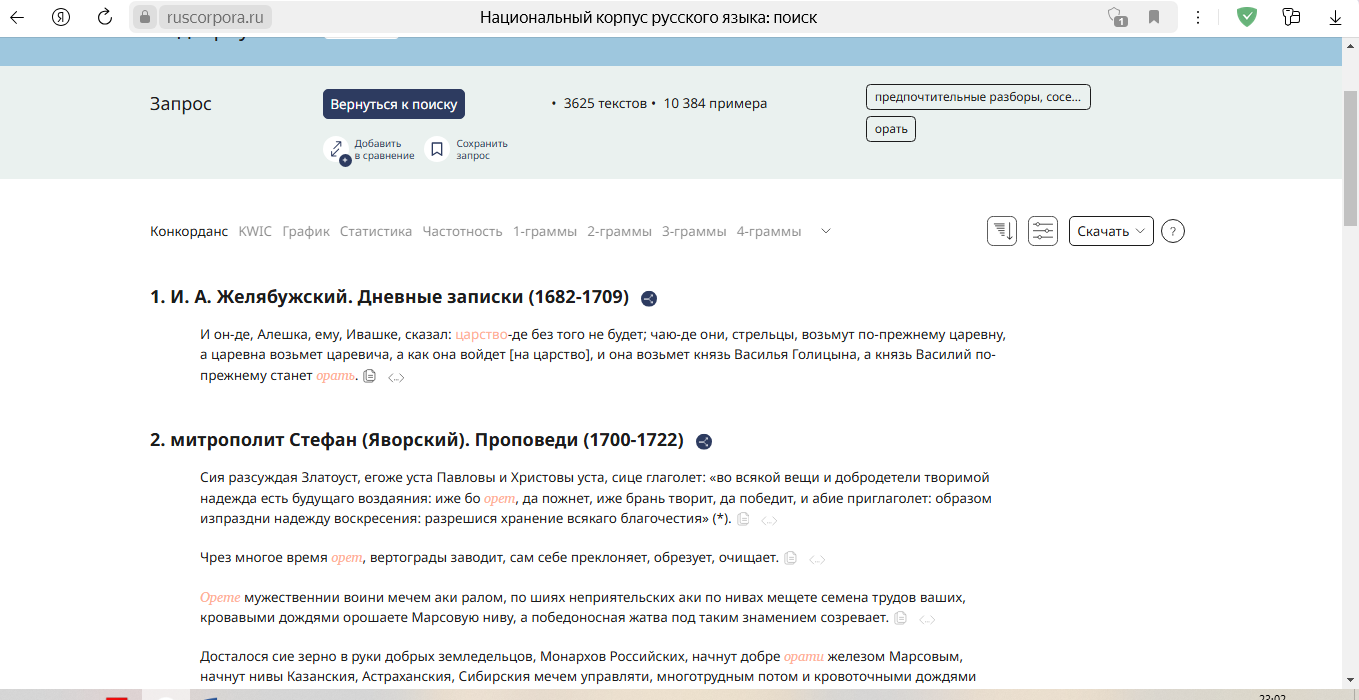
*Поля, могущие пропитать целый народ, едва дают прокормление одному семейству. Простой народ в сих странах не имеет участия и в бедности, то есть в землях никогда не оранных, которых там великое множество.* Пользуясь корпусом, выясните, какое значение имело слово *орать* до 19 века. Приведите другие примеры этого выражения из корпуса.

* Основное значение слова «орать» до XIX века — обрабатывать землю плугом, пахать. Это подтверждается множеством примеров, где «орать» используется в контексте сельского хозяйства и земледелия.
* Примеры:

«пахать землю и сажать всякое дерево» (Домострой)

“пашню орати” (Разные документы XVII века)

