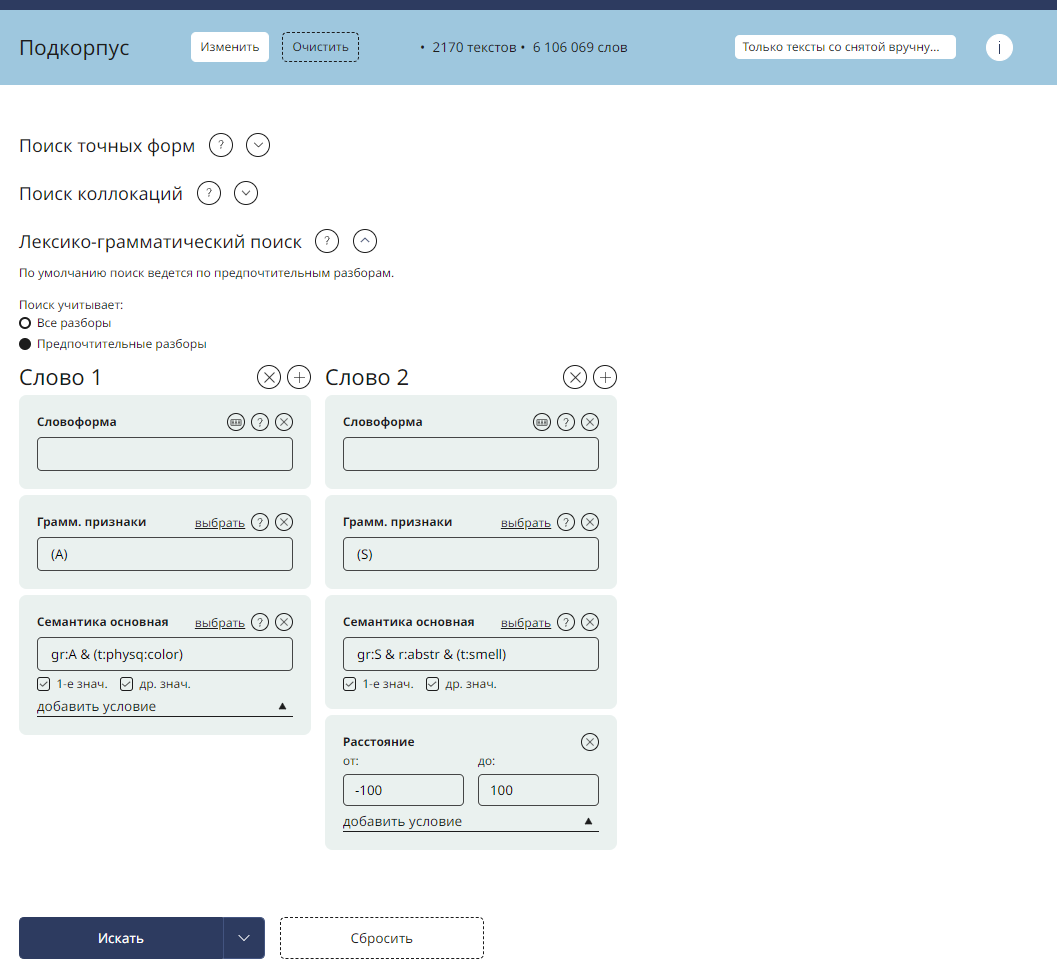
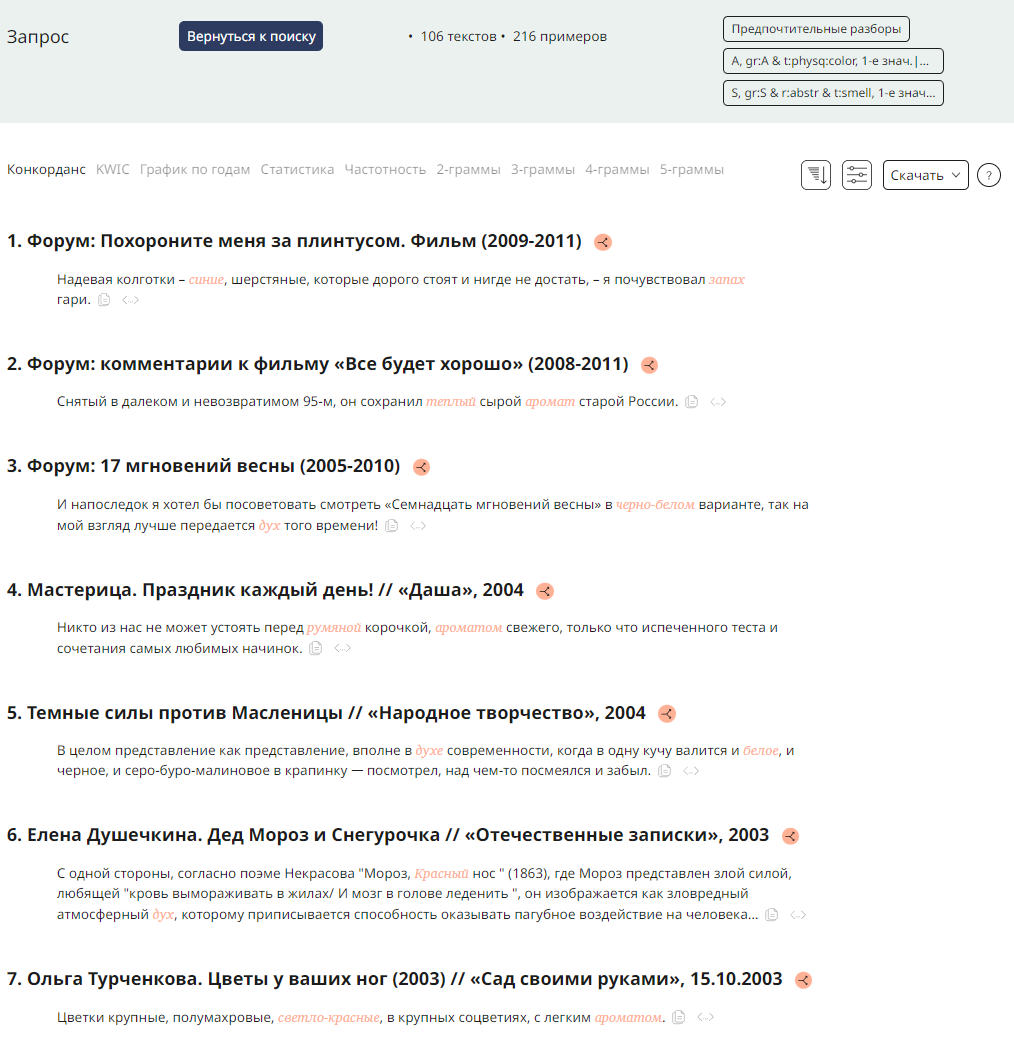
Уровень 1

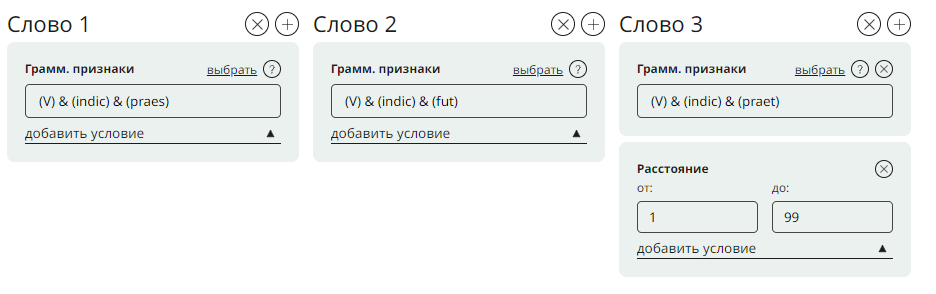
2) Зададим условие для поиска:



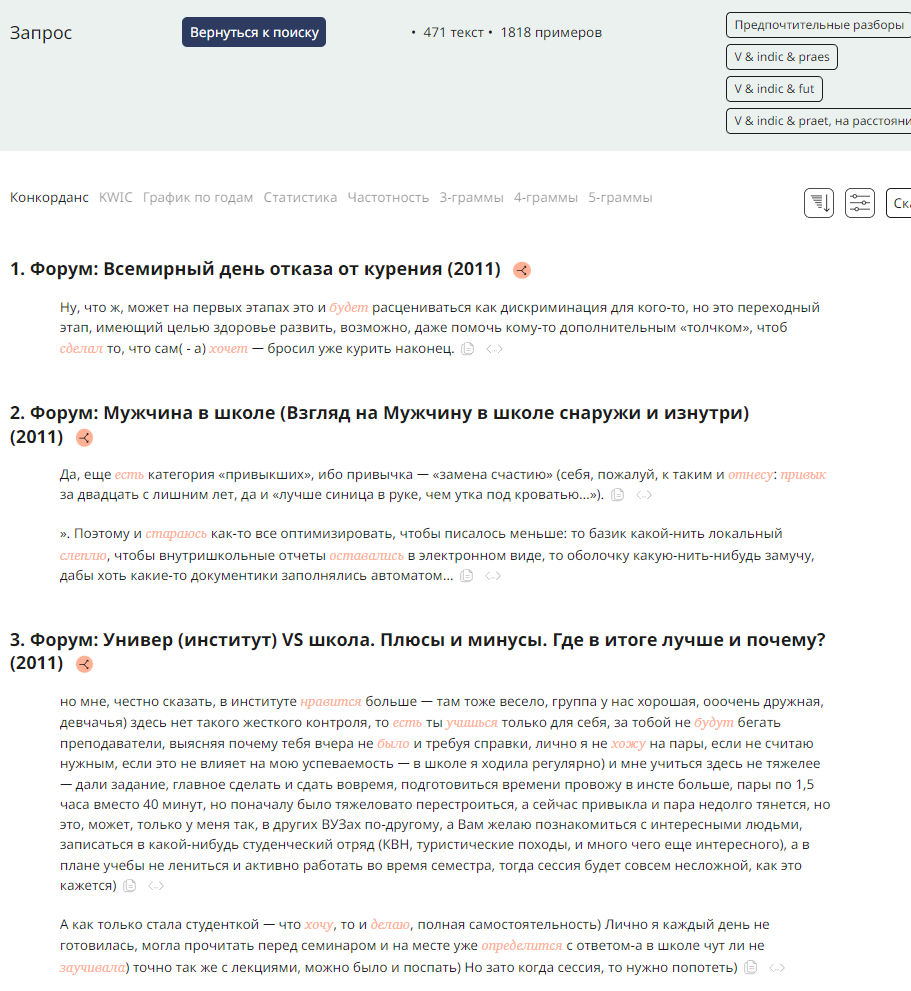
И количество вхождений:

