

Домашнее задание 5

(Камарова Эндже БИЯ1712)

Задание 1.

1. Первым делом в Национальном корпусе русского языка (в основном корпусе) я создала три подкорпуса, каждый из которых составляет документы определенных веков (18, 19, 20). В таблицу занесла количество слов в заданных подкорпусах, то есть количество слов за 18 век, за 19 век и за 20 век.

Объем основного корпуса			
	18 век	19 век	20 век
объем корпуса	5 325 444	55 166 493	158 277 632

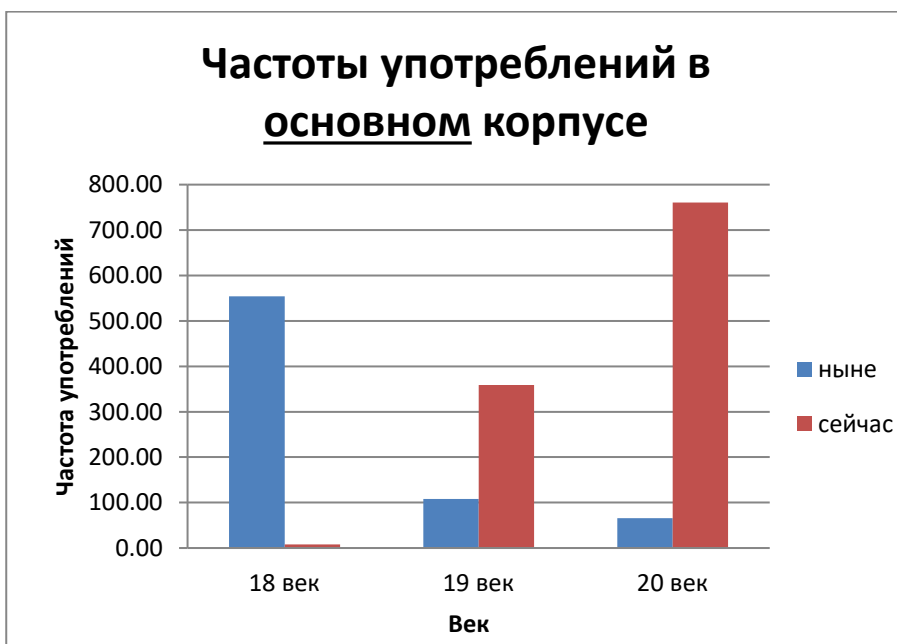
2. Далее нашла и добавила в таблицу количество вхождений слов «ныне» и «сейчас» в каждом из веков (в каждом подкорпусе).

Основной корпус (вхождения)			
	18 век (1701-1800)	19 век (1801-1900)	20 век (1901-2000)
ныне	2950	5 975	10 346
сейчас	43	19 787	120 408

3. По формуле (частота ipm определяется как количество употреблений слова за год, поделенное на объем корпуса за этот год (1 век) и умноженное на 1 миллион) рассчитала частоты употреблений слов «ныне» и «сейчас» в каждом из веков.

Основной корпус (частоты употреблений)			
	18 век	19 век	20 век
ныне	553,94	108,31	65,37
сейчас	8,07	358,68	760,74

4. Построила столбчатую диаграмму для частот употребления слов в разное время.



5. **Вывод (по основному корпусу):** Построив диаграмму, наглядно видно, что частота употребления слова «ныне» с каждым веком уменьшается, особенно видна разница между 18 и 19 веком, так как частота резко падает. С другой стороны, частота употребления слова «сейчас» растет с каждым веком. Если в 18 веке это слово употреблялось очень редко, то мы можем заметить, что в 20 веке употребление этого слова превысило употребление слова «ныне» в 18 веке. А также, мы можем сказать, что употребление слова «сейчас» увеличивалось почти в два раза с каждым веком.

Более того, по диаграмме видно, что в 18 веке частота употребления слова «ныне» намного превышала частоту слова «сейчас». Однако, в 19 и 20 веках ситуация изменилась: здесь частота употребления слова «ныне» намного меньше, чем частота слова «сейчас».