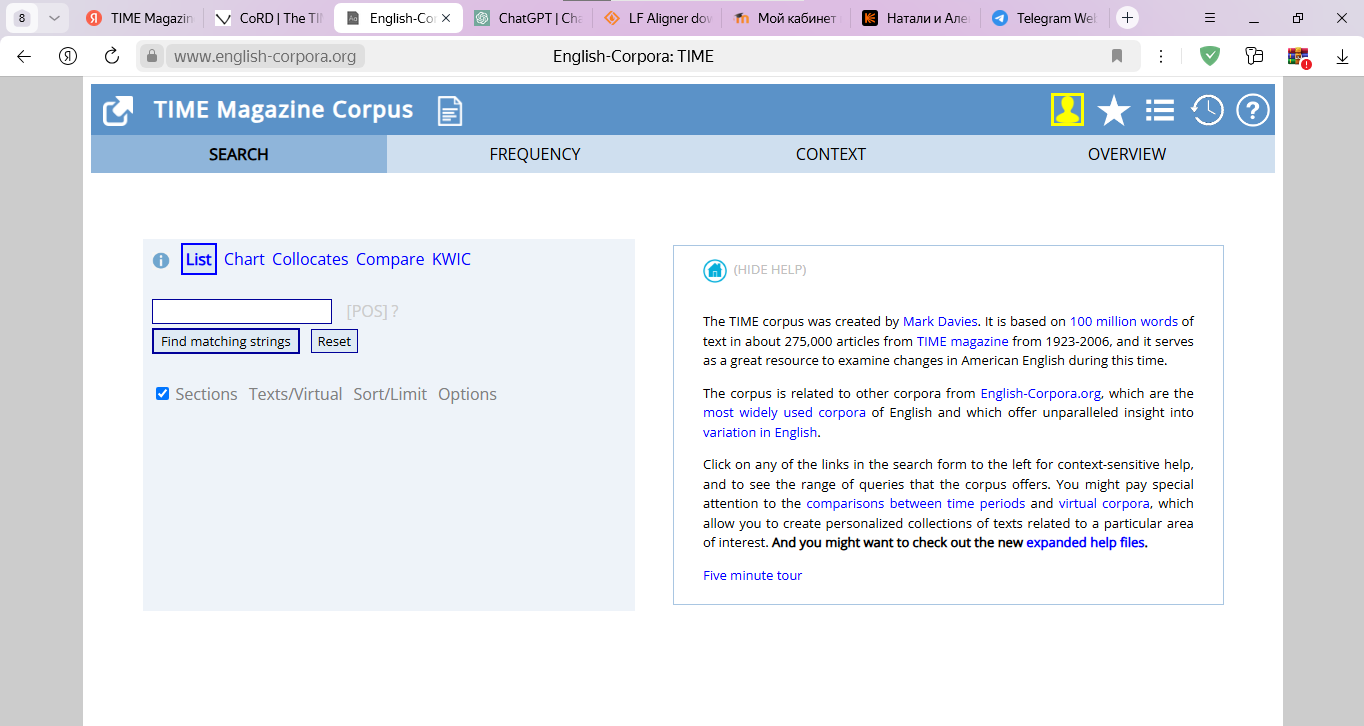
«крестьяне пахали землю» (Описание Кирилло-Белозерского монастыря)



Часть 2



**Домашняя страница проекта:** <http://corpus.byu.edu/time/>



**Корпус журнала TIME**

**Описание:** Корпус журнала TIME (Марк Дэвис, Университет Бригама Янга) — это большой корпус текстов, состоящий из всех статей, опубликованных в журнале TIME с 1923 по 2006 год. Он содержит более 100 миллионов слов и представляет собой исторический обзор американского английского языка и культуры на протяжении значительного периода времени. Важным аспектом является охват широкого спектра тем и стилей, отражённых в новостной и аналитической журналистике.

**1. Функционал TIME Magazine Corpus:**

Как и другие корпусы Марка Дэвиса, корпус журнала TIME предлагает следующие основные функции:

* **Поиск по ключевому слову (Keyword Search):** позволяет искать отдельные слова, фразы или последовательности символов в текстах журнала.
* **Поиск словосочетаний (Collocates):** эта функция определяет слова, которые часто встречаются рядом с заданным словом, помогая выявить типичные контексты его употребления.
* **Сравнение частотности (Compare):** позволяет сравнивать частотность употребления слова или фразы в разные периоды времени или в разных разделах журнала. Это особенно полезно для отслеживания изменений в языке и культуре.
* **Поиск по частям речи (Parts of Speech):** позволяет ограничить поиск словами определенной части речи (существительные, глаголы, прилагательные и т. д.) с помощью тегов частей речи (POS tags).
* **Ограничение поиска по годам:** позволяет исследовать, как использование языка менялось с течением времени.
* **Возможность просмотра контекста:** показывает результаты поиска в контексте предложений, в которых они встречаются, что важно для понимания значения и употребления слова или фразы.
* **Возможность загрузки результатов:** позволяет сохранять результаты поиска для дальнейшего анализа в других программах (например, в электронных таблицах или статистических пакетах).

