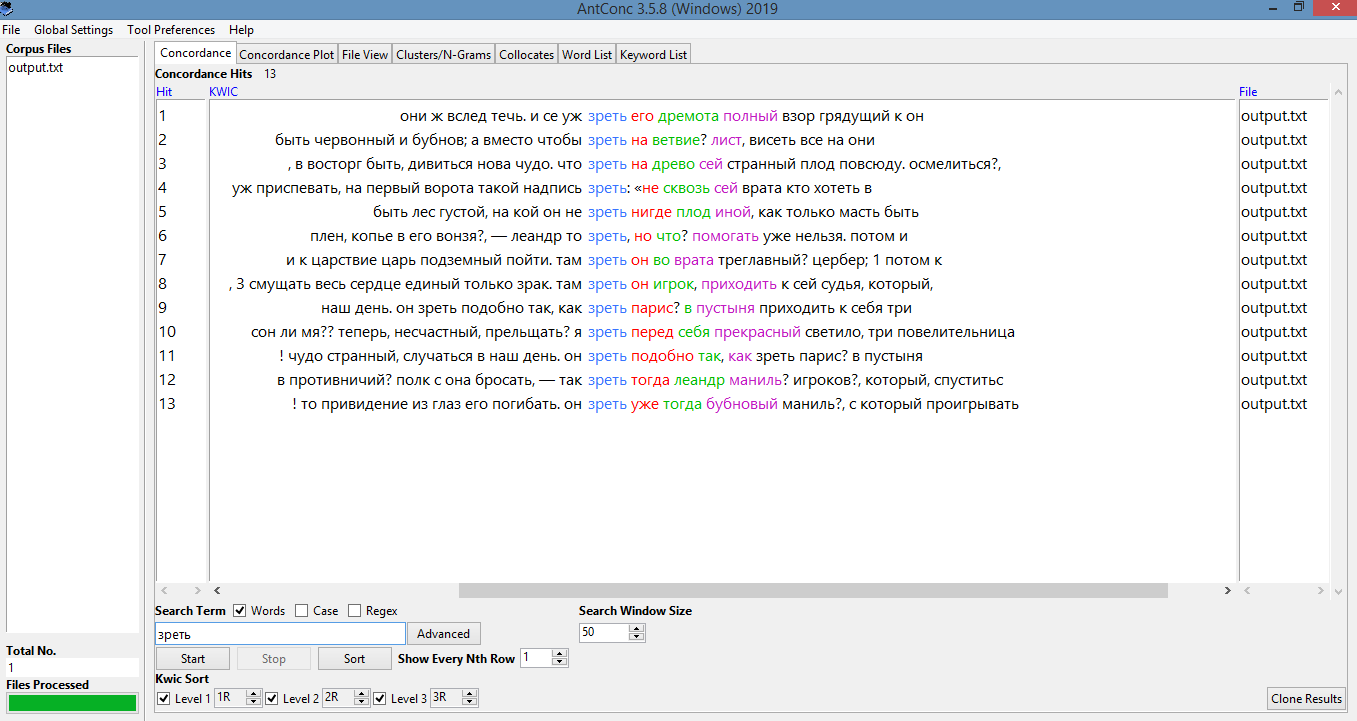
*Айгельди Эминова*

**Домашнее задание 3**

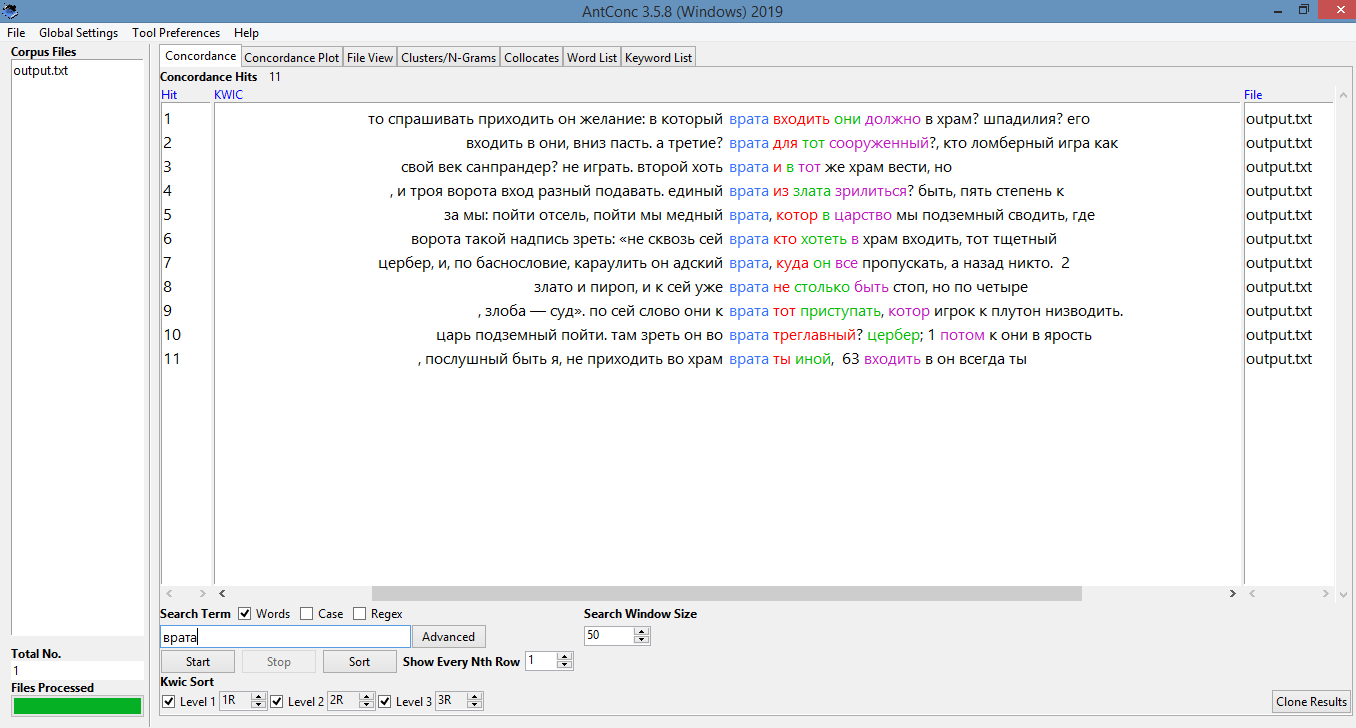
1. **AntConc**

*4. Кликнув на 3 устаревших слова, посмотрите их конкордансы.*

Устаревшее слово 1:

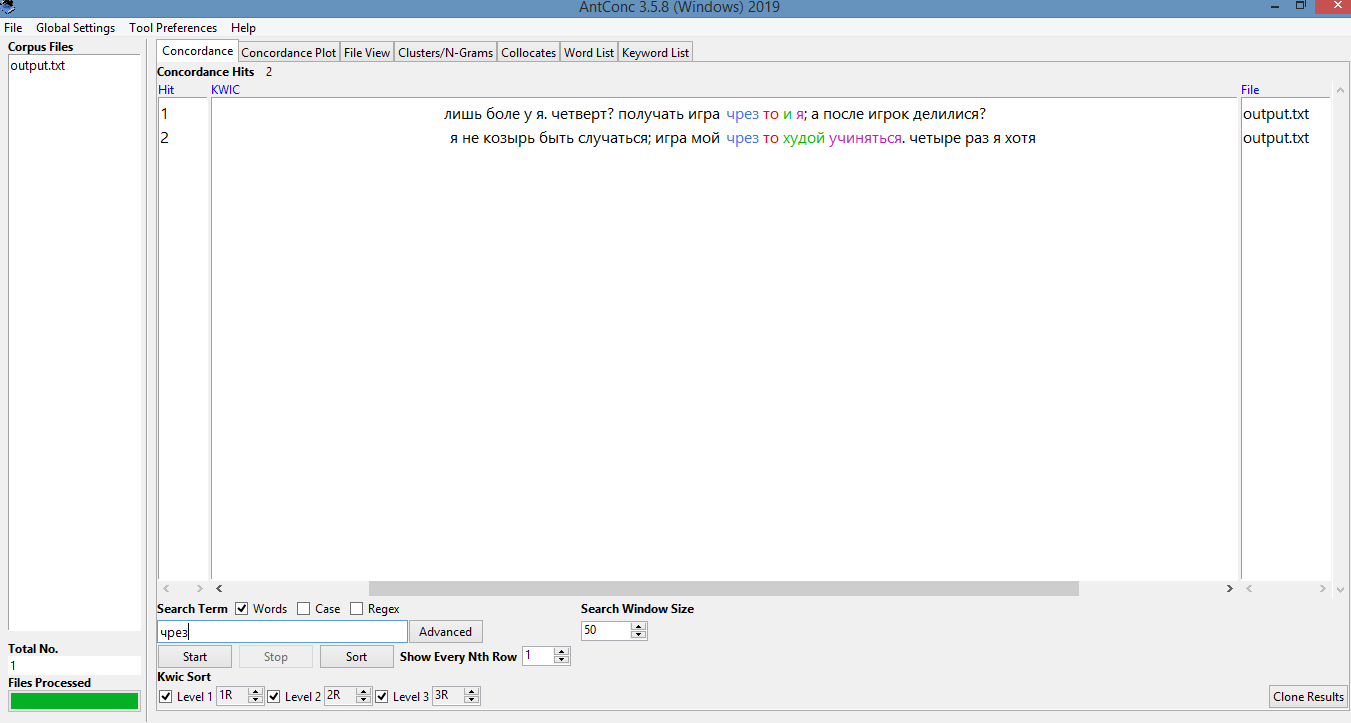


Устаревшее слово 2:



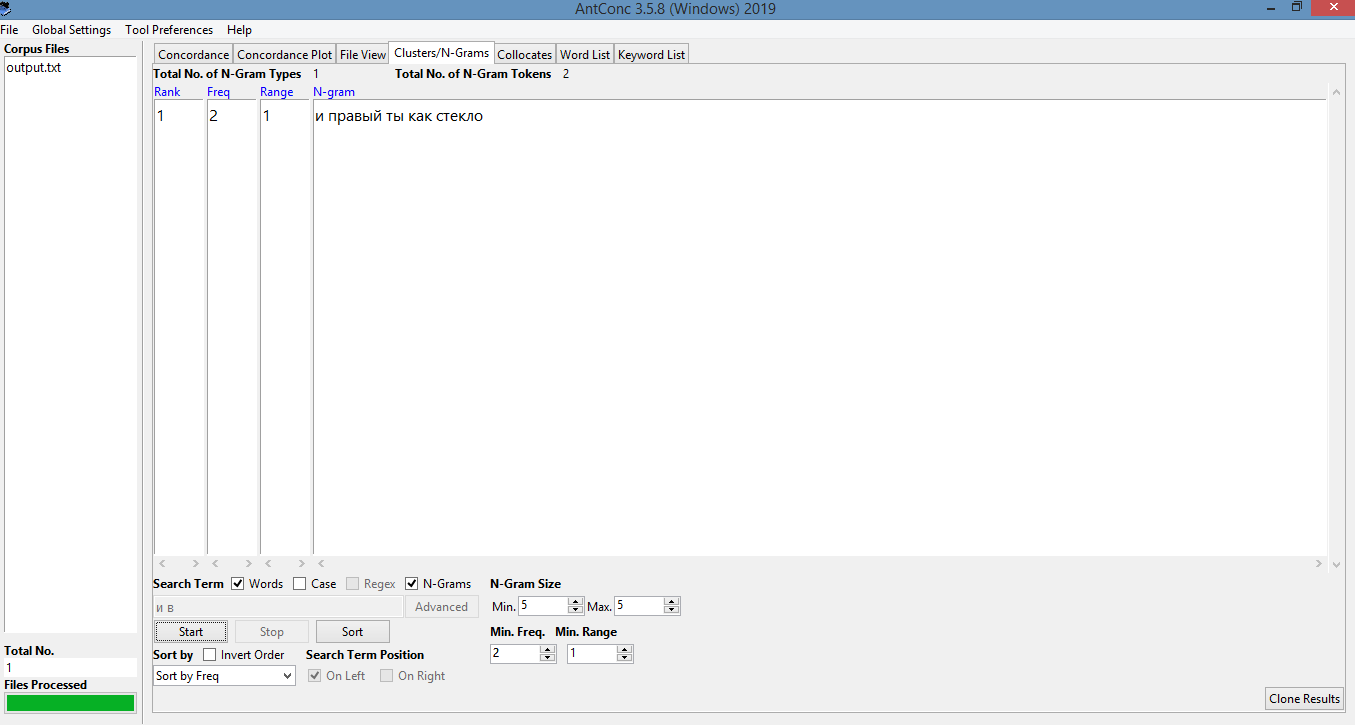
Устаревшее слово 3:

!*Через* входит в список стоп-слов, это не ошибка. Однако в нашем тексте употребляется его устаревший вариант *чрез*, который сохранился после применения стоп-слов.

