

Отчет hw5

Задание 1.

Основной подкорпус:

1. 18в.

- Я создала подкорпус с 1700 по 1799 год (18в) и отметила точное вхождение как обязательное. Объем подкорпуса 4 565 776 слов, число вхождений 3 625.

$$\text{Imp} = 3\,625 : 4\,565\,776 * 10\,000\,000 = 793.95047 \text{ (частота)}$$

Далее по аналогии:

- «сейчас»
 $\text{Imp} = 37 : 4\,565\,776 * 10\,000\,000 = 8.10377031$

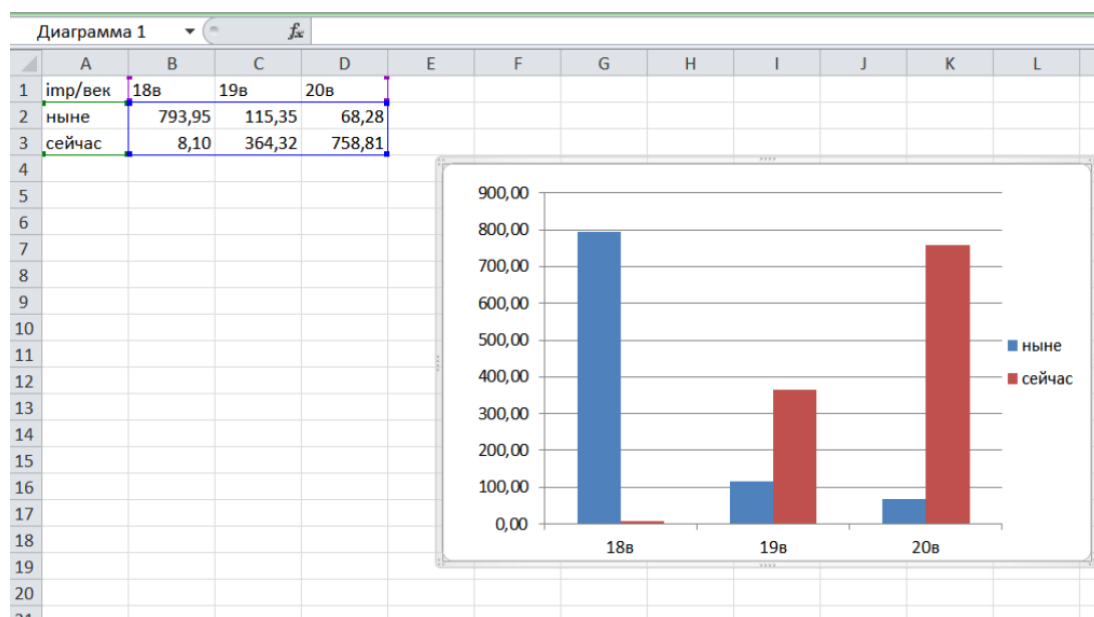
2. 19в.

- «ныне»
 $\text{Imp} = 6\,090 : 52\,797\,569 * 10\,000\,000 = 115.34622$
- «сейчас»
 $\text{Imp} = 19\,235 : 52\,797\,569 * 10\,000\,000 = 364.316016$

3. 20в.

- «ныне»
 $\text{Imp} = 10\,122 : 148\,244\,669 * 10\,000\,000 = 68.2790151$
- «сейчас»
 $\text{Imp} = 112\,490 : 148\,244\,669 * 10\,000\,000 = 758.813121$

Я создала диаграмму, отображающую частоту употребления данных слов с 18 по 20 века:



Поэтический подкорпус:

1. 18в.
 - «ныне» = 656.436375
 - «сейчас» = 13.194701
2. 19в.
 - «ныне» = 241.28766

Например: А. К. Толстой. Сказка про короля и про монаха (1840-1875)

Ныне ж, лишь только придут, ото всех отберет он бумаги,

Бросит под стол и велит принести побольше наливки.