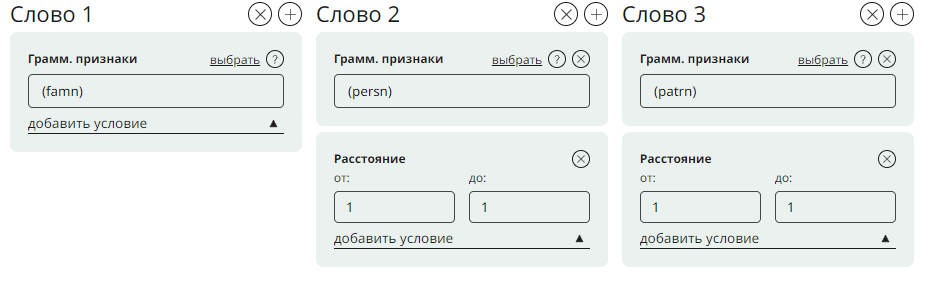


3) Далее найдём число вхождений трёх глаголов изъявительного наклонения разных времён. Условие:



Сами вхождения:

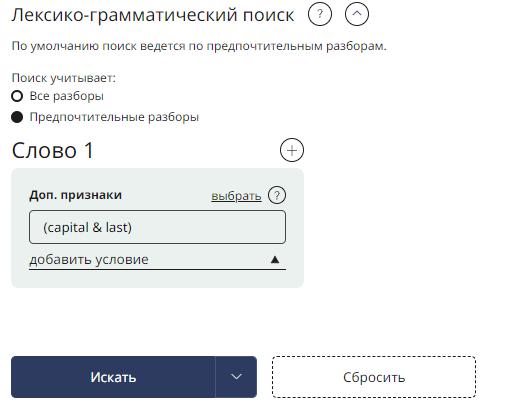


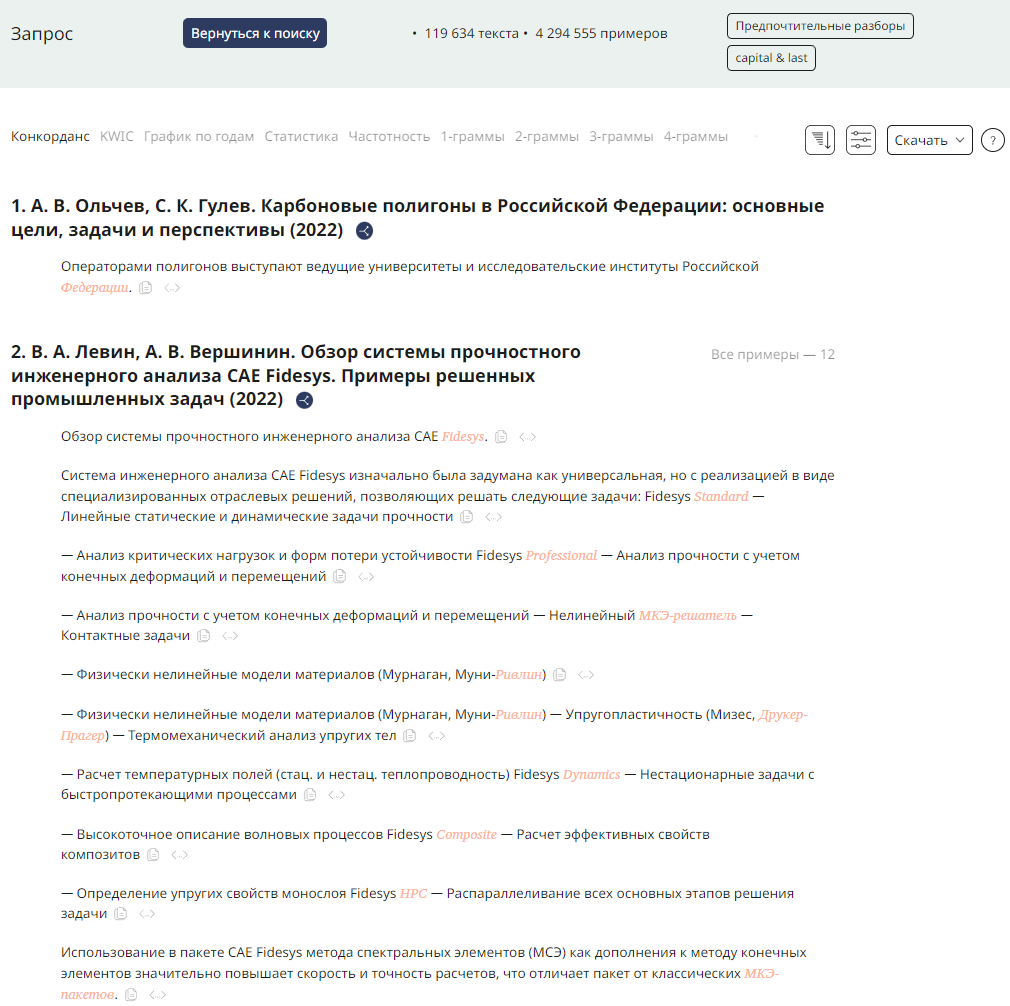
4) Работаем дальше с ФИО. Условие:  


Результат:

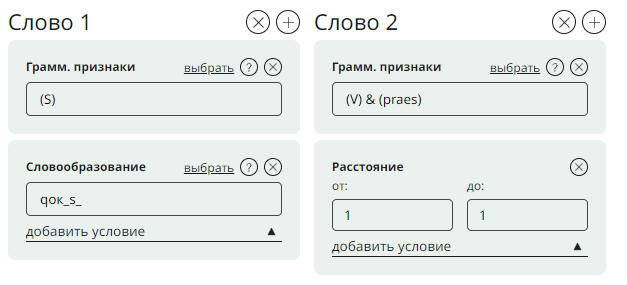


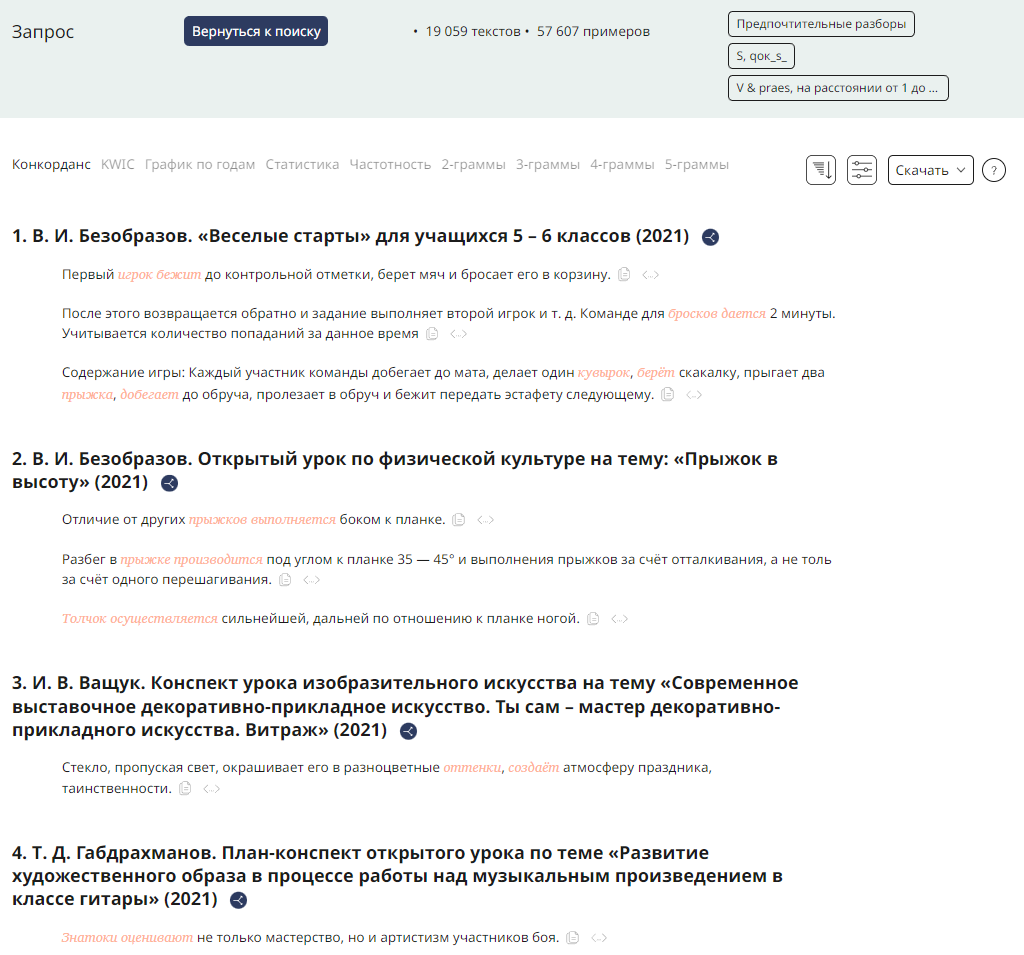
5) В скольких предложениях слово с заглавной буквы стоит в конце предложения?