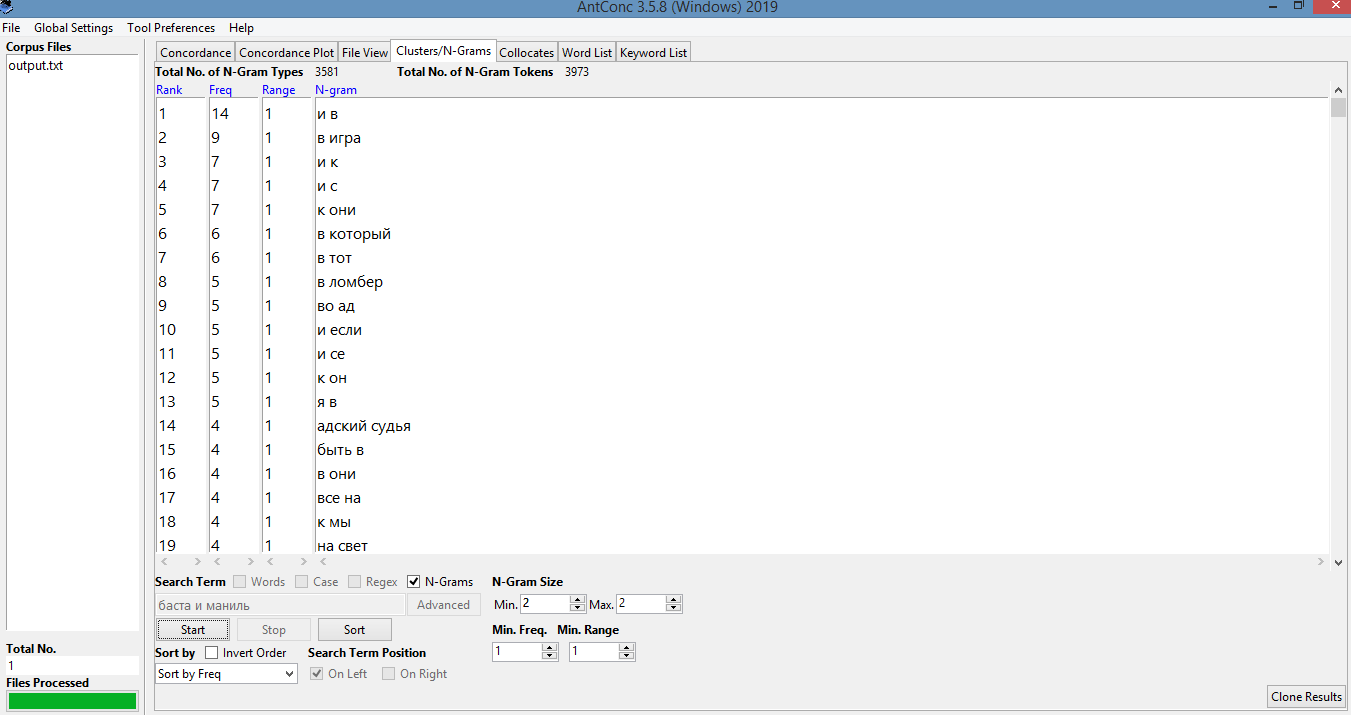


*5. Постройте частотный список двух-, трех- и т.д. -словных словосочетаний (вкладка Cluster/N-Grams, поставьте галочку на N-Grams, укажите, сколько слов в ngram-е вы хотите видеть, например, Min:3, Max:3, установите порог вхождений в корпусе, например, 10). Кликнув на n-грам, вы также можете попасть в его конкорданс.*