- «сейчас» = 97.1381074
3. 20в.
 - «ныне» = 137.440845
 - «сейчас» = 235.24118

Например: М. Вега. Атлантида (1968-1979) [омонимия не снята] Все примеры (2)

АТЛАНТИДА

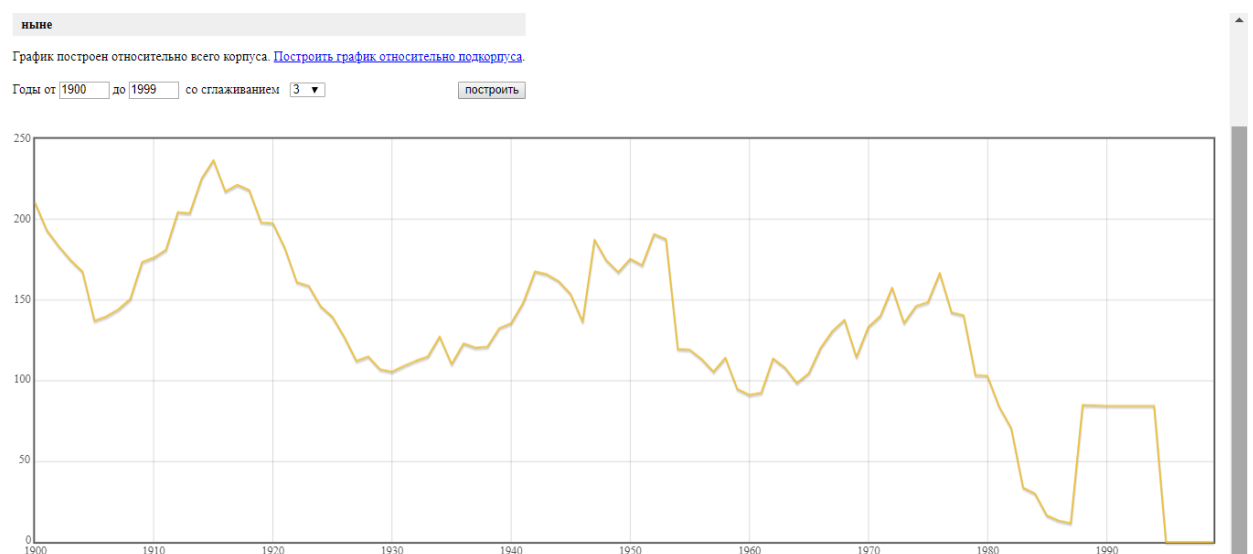
Сейчас — «старье», а завтра — «старина».