**2. Последовательность действий при работе с TIME Magazine Corpus:**

1. **Получить доступ к корпусу:** Обычно требуется регистрация.
2. **Ознакомиться с интерфейсом:** интерфейс обычно состоит из строки поиска и панели управления, позволяющей выбирать параметры поиска.
3. **Выбрать тип поиска:** определить, что исследовать: частотность слова, словосочетания, изменения во времени и т. д.
4. **Ввести поисковый запрос:** ввести слово, фразу или шаблон в поисковую строку, используя соответствующие метасимволы.
5. **Указать параметры поиска:** Задать параметры поиска, такие как период времени, части речи, диапазон коллокаций и т. д.
6. **Запустить поиск:** Нажать кнопку “Search” или аналогичную.
7. **Проанализировать результаты:**
   * Просмотреть список предложений, в которых встречается ваш запрос.
   * Изучить статистику частотности и коллокаций.
8. **Загрузить результаты:** Сохранить данные для дальнейшей обработки.

**3. Сферы возможного применения TIME Magazine Corpus:**

TIME Magazine Corpus особенно полезен для:

* **Историческая лингвистика:** изучение эволюции языка и изменения значений слов с течением времени.
* **Культурных исследований:** анализ отражения социальных, политических и культурных изменений в языке. Например, можно изучать, как менялось освещение определенных тем (феминизм, расовые отношения, экология) в журнале TIME.
* **Журналистика:** изучение стилей письма и тенденций в журналистике на протяжении десятилетий.
* **Лексикография:** отслеживание появления новых слов и изменения значений существующих слов.
* **Социолингвистики:** Изучение влияния социальных факторов на язык.
* **Анализ общественного мнения:** исследование того, как журнал TIME формировал общественное мнение по различным вопросам.

**4. Метасимволы/синтаксис, используемые в корпусе текстов журнала TIME:**

Синтаксис TIME Magazine Corpus аналогичен COCA и другим корпусам Марка Дэвиса. Основные метасимволы:

* **[слово]**: Поиск точного соответствия слову (чувствителен к регистру, если не указано иное).
* **[lemma слово]**: Поиск всех форм слова (лемматизация). Например, [lemma write] найдет “write”, “writes”, “wrote”, “written”, “writing”.
* **[NN] (пример):** Поиск по частям речи. Используются стандартные теги частей речи (NN — существительное, VB — глагол, JJ — прилагательное и т. д.). Список тегов можно найти в документации к корпусу.
* **\***: Подстановочный знак для любого слова. Например, the \* of найдет «значение», «важность», «историю».
* **?**: Подстановочный знак для одного символа.
* **[] (квадратные скобки с пробелами)**: поиск любого из перечисленных слов. Например, [man woman] найдет либо “man”, либо “woman”.
* **{слово}**: Поиск синонимов (использует встроенный тезаурус). Например, {good} найдет “good”, “excellent”, “fine” и т.д.
* **Year: from YYYY to YYYY (или просто Year: YYYY)**: Ограничение поиска определенным годом или диапазоном лет.

**Примеры запросов:**

* [lemma technology]: Найти все формы слова “technology”.
* [JJ] war: Найти все прилагательные, стоящие перед словом “war”.
* [NN] {computer} Year: from 1950 to 1970: Найдите существительные, связанные с понятием «компьютер» (используя тезаурус) в период с 1950 по 1970 год.

**Важные особенности:**

* **Исторический охват:** это главное преимущество корпуса. Он позволяет отслеживать изменения в языке и культуре на протяжении длительного периода времени.
* **Тематическое разнообразие:** журнал TIME охватывает широкий спектр тем, что делает его корпус данных полезным для исследований в различных областях.
* **Ограничения:** важно помнить, что корпус представляет собой только один источник — журнал TIME. Он не является репрезентативным для всего американского английского языка.

В заключение можно отметить, что TIME Magazine Corpus — это ценный ресурс для лингвистов и исследователей, интересующихся историей английского языка и культуры. Его функциональность и метасимволы позволяют проводить широкий спектр исследований, а его исторический охват делает его уникальным инструментом для изучения изменений во времени.