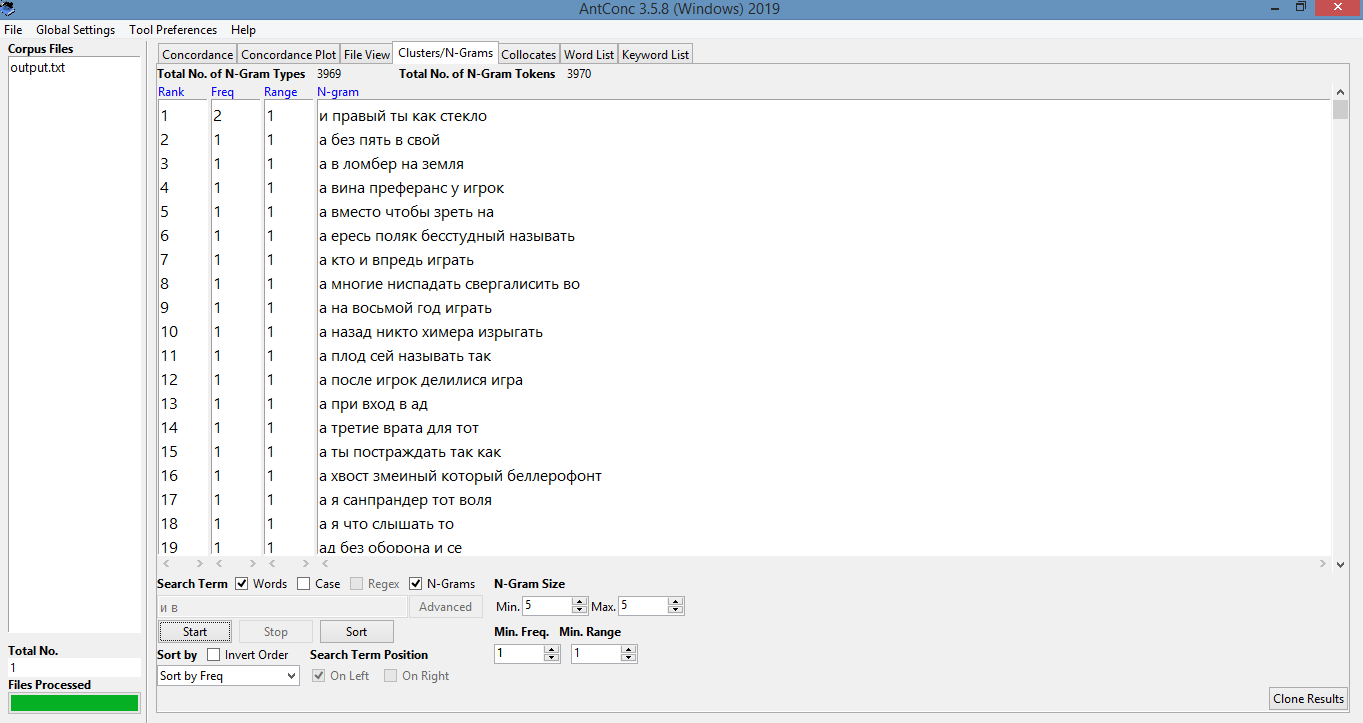
1. min, max-5, freq-2



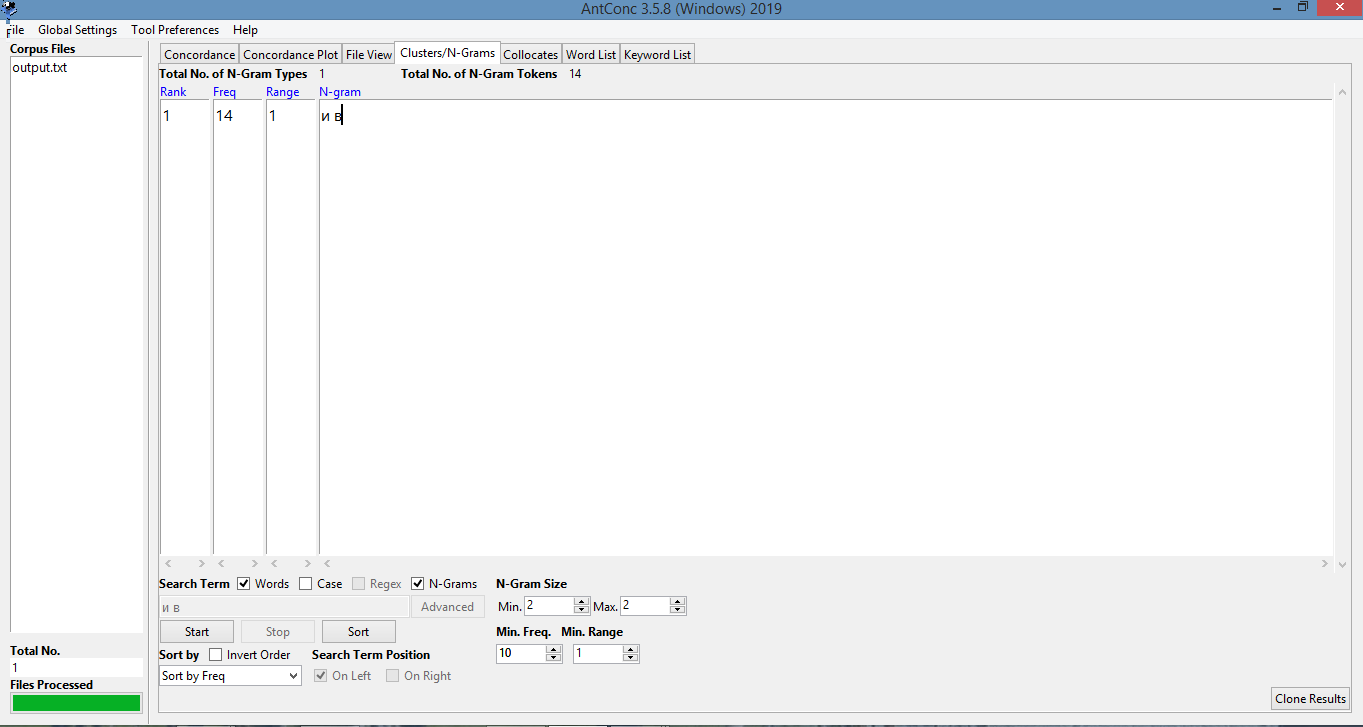
1. min,max 2



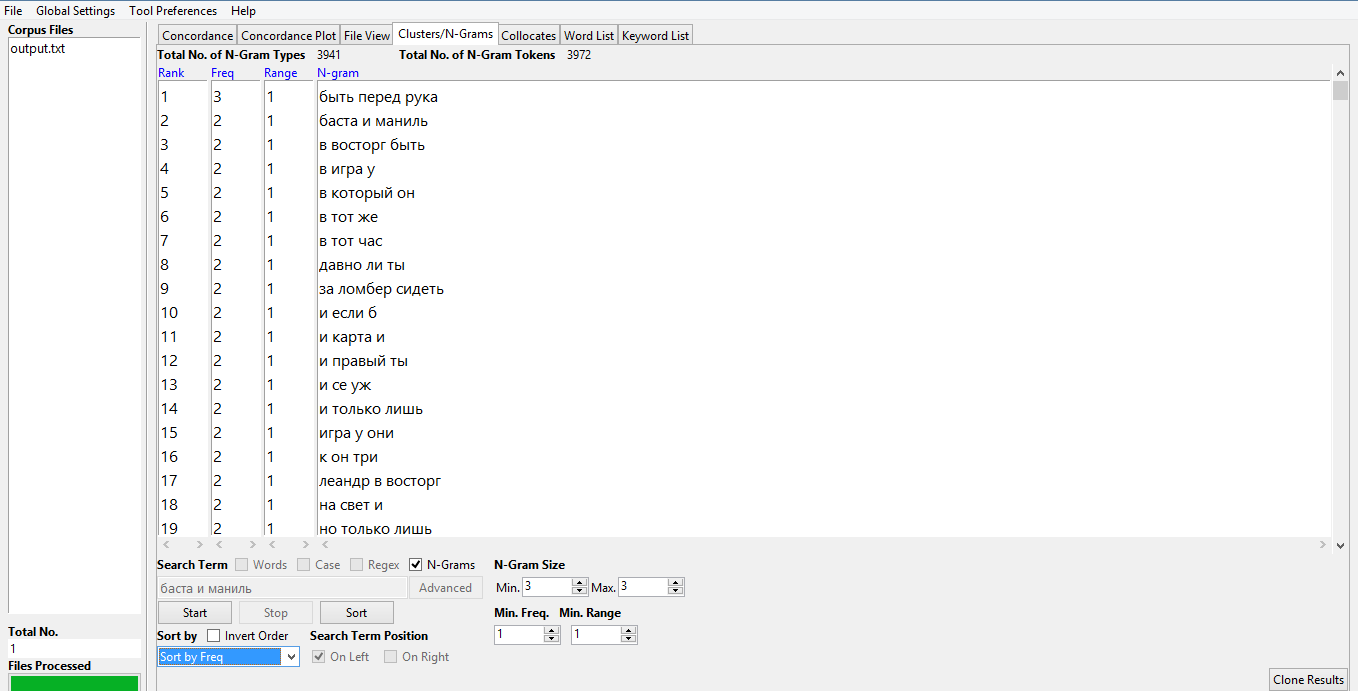
1. min,max – 5, freq – 1



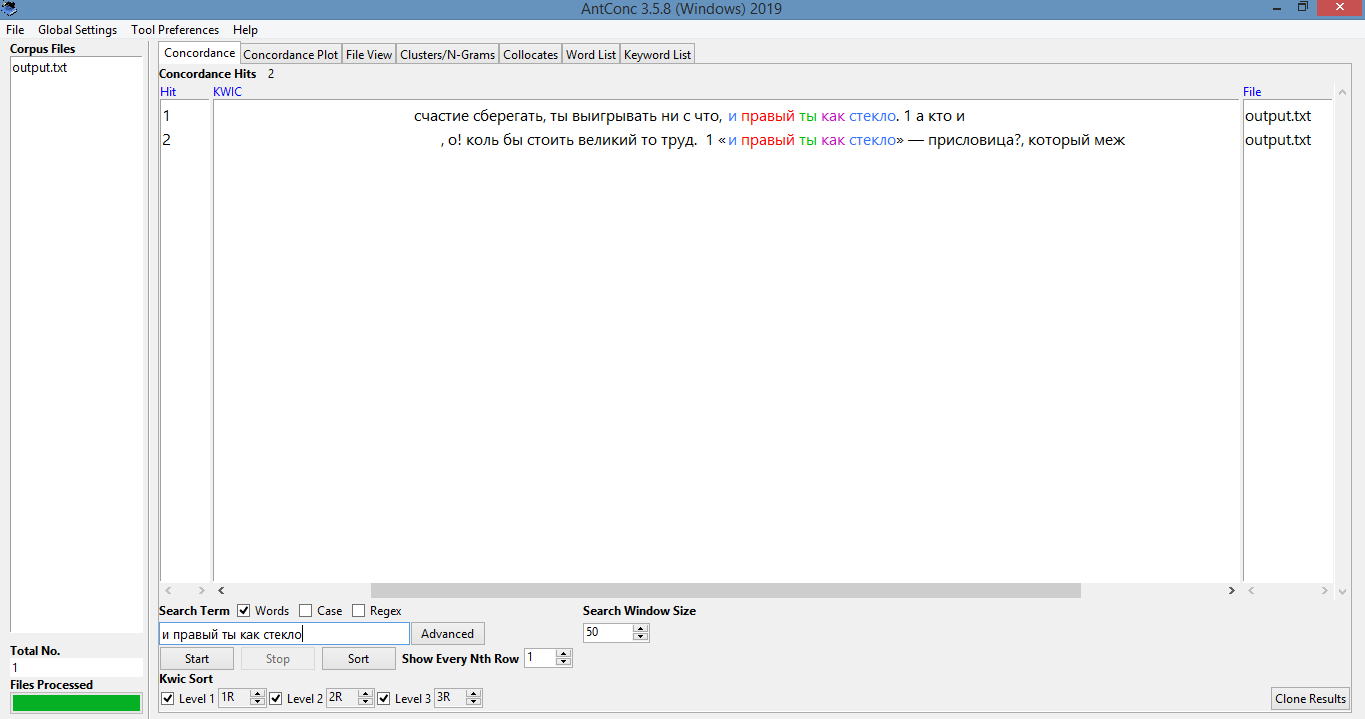