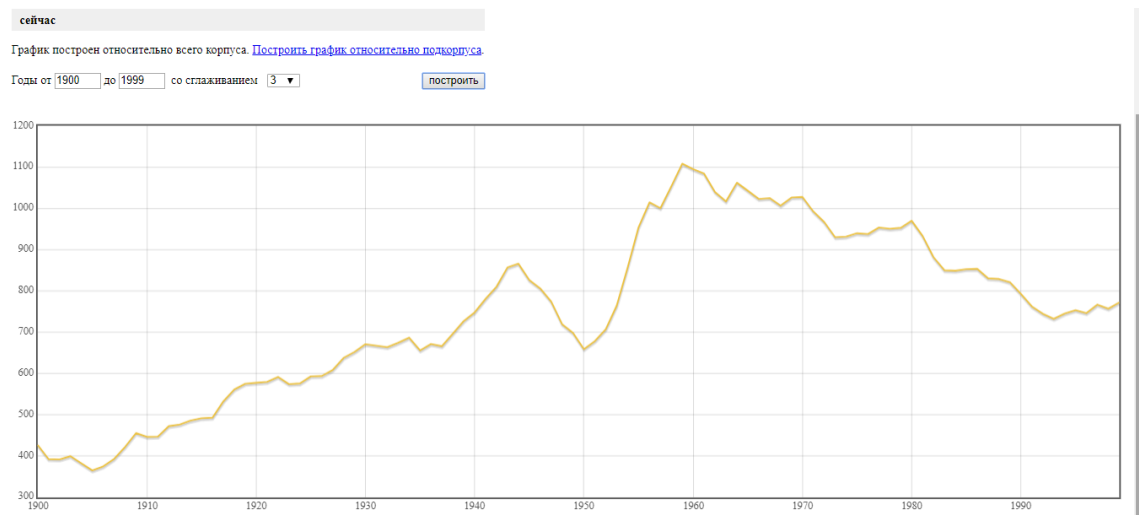
Сейчас не нужно, завтра драгоценно

Выводы:

Исходя из полученных данных мы видим, что:

- 1) Согласно графику, с 18 по 20 век число употреблений слова «ныне» постепенно падает, слово становится устаревшим (в основном корпусе в 1918 частота составляла 141.81, а в 1965 – 36.95) когда как «сейчас» становится всё популярнее (в 1904 составляла 400.16, а в 1960 - 1094). С помощью распределения по годам частоту можно увидеть наглядно:



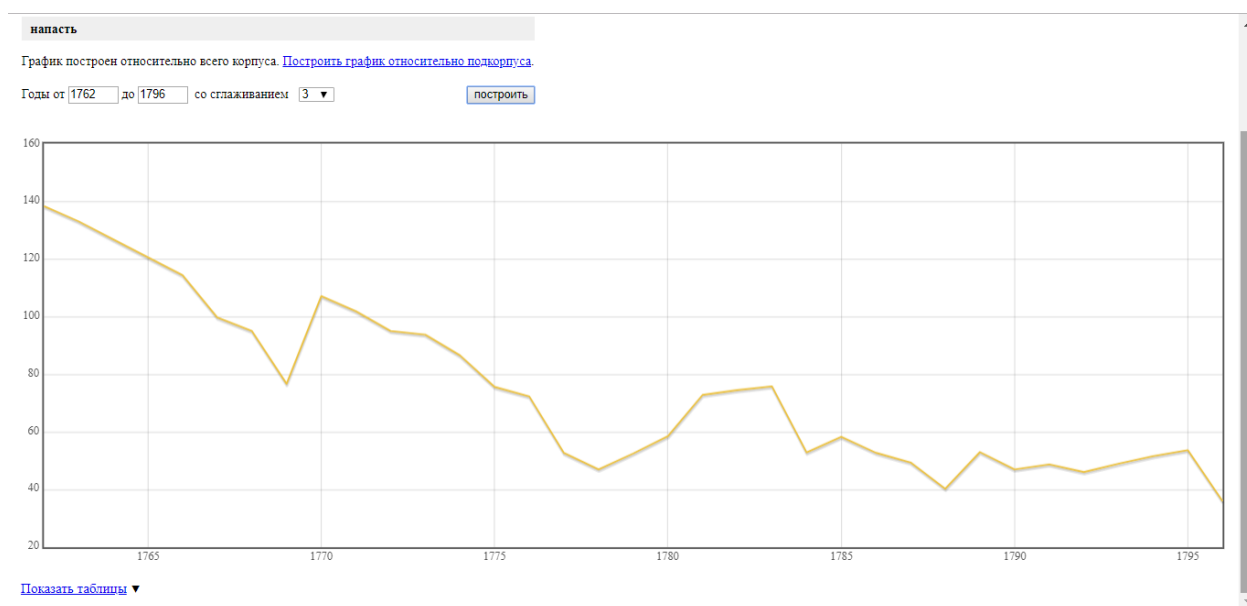


- 2) В поэтическом подкорпусе наблюдаются подобные процессы, однако популярность слова «ныне» не так стремительно снижается, из чего мы делаем вывод, что в поэзии устаревшие слова более используемы, чем в речи в целом.

Задание 2.