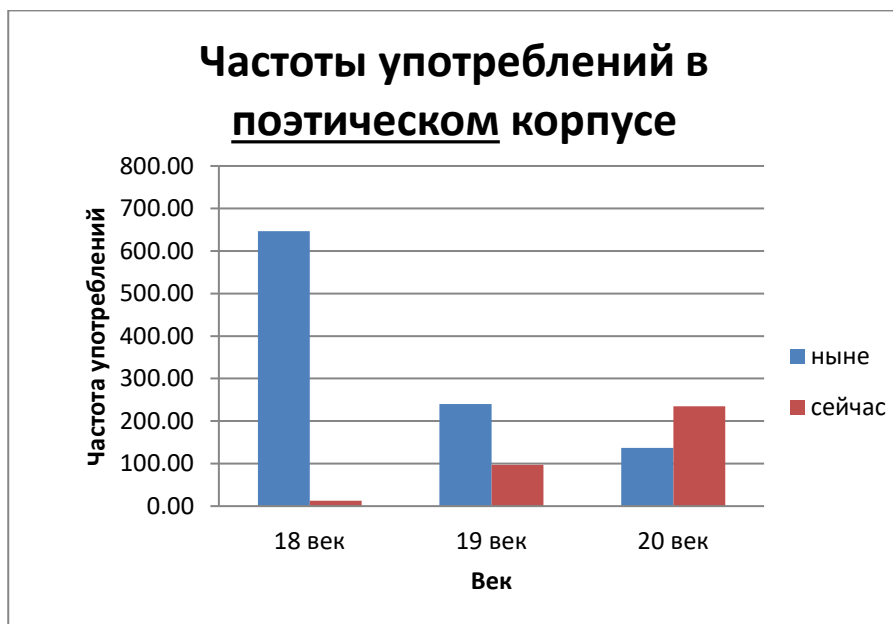
6. Далее проделала ту же работу, которая описывается выше в пунктах 1-4, но только в поэтическом корпусе.

Объём поэтического корпуса			
	18 век	19 век	20 век
объём корпуса	921 424	3 572 009	6 501 047

Поэтический корпус (вхождения)			
	18 век (1701-1800)	19 век (1801-1900)	20 век (1901-2000)
ныне	596	858	889
сейчас	12	347	1526

Поэтический корпус (частоты употреблений)			
	18 век	19 век	20 век
ныне	646,82	240,20	136,75
сейчас	13,02	97,14	234,73

7. Построила столбчатую диаграмму для частот употребления слов в разное время в поэтическом корпусе для большей наглядности в сравнении частот употребления с основным корпусом.



8. Вывод (по сравнению частот употребления слов в основном и поэтическом корпусах): во-первых, как и в основном корпусе, мы можем заметить, что частота употребления слова «ныне» уменьшается с каждым веком, а частота слова «сейчас», наоборот, растет. Во-вторых, в 18 веке частота употребления слова «сейчас» очень мала, а частота слова «ныне» довольно велика, как и в основном подкорпусе. Но видна небольшая разница между корпусами: мы видим, что частоты двух рассматриваемых слов в поэтическом корпусе в 19 и 20 веках отличаются не намного, хотя в основном корпусе разница между ними была намного значительнее. А также, если в основном корпусе в 19 веке частота употребления слова «сейчас» уже превышала частоту слова «ныне», то в поэтическом корпусе она превысит частоту слова «ныне» только в 20 веке. Более того, слово «сейчас» в поэтическом корпусе росло не такими большими темпами, как в основном корпусе.

Задание 2.

1. Во-первых, время правления Екатерины Второй - с 1762 по 1796 год.
2. Для начала я задала подкорпус с временным ограничением (1762-1796). Далее искала слова и словосочетания из диалогов и смотрела, употреблялись ли они в то время или нет (есть ли они в корпусе или нет).
- 3.

Первый диалог:

(— Отдай!

— Пиши пропало. Что пряжка, третьего дня эта поганка у меня с груди звезду Александра Невского уперла! Любит, сволочь, блестящее.)

Данный диалог в таком виде не мог встретиться во времена правления Екатерины Второй, так как не употреблялись такие слова и словосочетания, как: «пиши пропало», «поганка», «уперла». Остальные слова и словосочетания можно встретить по крайней мере 1 раз.

Второй диалог:

(— Ох, ну просто немыслимо! Нет, как можно быть таким лопухом?!)

Данный диалог в таком виде не мог встретиться во времена правления Екатерины Второй, так как не употреблялись такие слова и словосочетания (их нет в заданном мною подкорпусе), как: «ну просто», «немыслимо», «лопух».