4.min,max – 2 freq – 10



5.min,max – 3,freq – 1

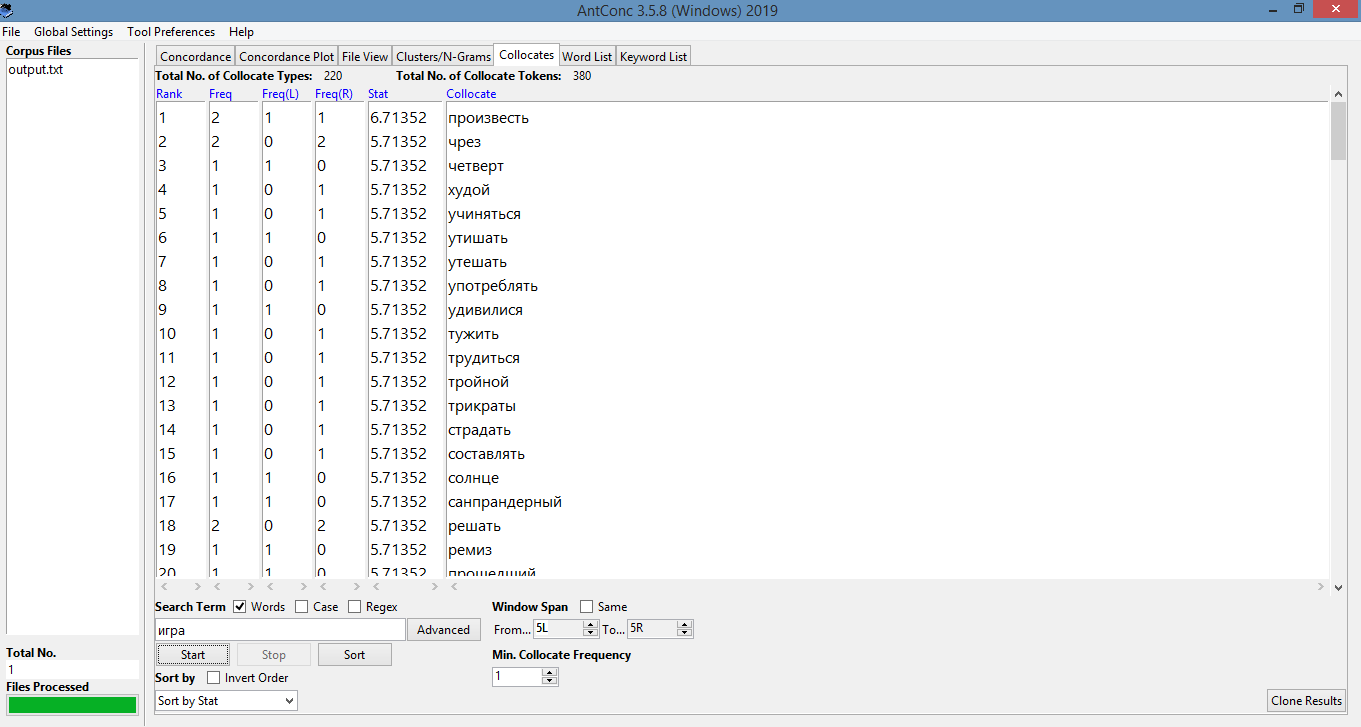


6. Конкорданс через ngram – пример



*6. Постройте списки коллокатов выбранного вами слова (вкладка Collocates), указав границы окна справа / слева*