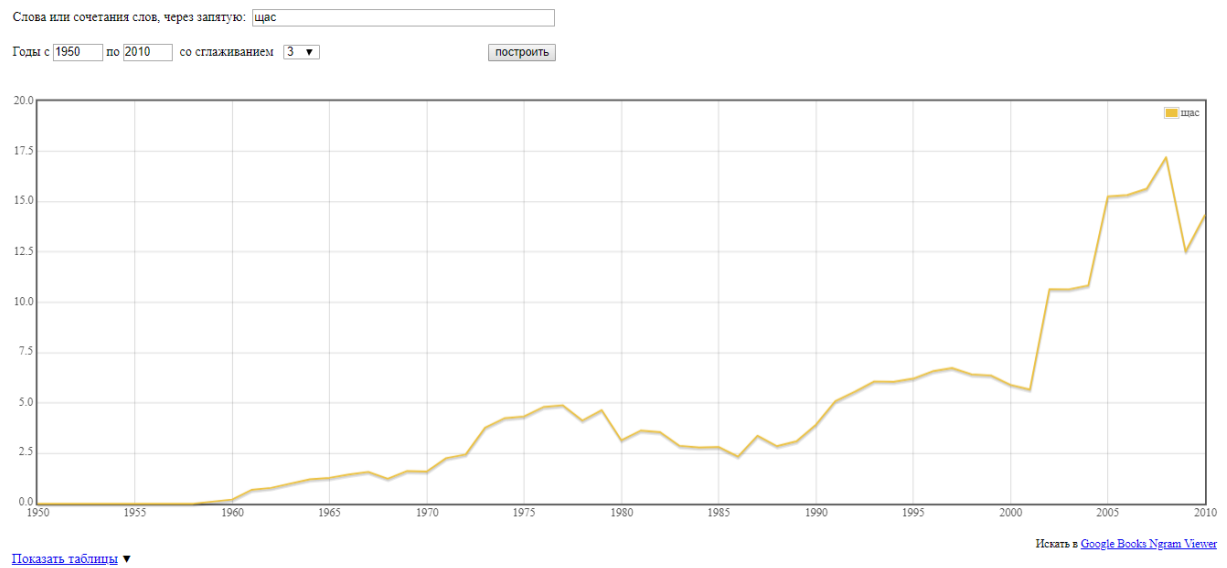
Екатерина 2 правила с 1762 по 1796. Я задала подкорпус, ограниченный этими датами.

- 1) Первая цитата не может относиться к этому времени поскольку вхождений не найдено ни у фразы «пиши пропало», ни у слова «уперла».
- 2) Вторая цитата также не подходит, потому что слова «лопух» и «немыслимо» не были в употреблении (вхождений не найдено).
- 3) Третий диалог мог встретиться в таком виде во времена Екатерины 2, потому что использовалось слово «растолковать» (33 вхождения), «напасть» (242 вхождения) и др. Ниже представлен график использования слова «напасть» в данный период:



Задание 3.

График употребления слова «щас» с 1950 по наши дни(2010):



Частота употреблений данного слова постепенно росла в течение 60 лет. Возможно, это объясняется развитием технологий и растущей популярностью сокращений.

- 1) Мы видим, что в жанре детектива и боевика число вхождений довольно высокое – 85.

Частота = $85 : 6\,928\,811 \text{ слов} * 10\,000\,000 = 12.2676$

Это объясняется высоким количеством жаргонизмов в текстах данного жанра.

- 2) В детской литературе слово «щас» используется намного реже (10 вхождений), и во всех случаях данное слово – часть цитаты. Следовательно, это слово не популярно для использования в детских книгах.
- 3) В фантастических и приключенческих книгах (вместе) найдено 56 вхождений. Частота = 7.186 может объясняться развлекательным характером данного жанра литературы.
- 4) В исторической и документальной же прозе (вместе) число вхождений самое низкое (6) и все они также включены в цитаты. Это объясняется официальным и научным характером данных текстов.