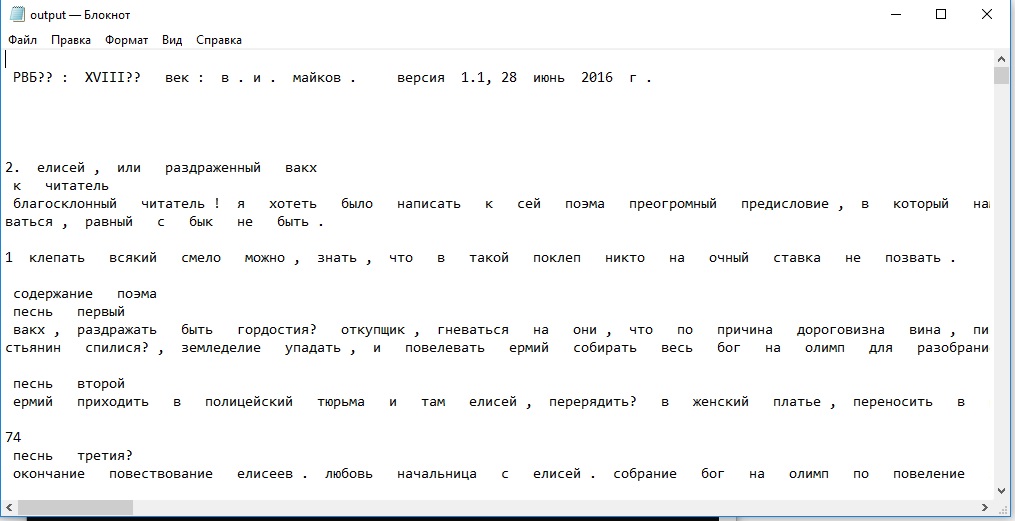
Горбунова Елизавета  
БФИЛЛ185

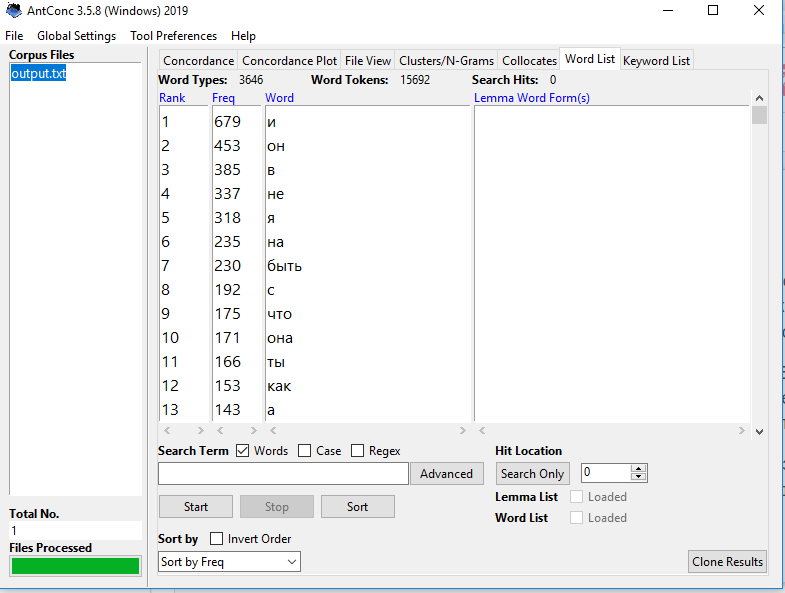
Цифровая грамотность. ДЗ №3

1. AntConc

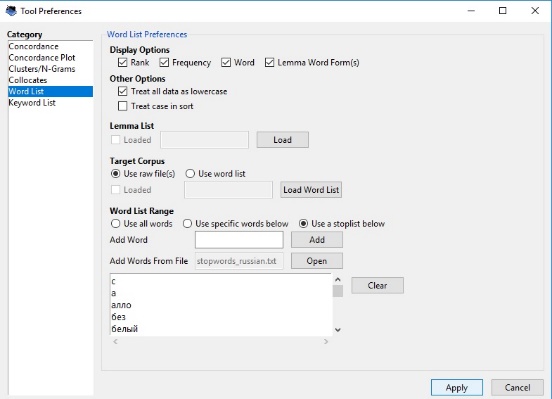
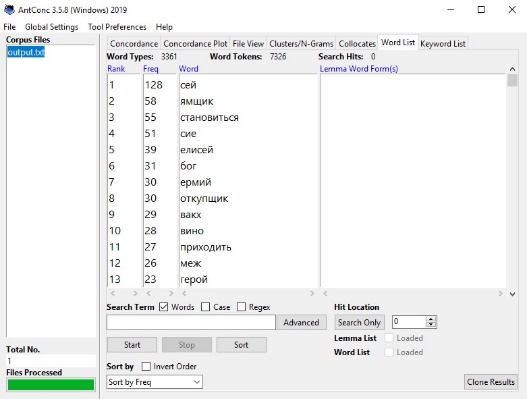
1. Создайте txt с вашим текстом (полное произведение), лемматизируйте