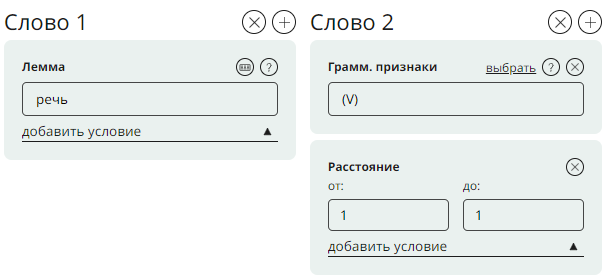


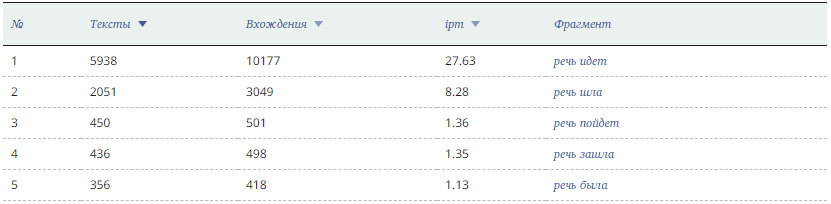
6) Сколько предложений с существительными, оканчивающимися на суффикс «-ок», за которым следует глагол в настоящем времени?



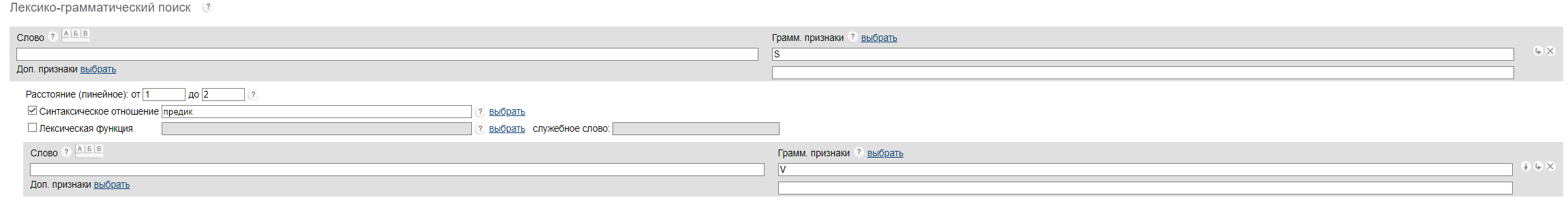


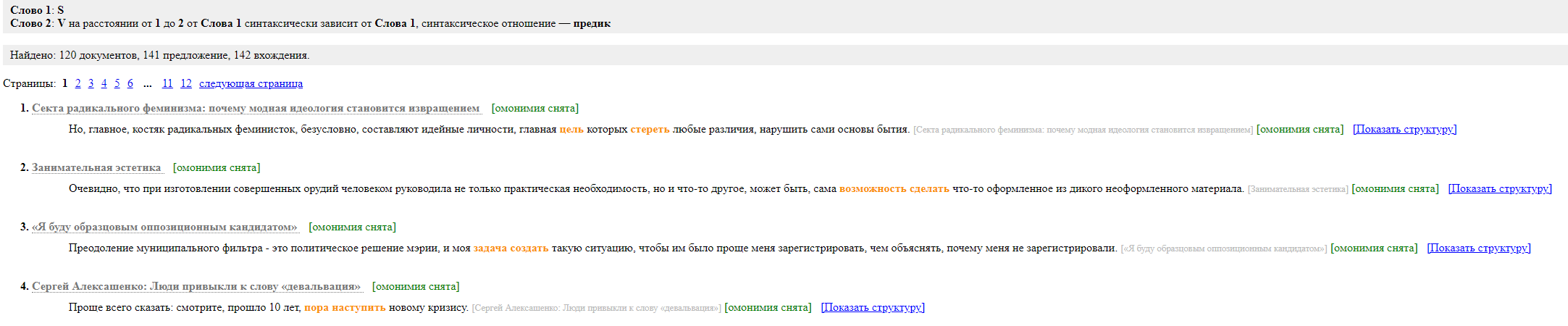
7) Какая самая частая биграмма с существительным "речь", если второе слово - глагол?





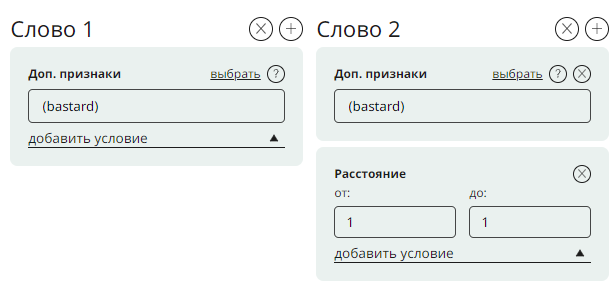
Самая частая – речь идёт.

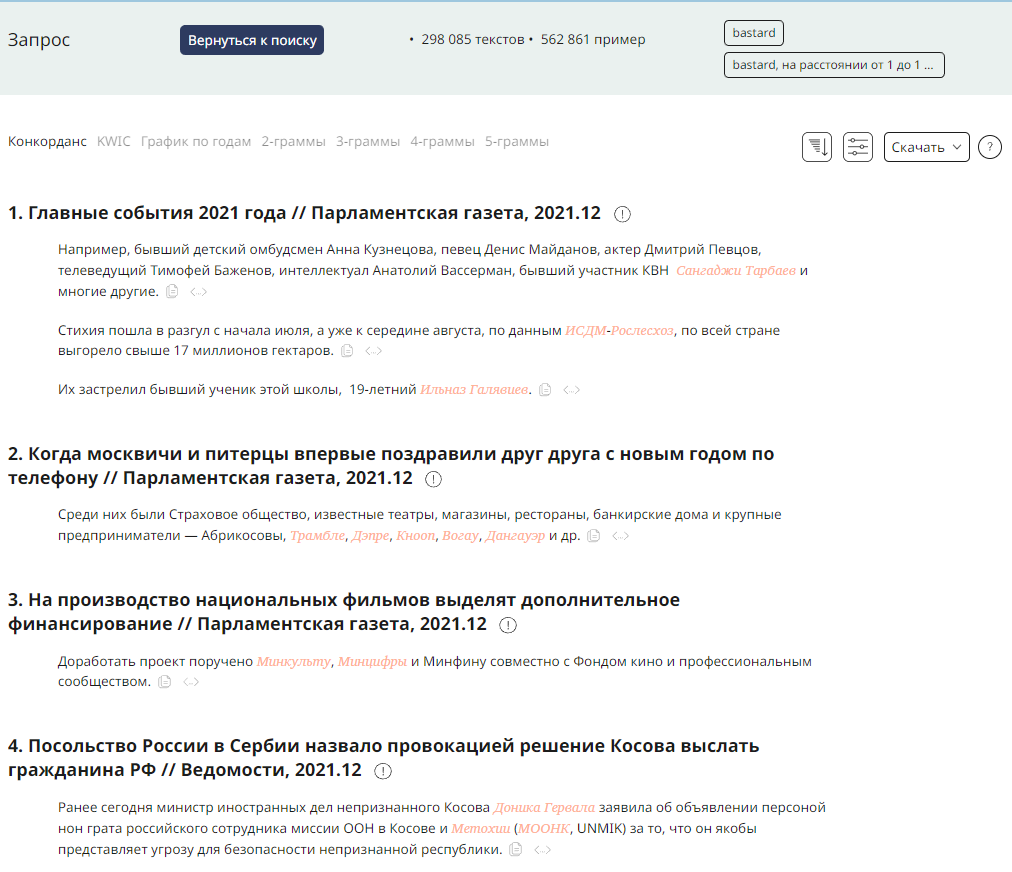
8) Сколько в синтаксическом корпусе предложений, в которых первое слово – существительное, второе – глагол, а связаны они предикативным синтаксическим отношением и «расстояние от родителя» от 1 до 2. Запрос:  


Результат:  


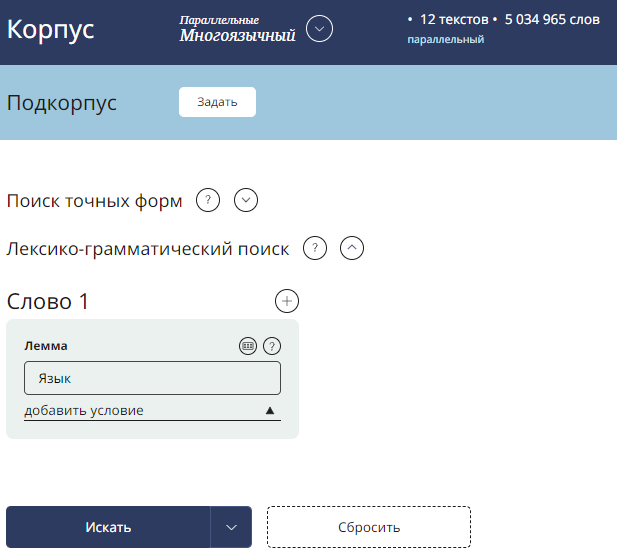
То есть 141 предложение.

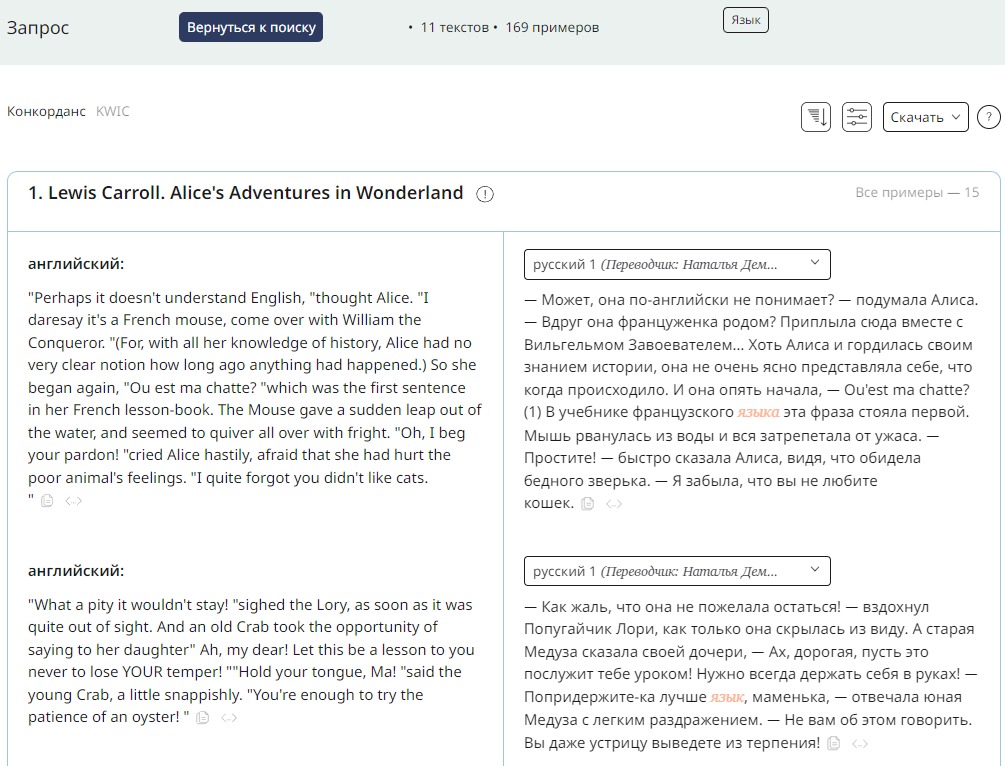
9) Сколько в газетном корпусе предложений, в которых встречаются два идущих подряд несловарных слова?