Выбранное слово – **игра**

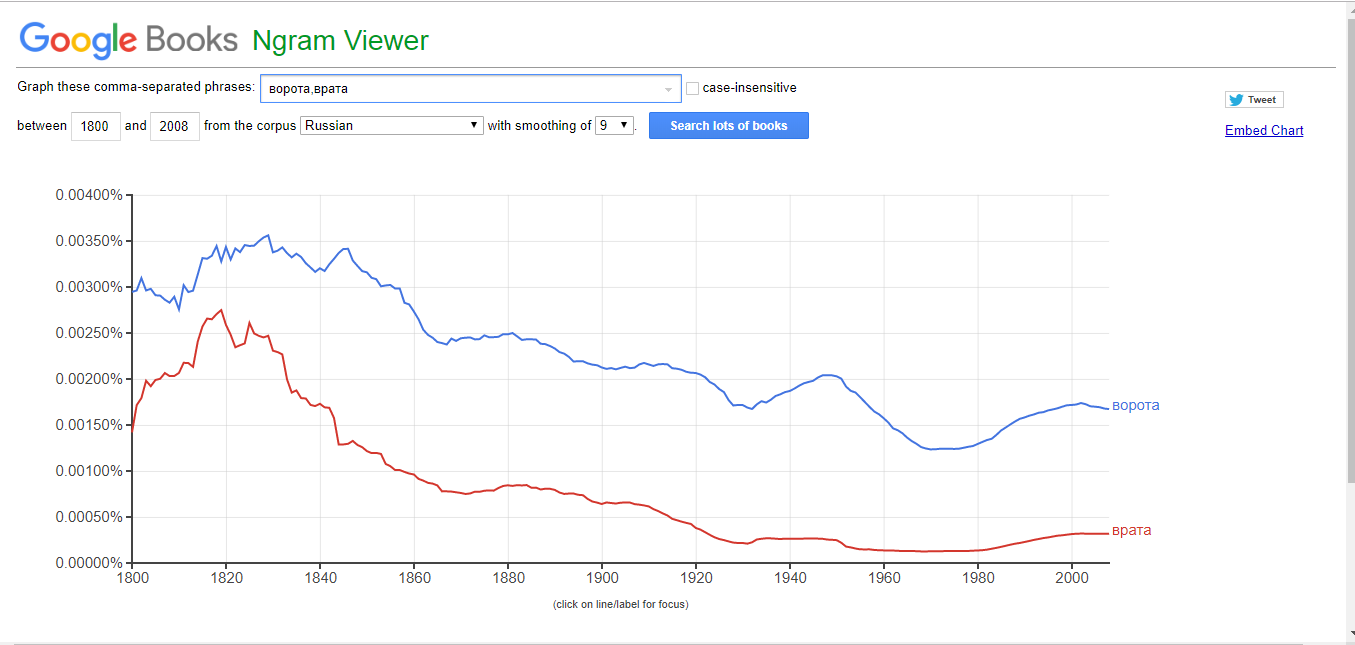
****

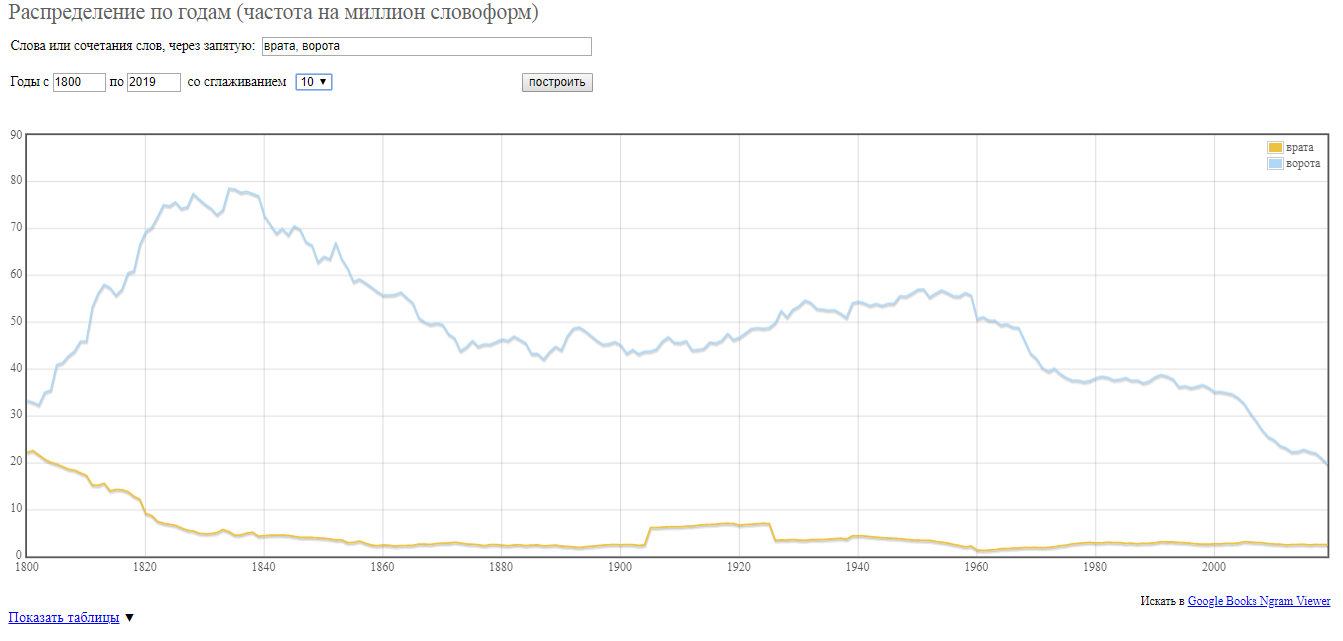
**2.  Google Ngrams и НКРЯ**

1*.Изучите фрагмент вашего текста, выделите устаревшие слова.*

*2. Сравните их употребление с современными аналогами в Google Ngrams и НКРЯ. Сделайте выводы.*

1)Врата vs ворота (1 скрин – Google Ngrams, 2 – НКРЯ)





Врата-ворота

Google Ngrams. График показывает, как два слова практически одинаково употреблялись с 19 века по настоящее время. Пик этих слов был в первую половину 19 века,

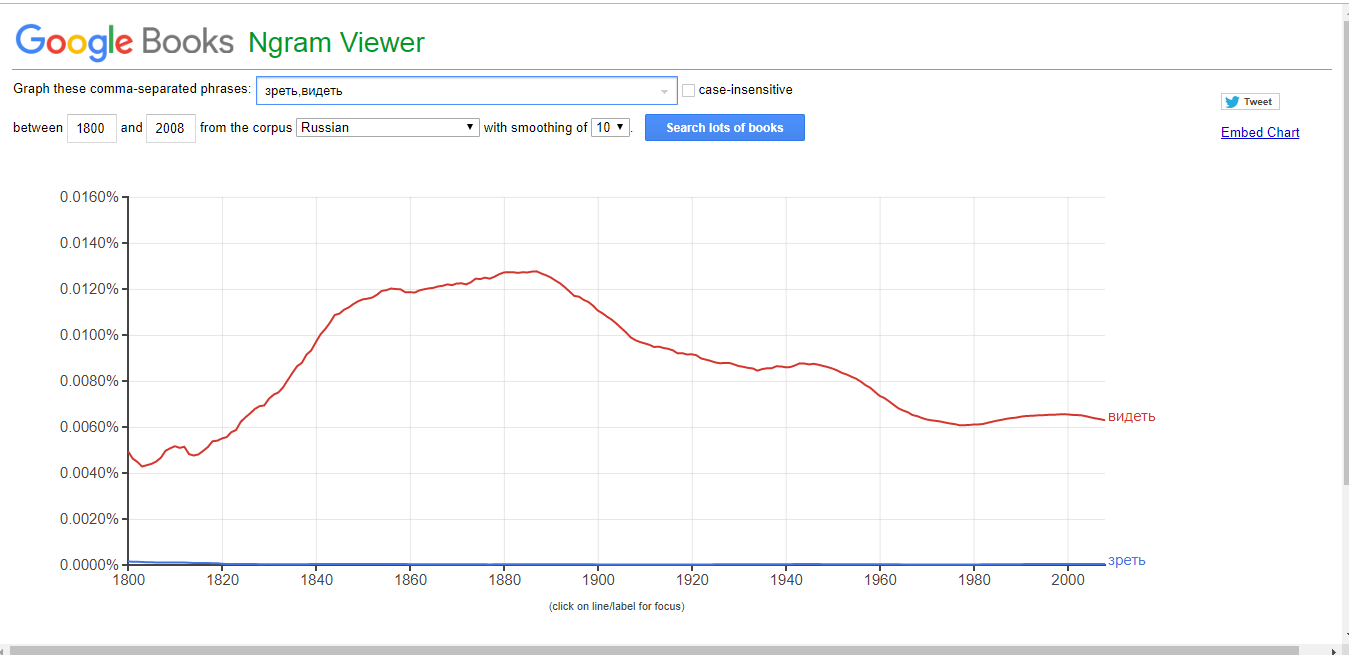
далее они стали употребляться чуть реже. Существенное отличие в употреблении этих слов: врата менее употребительны чем ворота.