его с помощью Mystem (презентация 8). Формат текста - UTF-8, проверить можно в любом текстовом редакторе (Sublime Text, NotePad++)



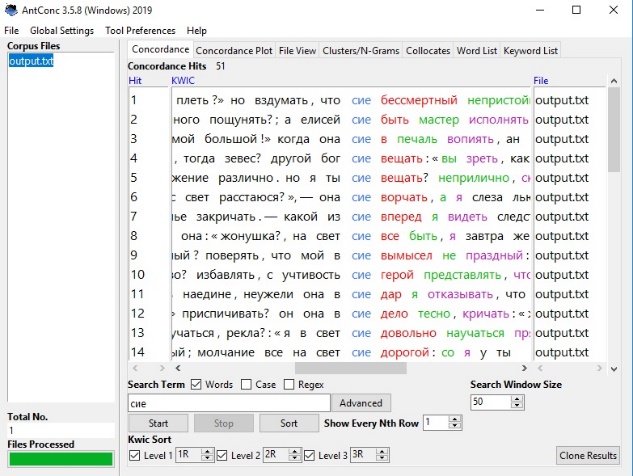
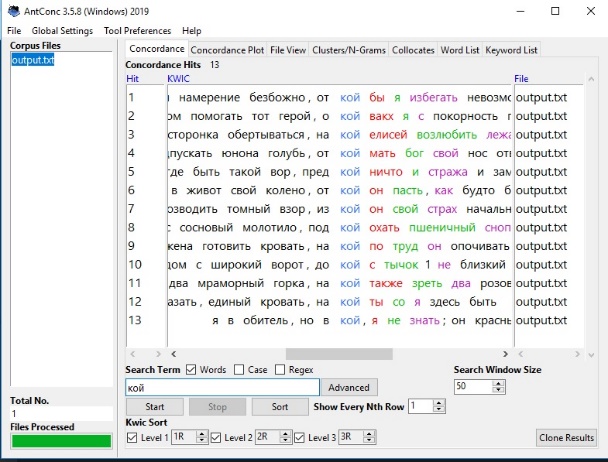
1. Откройте его в AntConc, посмотрите самые частотные слова (Word List).