Часть 3

1. Выберите два слова, находящихся в одном семантическом поле (парадигме, где все компоненты имеют хотя бы один общий семантический признак, например, *движение*: *бег*, *плавание*, *ходьба…*): гипотетически лингвоспецифичное и неспецифичное. Используя вышеперечисленные параллельные корпусы (Параллельные корпуса НКРЯ, OPUS, Linguee) и меры разброса, подтвердите или опровергните свои гипотезы относительно их лингвоспецифичности или ее отсутствия.

Семантическое поле — это группа слов, объединённых общим смыслом, общей темой.

Лингвоспецифичность означает, что какое-то понятие, выраженное словом, тесно связано с культурой, историей или образом жизни конкретного народа. Иными словами, это понятие может быть трудно точно перевести на другие языки, потому что в других культурах нет точного аналога.

* **Пример лингвоспецифичного слова:** русское слово «тоска». Это не просто грусть или печаль, а особое сочетание тоски, душевной боли, неопределённой тревоги и стремления к чему-то большему. Прямого эквивалента в английском или немецком языке нет.
* **Пример неспецифичного слова:** слово «солнце». Солнце есть везде, и во многих языках есть простые и понятные слова для его обозначения (sun, Sonne, soleil).

**Пример: слова “гостеприимство” и “уют”**

* **Семантическое поле:** оба слова относятся к сфере *отношений между людьми, дому и образу жизни*. Их общий семантический признак — создание благоприятной атмосферы для человека.
* **Лингвоспецифичность (гипотезы):**
  + **«Уют»**: я предполагаю, что слово «уют» является более специфичным с лингвистической точки зрения. «Уют» — это не просто комфорт, а нечто большее: ощущение тепла, защищённости, душевного покоя, связанное с домом, семьёй, традициями. В других культурах может не быть одного слова, которое точно передавало бы это ощущение.
  + **“Гостеприимство”**: я предполагаю, что слово “гостеприимство” менее специфично с точки зрения языка. Принимать гостей — распространенная практика во многих культурах, поэтому во многих языках должны быть слова для обозначения этого понятия.