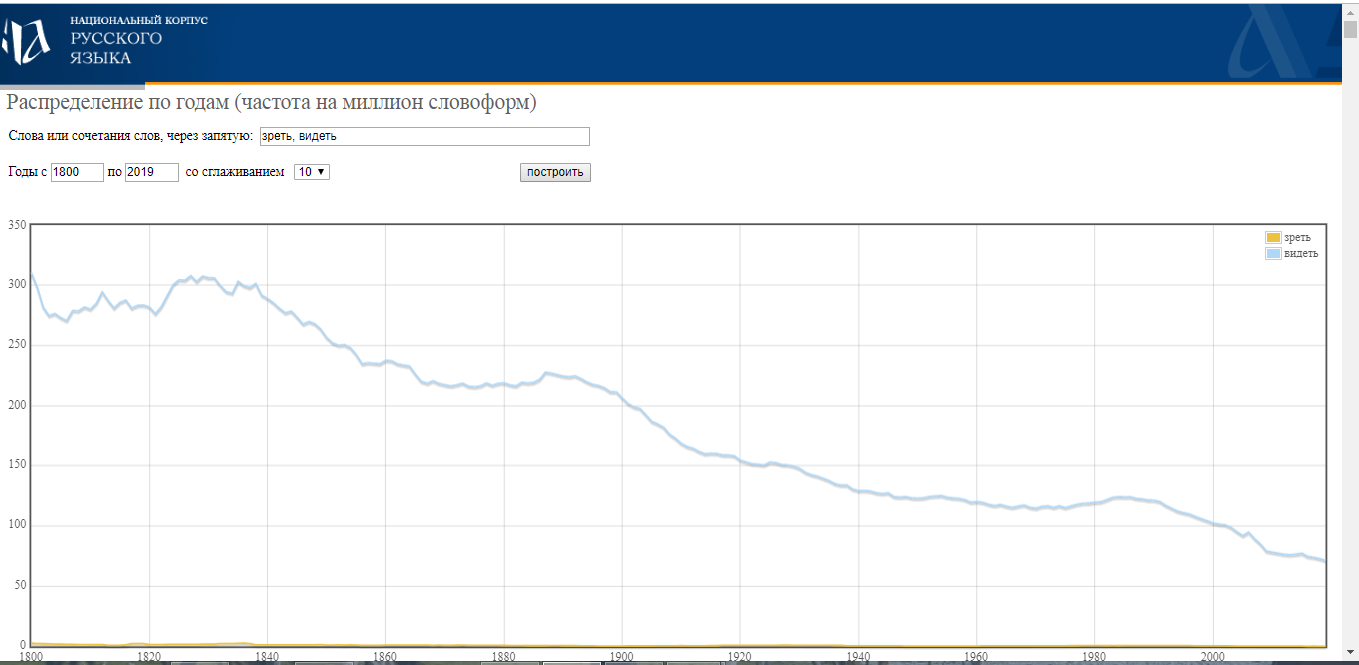
НКРЯ: График в НКРЯ представляет немного иную картину: пик слова врата был в начале 19 века, далее его употребление пошло на спад.Слово ворота было намного

употребительным.

Вывод: Несмотря на некоторую разницу в двух графиках, можно сделать очевидный вывод, что слово врата выходило из употребления уже в 19 веке,что может свидетельствовать о сокращении употребления церковно-книжных слов