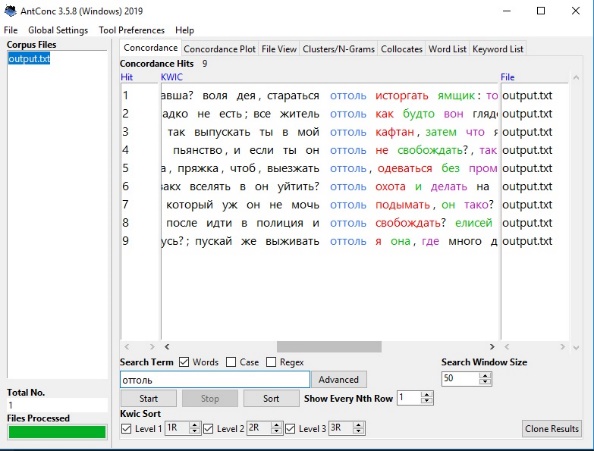


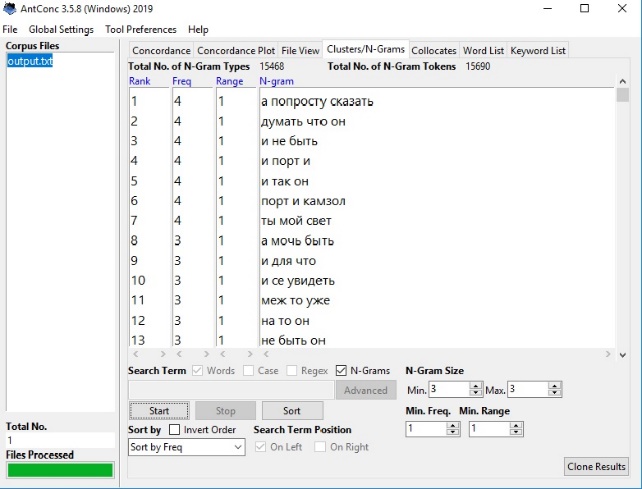
1. Добавьте список стоп-слов

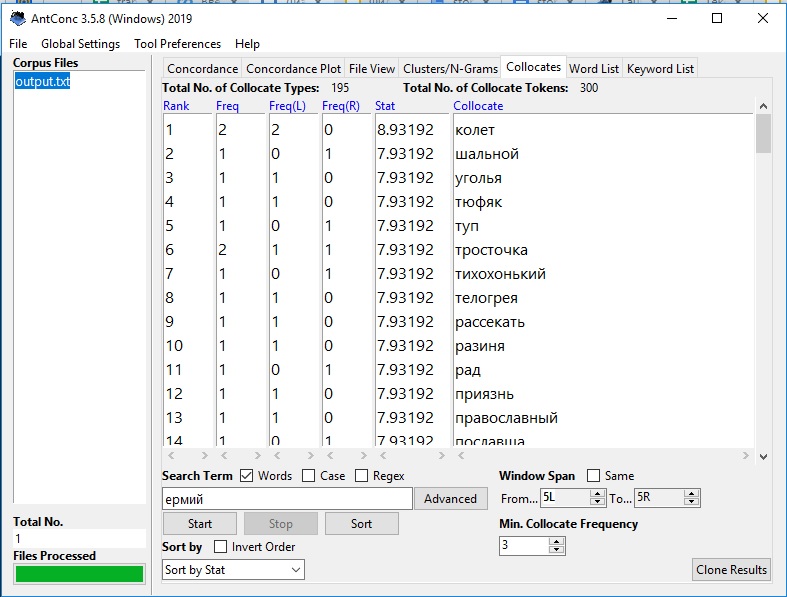
1. Кликнув на 3 устаревших слова, посмотрите их конкордансы.