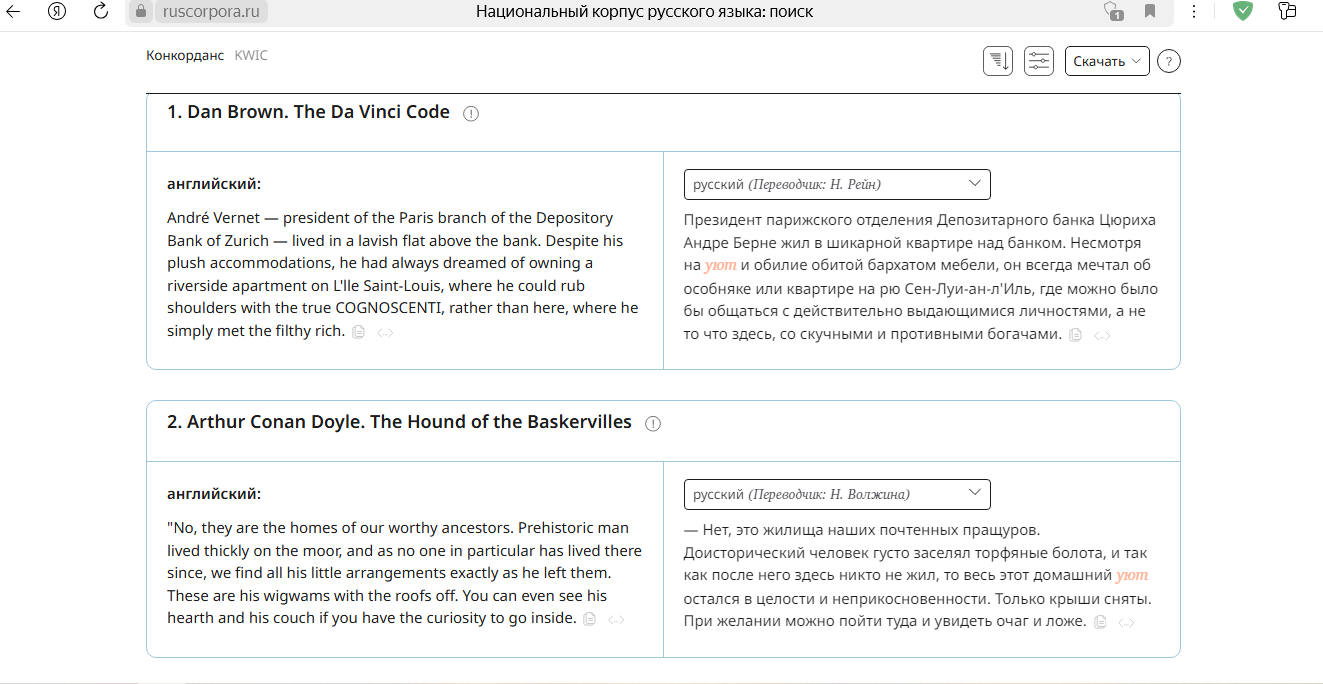
**Ожидаемые результаты и примеры:**

* **“Уют”**:
  + **Английский:** Возможные переводы: “coziness”, “comfort”, “snugness”, “atmosphere of warmth and comfort”, “homeliness”. Видим, что нет одного идеального слова, и часто используются описательные фразы.
  + **Немецкий:** “Gemütlichkeit” (близкое по значению, но также имеет свои культурные коннотации), “Behaglichkeit”, “Wohnlichkeit” (больше относится к физическому комфорту).
  + **Французский:** “confort”, “intimité”, “ambiance chaleureuse” (приходится использовать сочетания слов).
  + **Пример:** В параллельном корпусе находим, что фразу “В доме было очень уютно” переводят на английский как “The house had a very warm and welcoming atmosphere”. Это значит, что переводчик использовал описательный перевод, так как не нашел одного точного слова.
* **“Гостеприимство”**:
  + **Английский:** “hospitality”.
  + **Немецкий:** “Gastfreundschaft”.
  + **Французский:** “hospitalité”.
  + **Пример:** Фраза “Русское гостеприимство известно во всем мире” переводится на английский как “Russian hospitality is known throughout the world”. Здесь используется прямое соответствие.