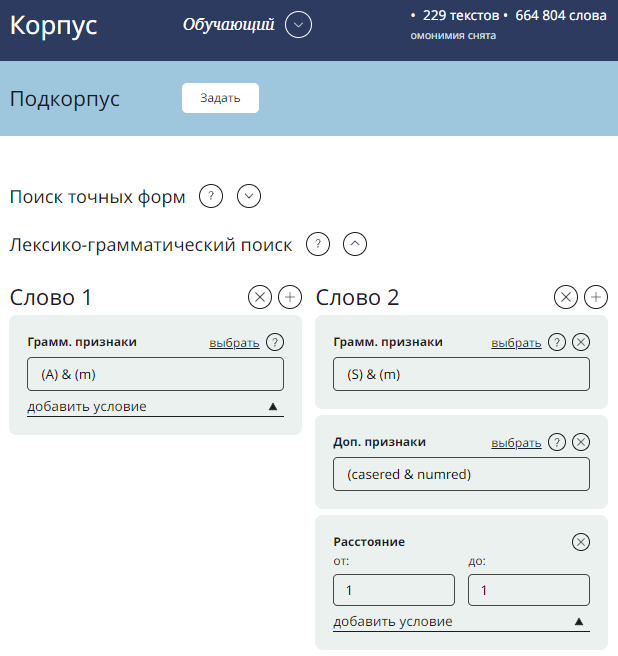


10) В скольких предложениях в многоязычном корпусе встречается слово «язык»?



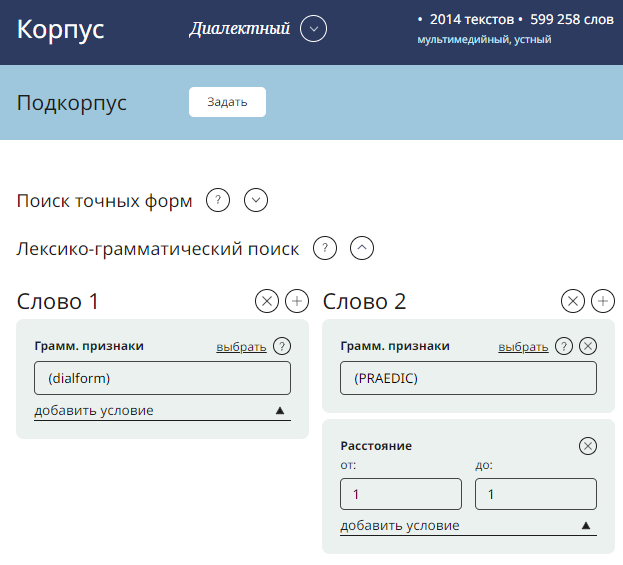


11) В скольких предложениях в обучающем корпусе за прилагательным мужского рода стоит существительное мужского рода, причем они должным быть в одном и том же числе и падеже?



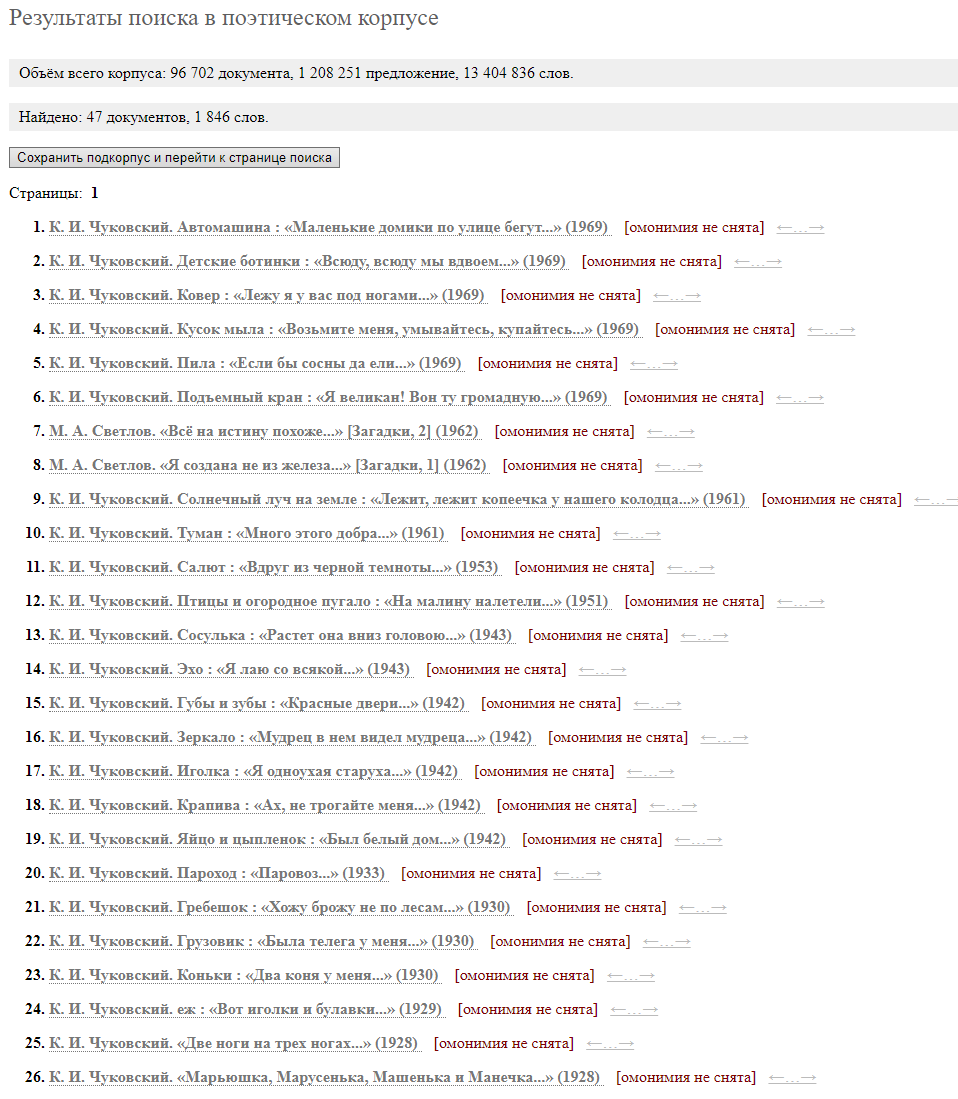


12) Сколько предложений в диалектном корпусе с диалектными лексемами, за которыми следует предикатив?

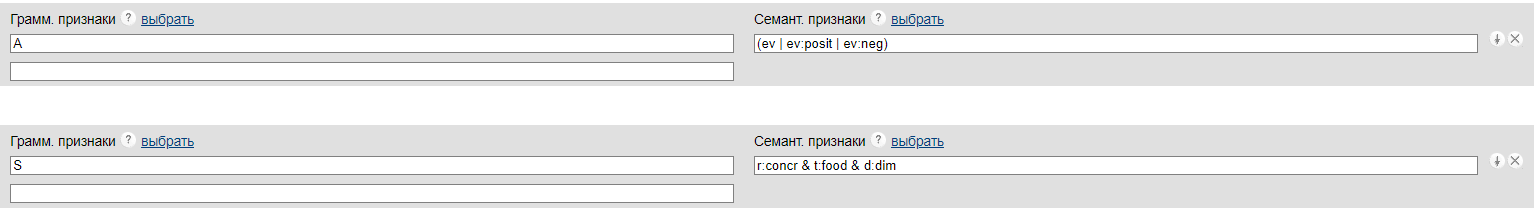


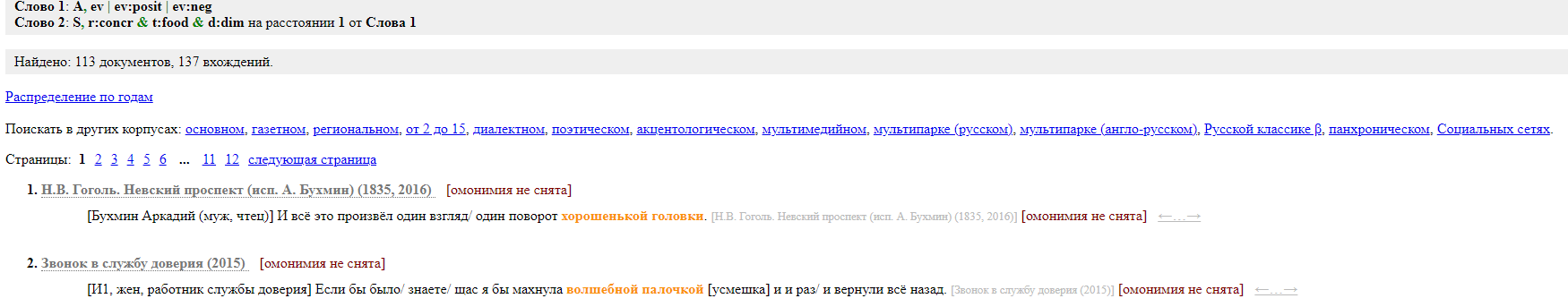


13) Сколько в поэтическом корпусе загадок?