2) Зреть vs видеть





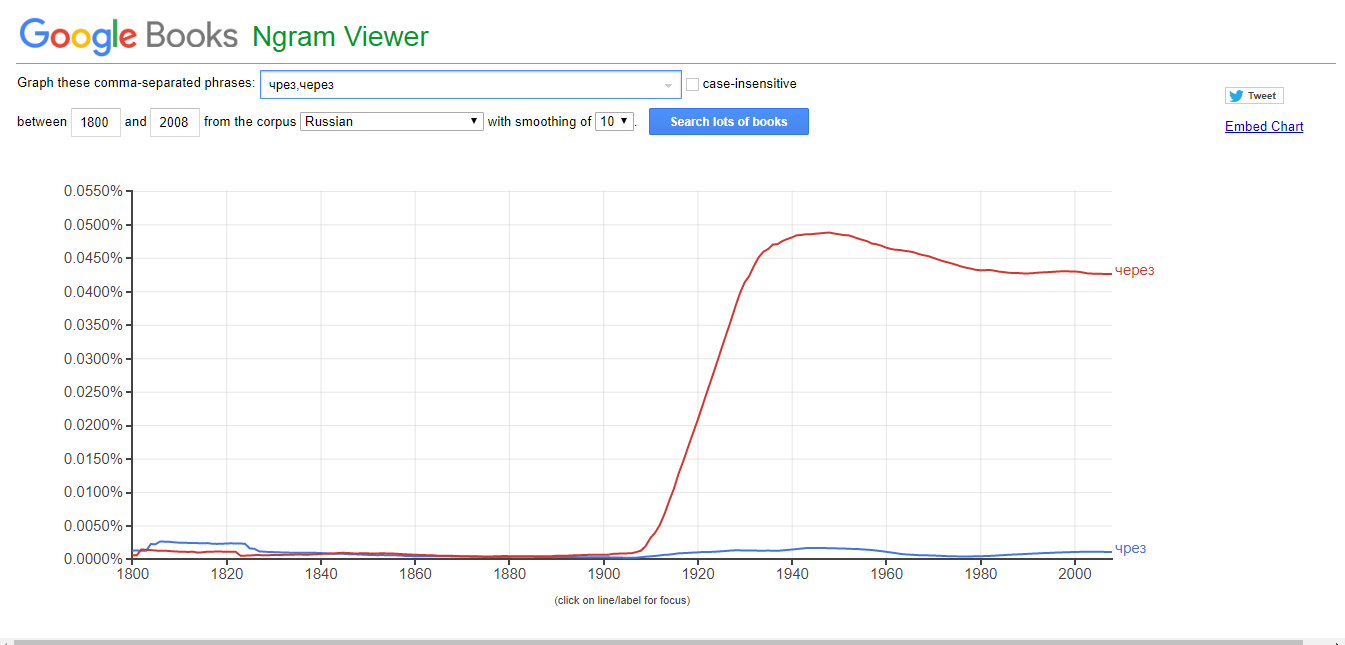
Зреть-видеть

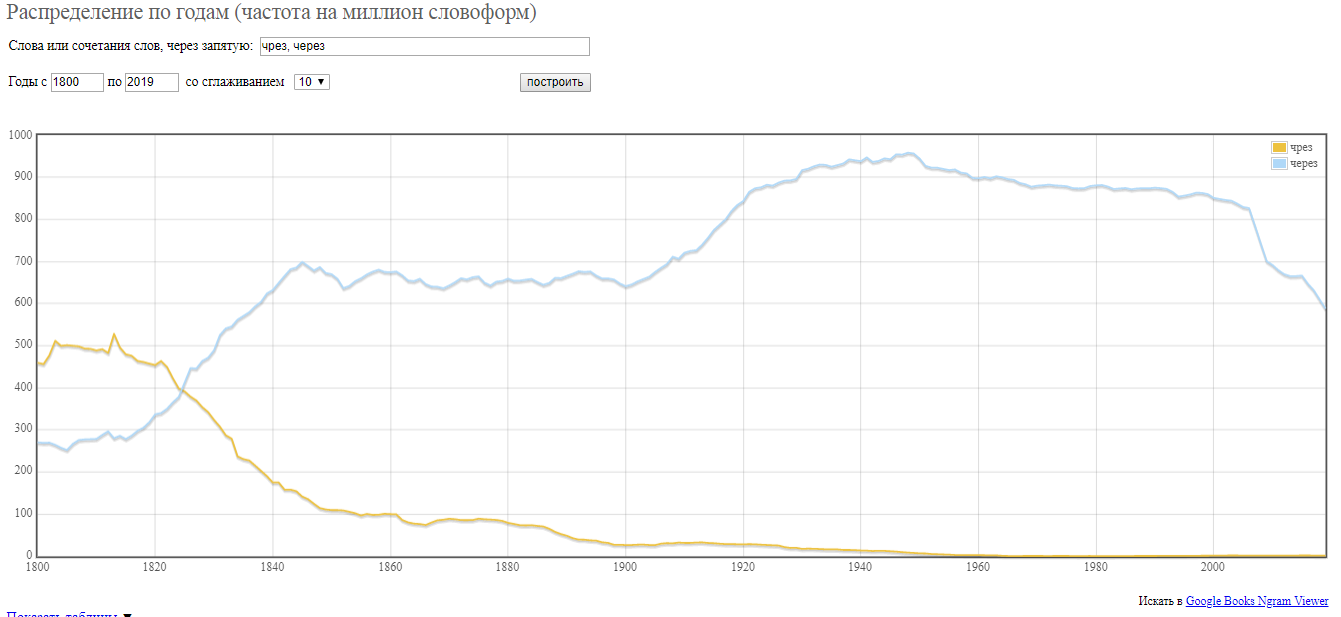
Google Ngrams. Слово зреть по графику практически никогда не употреблялось, а вот слово видеть имело периоды наибольшего (конец 19 века) и наименьшего(начало 19 века) употреблений

НКРЯ: Слово зреть по графику НКРЯ точно так же не употреблялось (за искл начала 19 века). Однако график слова видеть расходится с графиком Google Ngrams в том, что оно пошло на спад после начала 19 века.

Вывод: Оба графика иллистрируют то, что слово устаревшее слово зреть не употреблялось даже в 19 веке. Расхождения в употреблении слова видеть, скорее всего, основаны на различных текстах, из которых состоят корпуса.

3) Чрез – через (через входит в стоп-слова, но в нашем тексте был употреблен его устаревший вариант)





Чрез-через

Google Ngrams: График показывает, что до начала 20 века эти предлоги употрелялись почти одинаково. В период с 1800 по 1820 слово чрез даже преобладало, однако после начала 20 века

заметен резкий взлет в употреблении предлога через.

НКРЯ: Такое же наблюдение можно сделать, глядя на график НКРЯ

Вывод: Оба графика являются наглядным примером того, как неполногласный вариант слова стал вытесняться полногласным и почему этот процесс более заметен в начале 20 века.

**Общий вывод**: Хоть и графики Google Ngrams и НКРЯ иногда не согласуются полностью, они позволяют хронолгически и частотно определить причины в наибольшем и наименьшем употреблении слов

и вытеснении одного слова другим.

*3Посчитайте ipm для каждого слова по основному корпусу в НКРЯ*

IPM

**1. зреть** - 170 словоупотреблений, количество слов в корпусе 283 431 966

IPM=0, 5997912035

Округленно=0,6

vs

**видеть** - 40 198 словоупотреблений

IPM=141,825922345

Округленно=142

**2. чрез** -12 260 словоупотреблений

IPM=43, 2555303236

Округленно=43,3

vs

**Через**- 222 655

IPM=785, 5677083368

Округленно=786

**3. врата**

IPM=3, 70106544

Округленно=3, 7

vs

**Ворота**-12 186

IPM=42, 9944447409

Округленно=43

**3Использование инструментов корпусного анализа в профессиональной деятельности.**

Исследование корпуса в понимании текстп помогло на нескольких уровнях.

Во-первых, при лемматизации текста с помощью Mystem и дальнейшем просмотре частотных слова при добавлении стоп-слов, мы можем приблизительно определить по

нескольким первым десяткам слов основную тему.

Во-вторых, конкордансы устаревших слов дают список их употреблений в определенном контексте.

В-третьих установление количества слов в ngram-е, построение списка коллокатов определенного слова помогает определить взаимное употребление слов и

устойчивые сочетания слов в этом тексте.

Анализ графиков в Google N-Grams и НКРЯ было полезным в определении того, как и почему было использовано устаревшее слово и употреблялось ли оно в этот период.

Вычисленная нами частота ipm помогла в рассмотрении употребления слова другими авторами.

Таким образом, исследование корпуса и применение инструментов в некоторой степени было новым в плане изучения текста путем применения технологий. Определение

темы, ключевых слов, слов в контексте и т.д существенно экономит время исследователя.