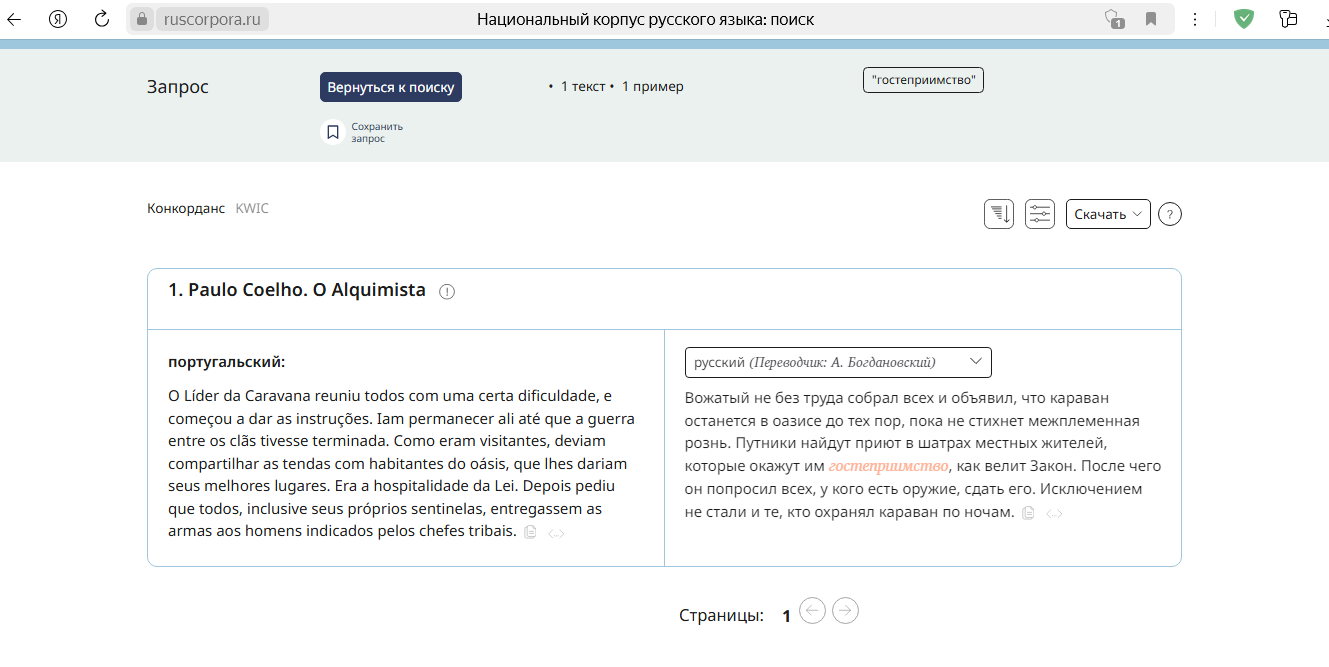
**Вывод:**

Если наш анализ покажет, что слово “уют” действительно требует более сложных переводов и часто заменяется описательными фразами, а слово “гостеприимство” имеет прямые и понятные эквиваленты во многих языках, то это подтвердит нашу гипотезу о том, что “уют” является более лингвоспецифичным понятием.

1)Уют:



В этих примерах слово «уют» заменяется другими фразами

2)Гостеприимство:  


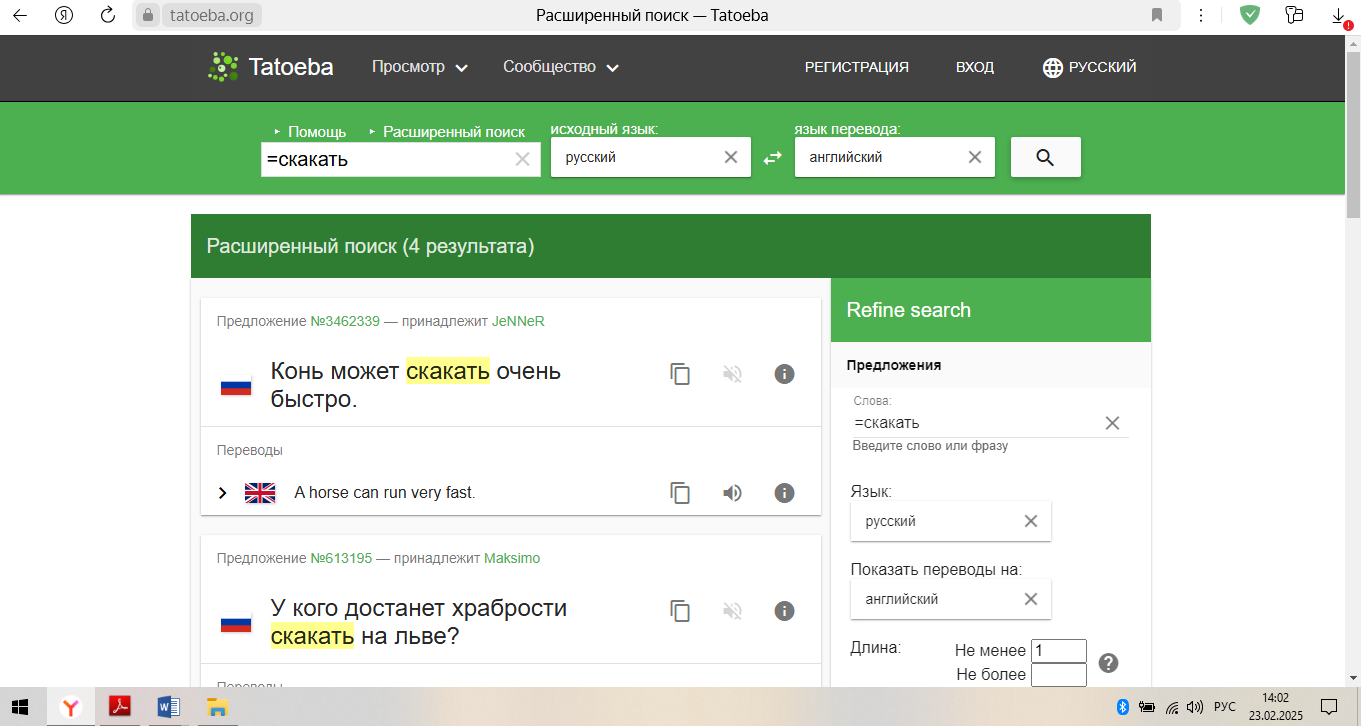


Всё оказалось так, как я и предполагала.