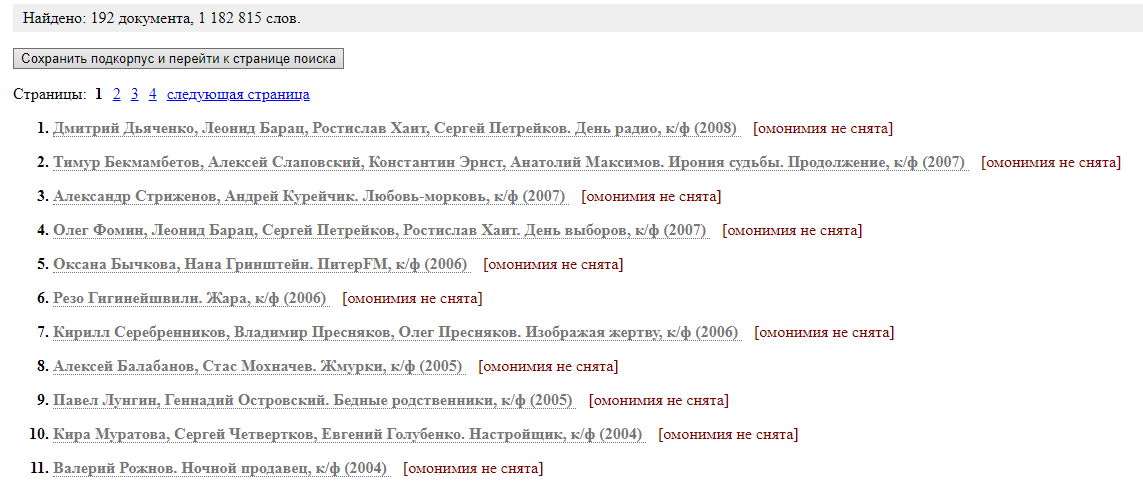


14) В скольких предложениях в «устном корпусе» встречается сочетание прилагательного с семантическим признаком «оценка» и существительного, обозначающего еду или напитки и являющегося диминутивом?

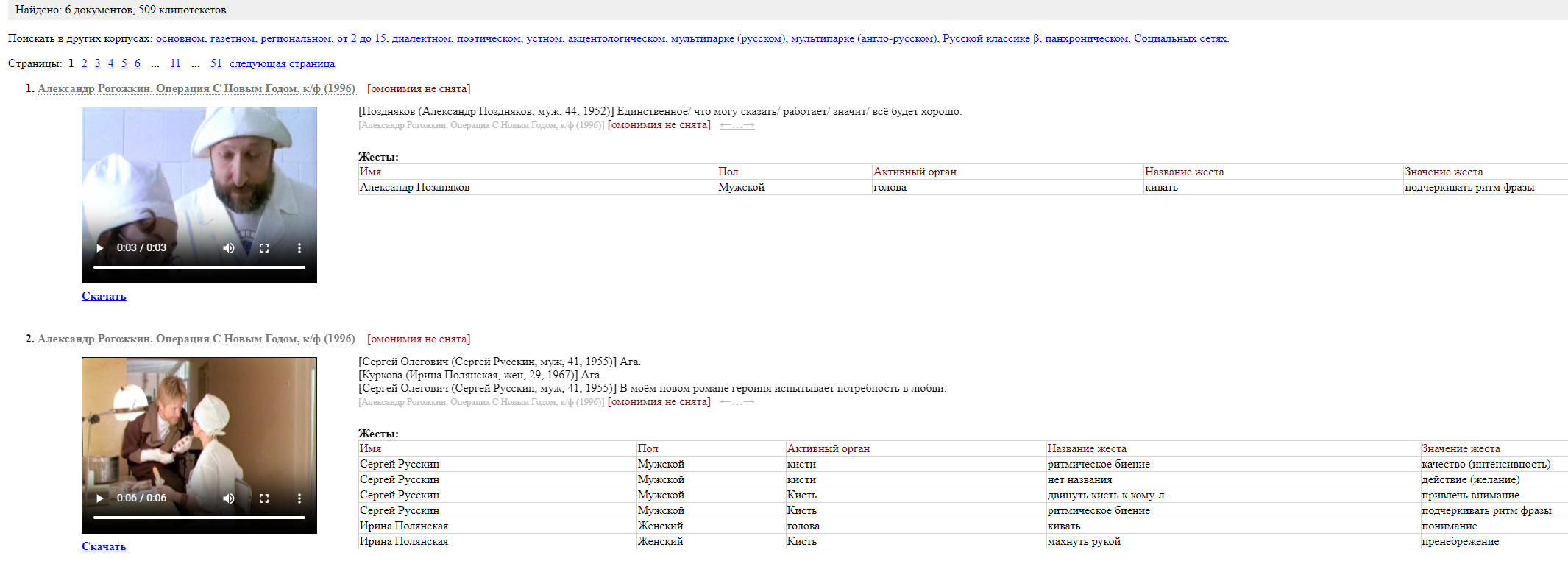




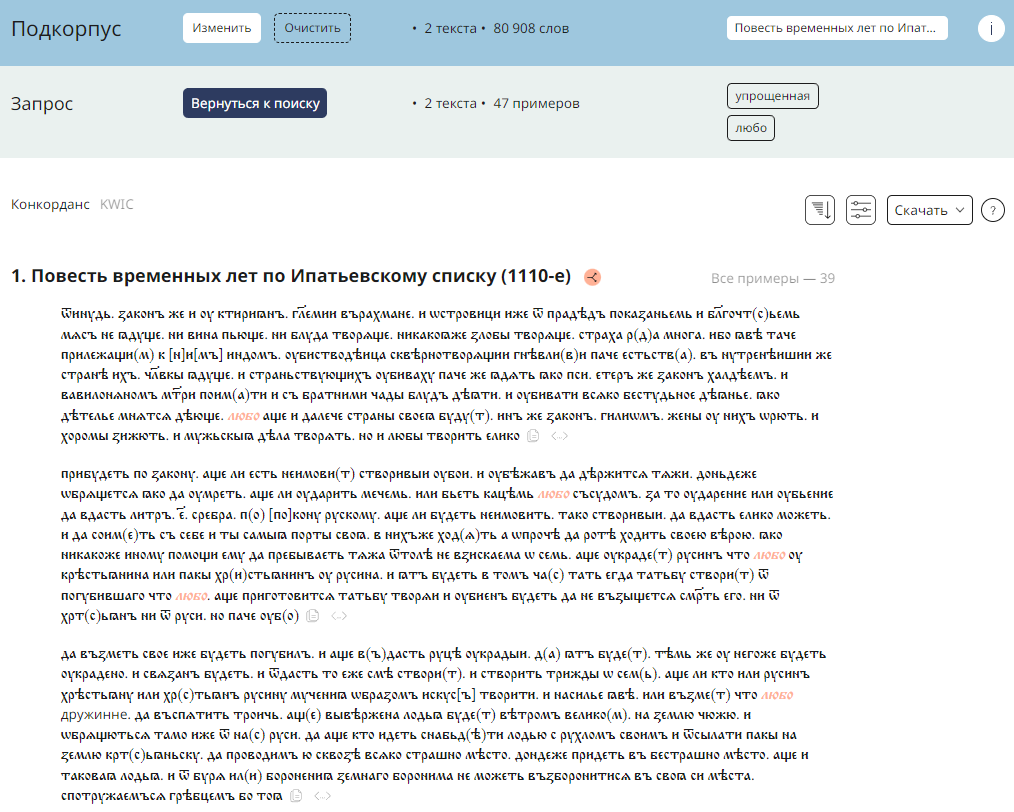
15) Сколько в акцентологическом корпусе кинокомедий?

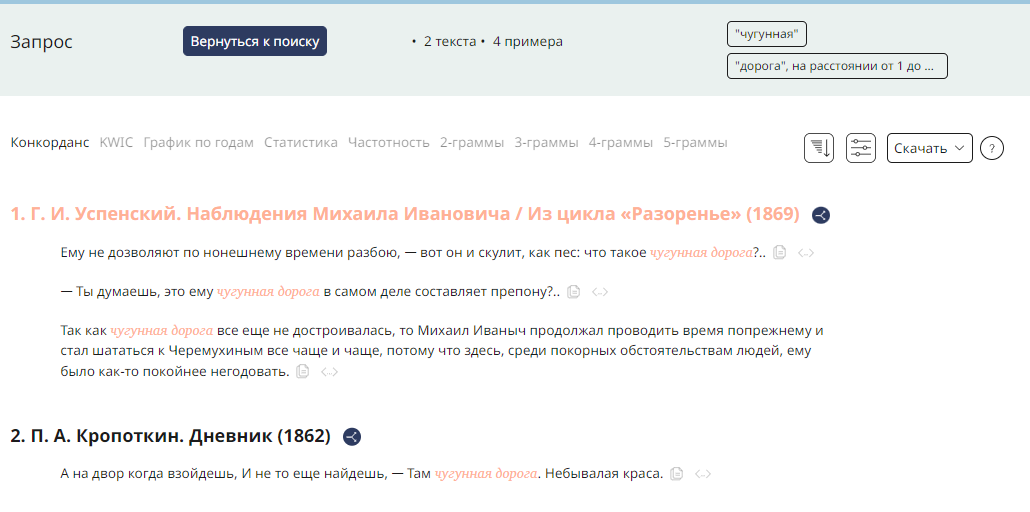


16) В скольких документах мультимедийного корпуса встречаются сцены с аппелятивами, сопровождающиеся риторическими жестами?



17) Сколько раз в историческом корпусе в "Повести временных лет" встречается слово "любо"?



18) Определить значение фразы чугунная дорога и привести ещё пример.

Видим, что чугунной дорогой в то время называли нашу железную дорогу.