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 

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



2. Google Ngrams и НКРЯ

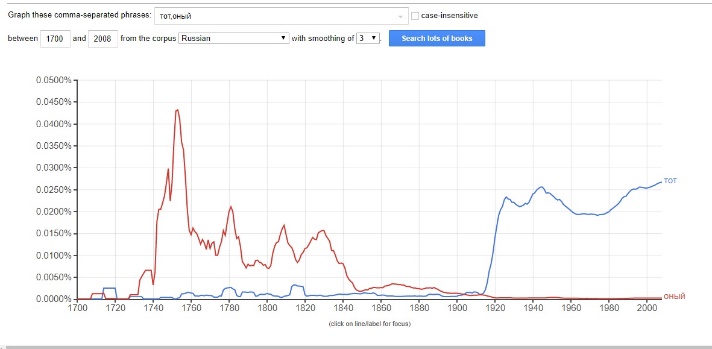
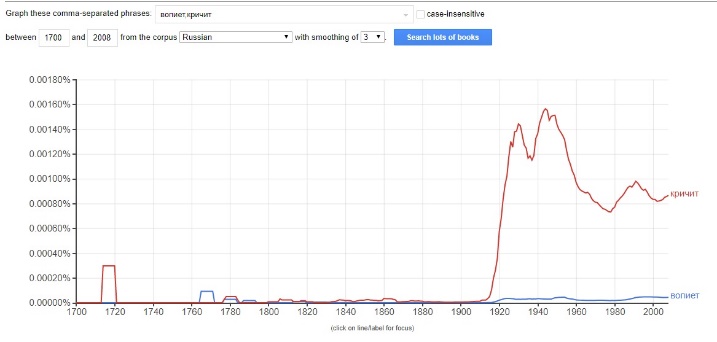
1. Устаревшие слова из отрывка: сие, порты, вопиет, оный

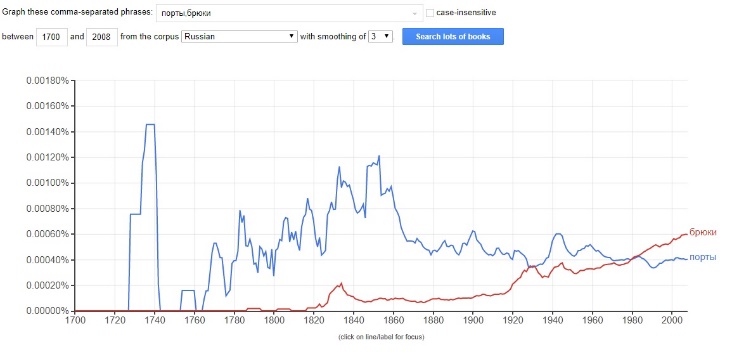
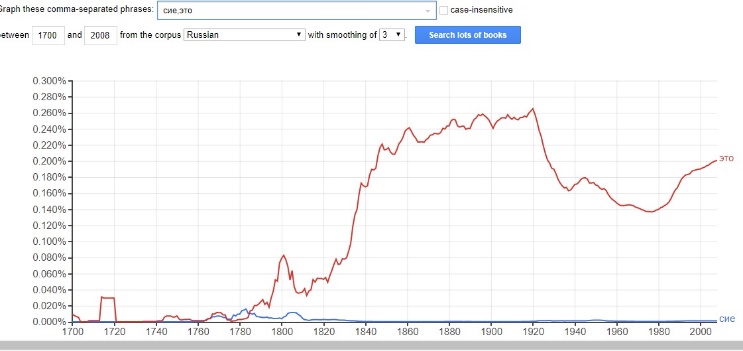
2. До 1990 слово «оный» использовалось больше, чем слово «тот»

Слово «вопиет» было менее употребимым практически все время, кроме 1760-1780х годов.

Для обозначения брюк слово «порты» было более популярным вплоть до 1980-х годов

Слово «это» всегда было более упортебимым, чем «сие»

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

сие = 17 031/283 431 966 \* 1 000 000 = 60,08

порты = 1 427/283 431 966 \* 1 000 000 = 5,03

вопиет = 405/ 283 431 966 \* 1 000 000 = 1,42

оный = 2 331/ 283 431 966 \* 1 000 000 = 8,22

3. Использование инструментов корпусного анализа в профессиональной деятельности.

Использование корпуса упрощает намного исследование, так как многие процессы автоматизированы и на них не приходится затрачивать ресурсы. Я считаю НКРЯ несомненно полезным ресурсом для филологов и лингвистов.