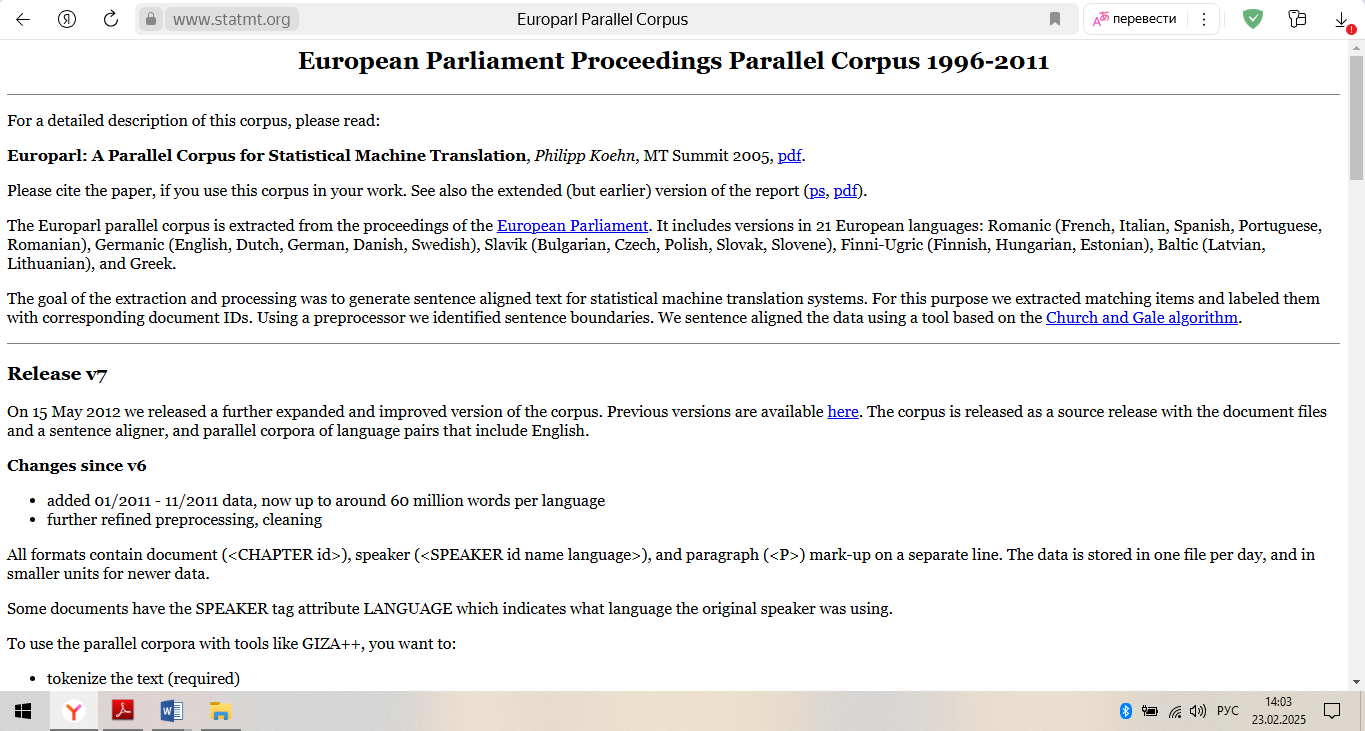
**2. Примеры аналогичных ресурсов с параллельными текстами:**

Вот 3 примера ресурсов с параллельными текстами, аналогичных перечисленным:

1. **Tatoeba:** (tatoeba.org) — это большая база данных предложений, переведенных на множество языков. Она ориентирована на изучение языков и содержит примеры употребления слов и фраз в контексте. Хотя это и не корпус текстов в строгом смысле, она предоставляет параллельные переводы предложений.



1. **Europarl:** ([www.statmt.org/europarl/](http://www.statmt.org/europarl/)) — это корпус, состоящий из протоколов заседаний Европейского парламента. Он доступен на 21 европейском языке. Это ценный ресурс для исследований в области политической лингвистики и машинного перевода.





1. **OpenSubtitles:** ([www.opensubtitles.org](http://www.opensubtitles.org/)) — этот корпус содержит субтитры к фильмам и телепередачам, переведённые на множество языков. Он полезен для изучения разговорной речи и культурных особенностей разных стран.