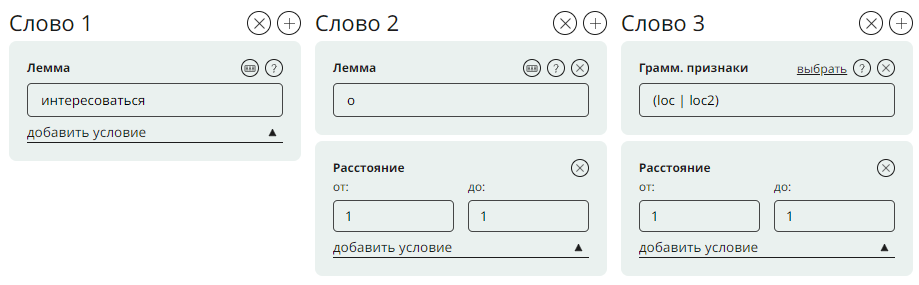
19) Пользуясь поэтическим корпусом, докажите, что в 18 веке слово «приличен» могло иметь ударение на последнем слоге: «приличЁн». Приведите этот пример (эти примеры).

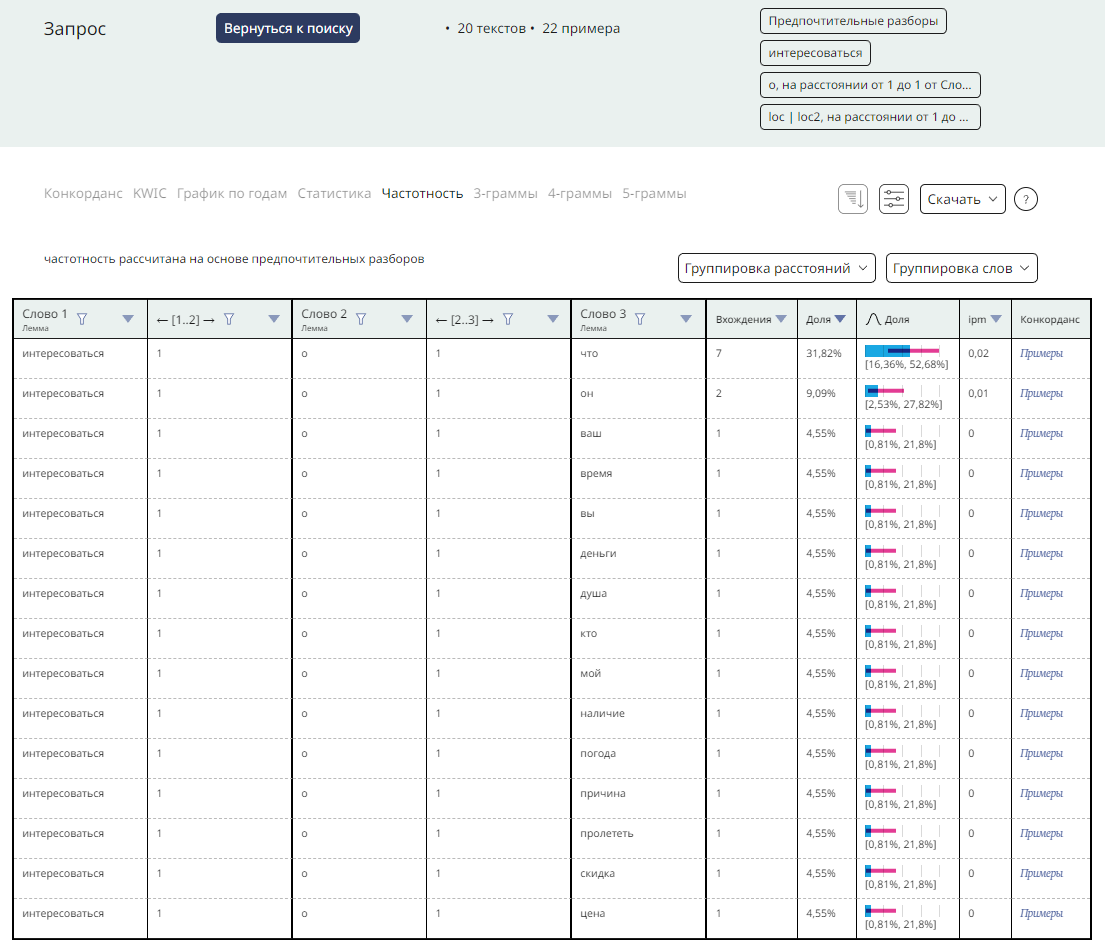
В этом задании как бы я не пытался я не смог найти форму приличён.

20) Из социальной сети Facebook нередко приходит сообщение: *«Интересуюсь о Вас на*

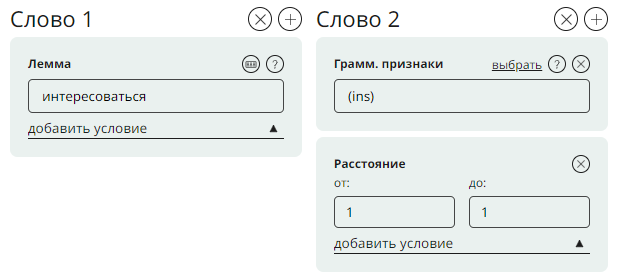
*Facebook. Здравствуйте, Natasha! Я Вас приглашаю посмотреть мою личную страницу на Facebook...»* Воспользовавшись корпусом, оцените, насколько конструкция интересоваться о ком-либо» распространена, в каких текстах она встречается. Сравните ее частотность с частотностью конструкции «интересоваться кем-либо».\_\_

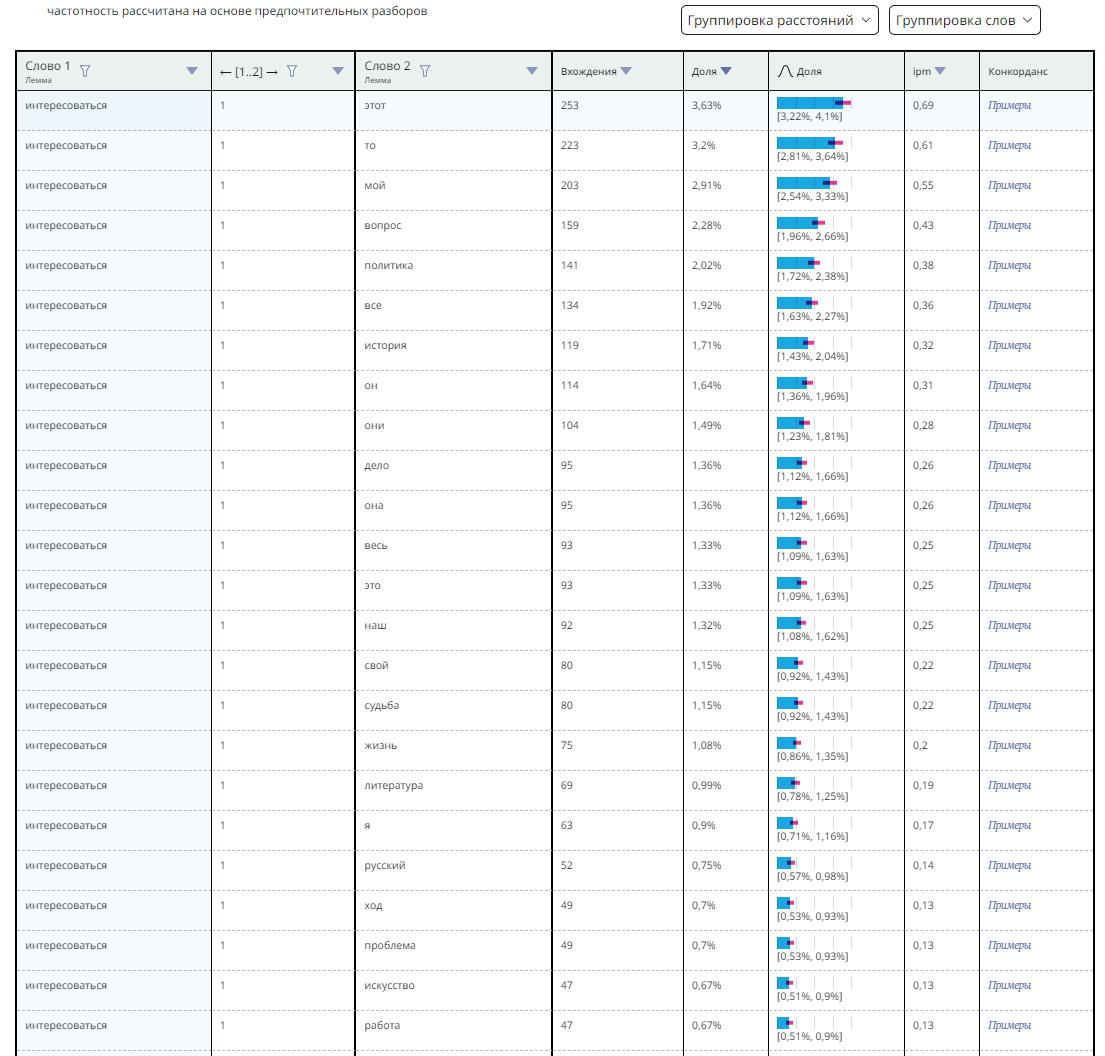
Конструкция “интересоваться о ком-либо”:





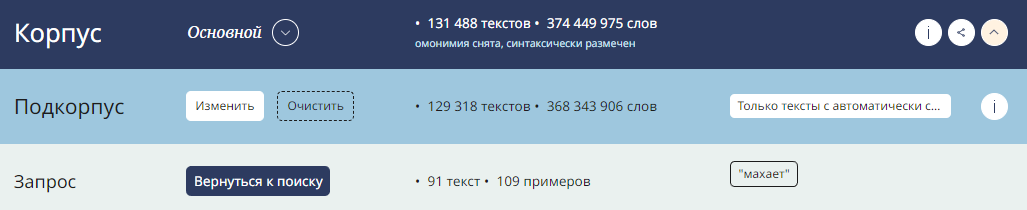
Конструкция “интересоваться кем-либо”:

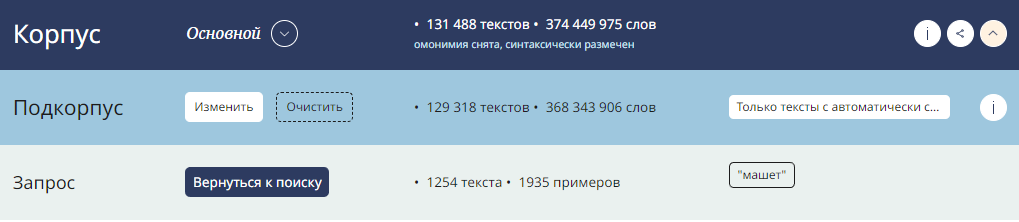


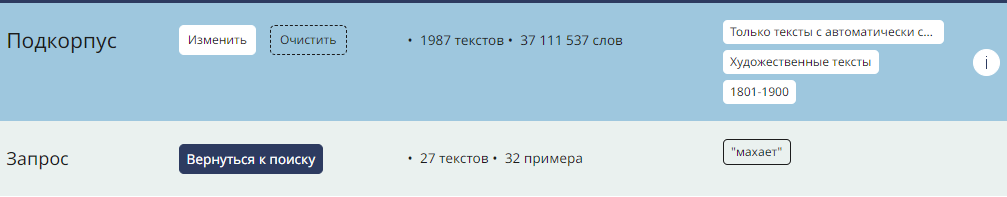


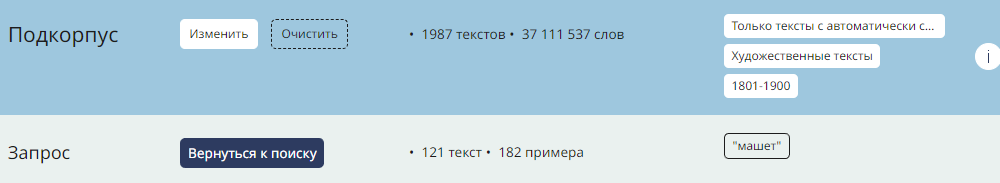
21)

| Результаты | НКРЯ, основной корпус | | | Подкорпус худ. литературы | | |
| --- | --- | --- | --- | --- | --- | --- |
|  | Размер: 283 431 966 | | | Размер: 150 820 057 | | |
|  | Количество текстов | Абсолютная частота | Ipm | Количество текстов | Абсолютная частота | Ipm |
| махает | 91 | 109 | 0,39 | 27 | 32 | 0,21 |
| машет | 1254 | 1935 | 6,83 | 121 | 182 | 1,21 |

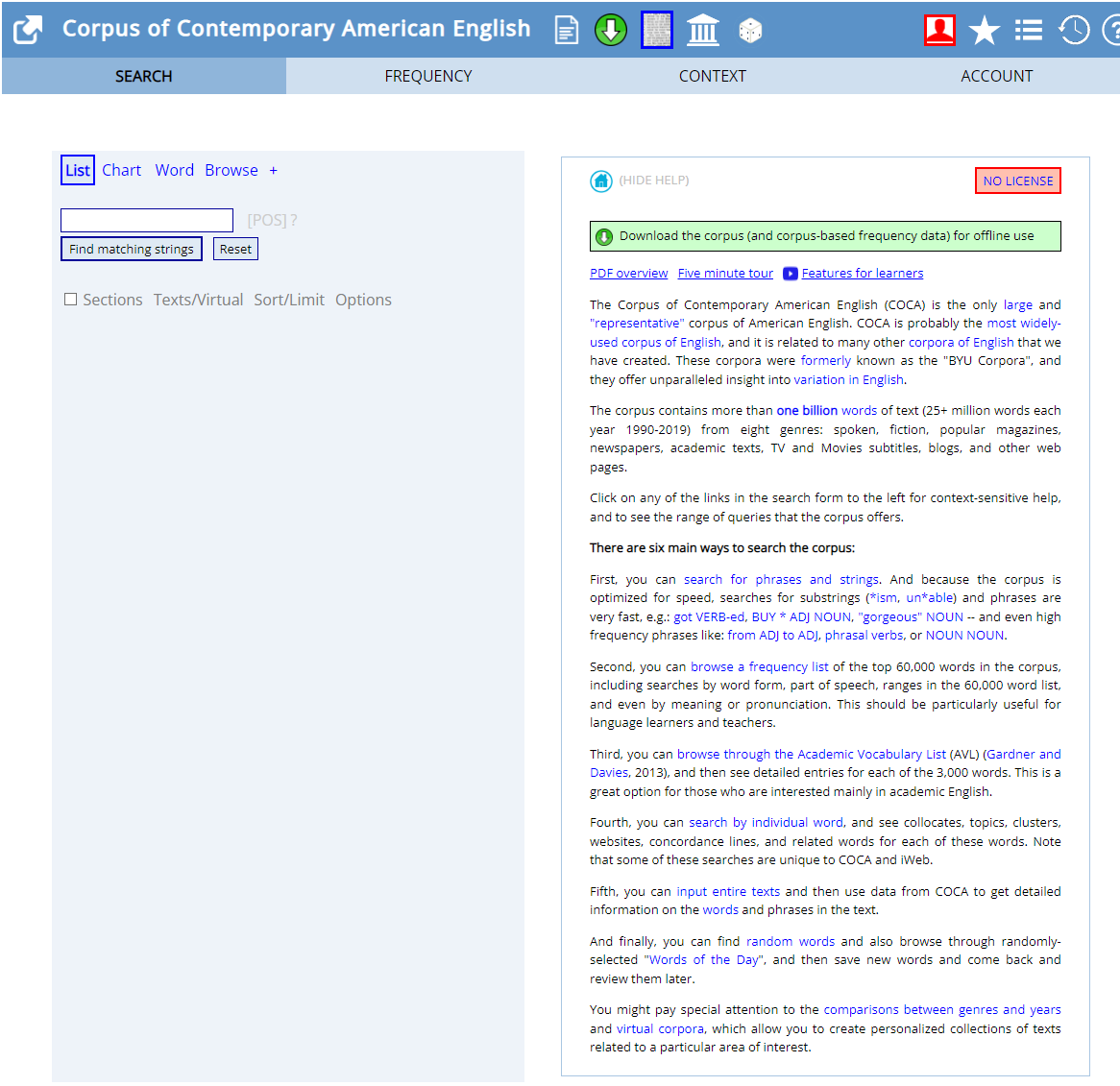




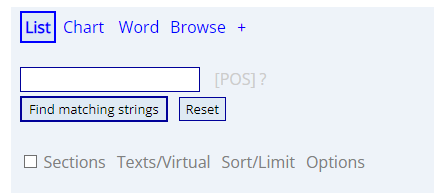


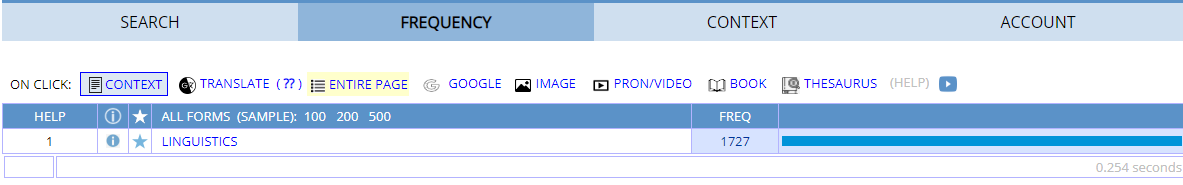


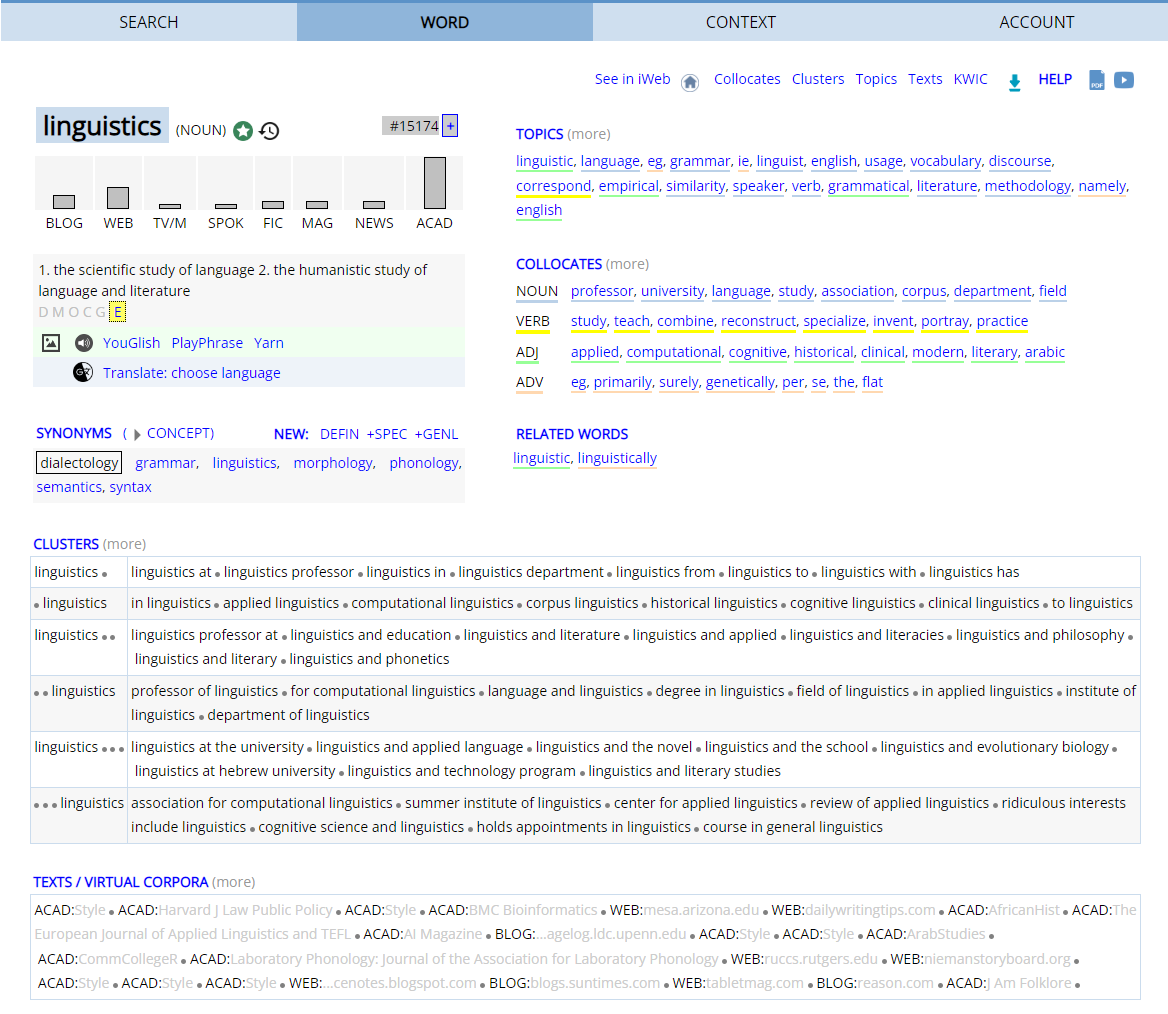
22) Возьмём самый известный, пожалуй, корпус Марка Дэвиса – COCA. Главная страница сайта:

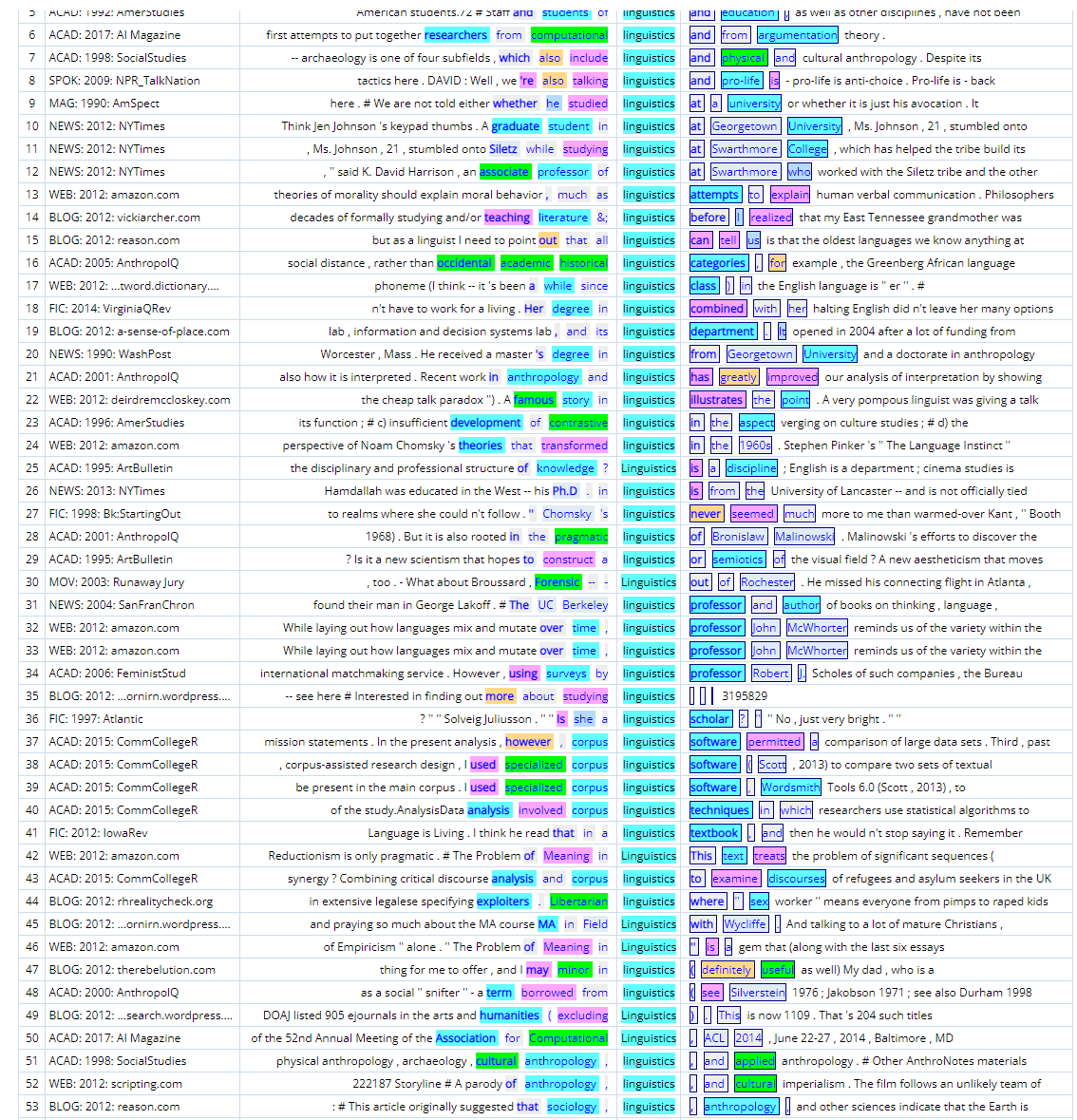


В этом корпусе есть много возможностей, таких как поиск слов, анализ контекстов, определение частотности и сбор статистики, а также создание собственных подкорпусов.







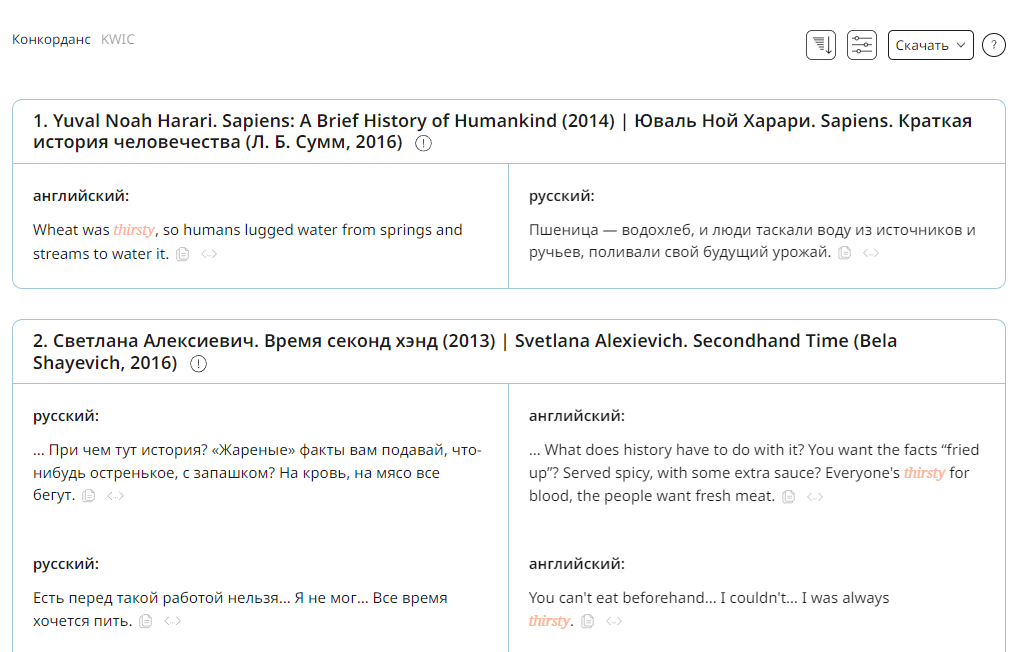


В COCA используется следующий синтаксис (примеры): Вопросительный знак (?) используется для обозначения одного произвольного символа. Кавычки (") используются для поиска точных фраз. Диапазон ([]) используется для указания диапазона символов. Ударение (/) используется для указания ударения. Логические операторы AND, OR, NOT используются для комбинирования запросов.

Работа с параллельными корпусами.

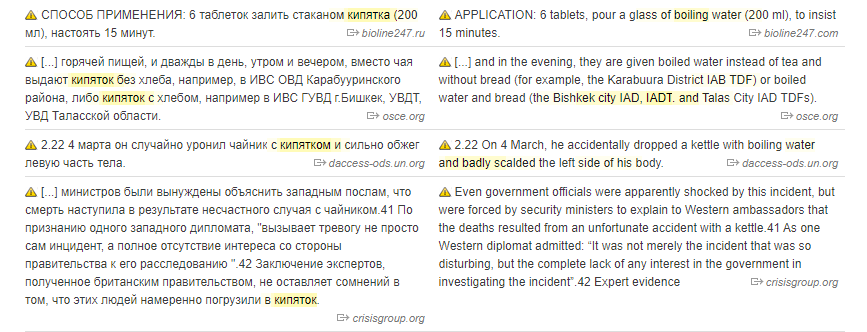
* НКРЯ:

Если возьмём слово “thirsty”, которому нет аналогов в русском языке, то получим следующее:



* Linguee:

Пример с кипятком:



Мои примеры аналогичных ресурсов с параллельными корпусами:

1. StudyEnglishWords
2. Parallango
3. COCA

Далее сравним программы по выравниванию текстов:

| Название | Характеристики и особенности | Простота в использовании | Доступность | Внешний вид |
| --- | --- | --- | --- | --- |
| LEOBILINGUA 1.0 | Автоматический, поддержка файлов TXT, XLIFF | Достаточно понятно | Доступен онлайн без необходимости установки, бесплатно | Не красива, но удобна |
| Hunalign LF Aligner | Поддержка текстовых файлов | Есть трудности, но в целом понятно | Доступен для загрузки и использования с командной строки, бесплатно | Имеет простой текстовый интерфейс командной строки |
| Champollion | Поддержка текстовых файлов | Достаточно понятно | Доступен для загрузки и установки на локальную систему, бесплатно | Имеет пользовательский графический интерфейс с удобной навигацией |
| Skuuper Cleaner | Автоматически, может обрабатывать большие объёмы текстов | Достаточно просто | Доступен онлайн без необходимости установки, бесплатно | Нет лишней информации, всё легко самому понять |
| YouAlign | Автоматический, бесплатная версия позволяет выравнивать тексты до 1Мб и скачивать их в специальном xml формате | Также несложно | Доступен онлайн без необходимости установки, есть платная версия | Дизайн чуть более приятный, а главное понятный |