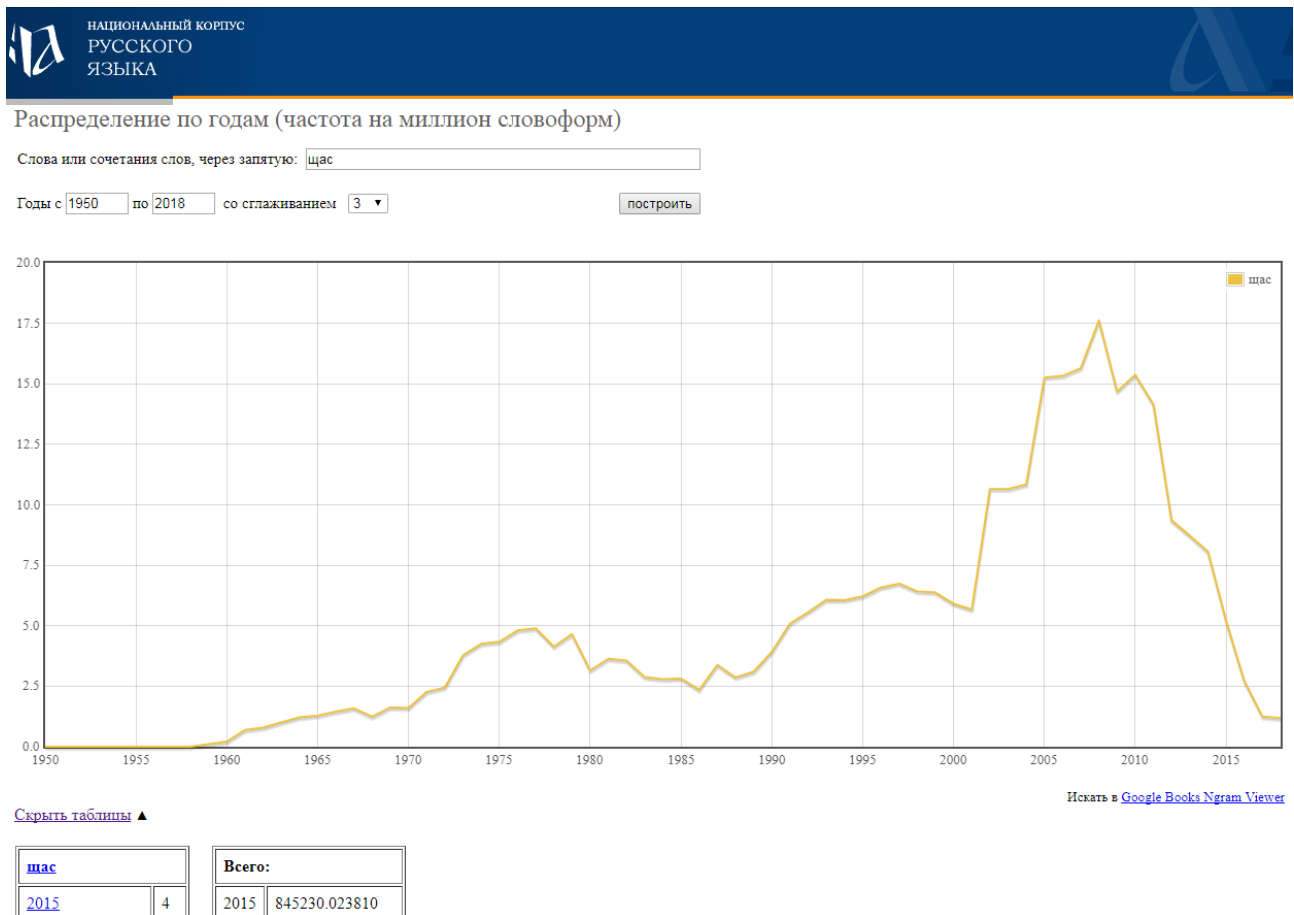
Третий диалог:

(— Я гнева вашего никак не растолкую. Он в доме здесь живет, великая напасть!)

Данный диалог в таком виде не мог встретиться во времена правления Екатерины Второй, хотя в этом диалоге в отличие от других все отдельные слова уже можно было встретить, но некоторые сочетания не нашлись в заданном мною подкорпусе, такие как: «великая напасть». Следовательно, можно предположить, что именно в таком виде, в таких сочетаниях, они встретиться не могли.

Задание 3.

1. Чтобы увидеть динамику употребления слова «щас» в период с 1950 года до наших дней (2018 года), построила график в основном корпусе.



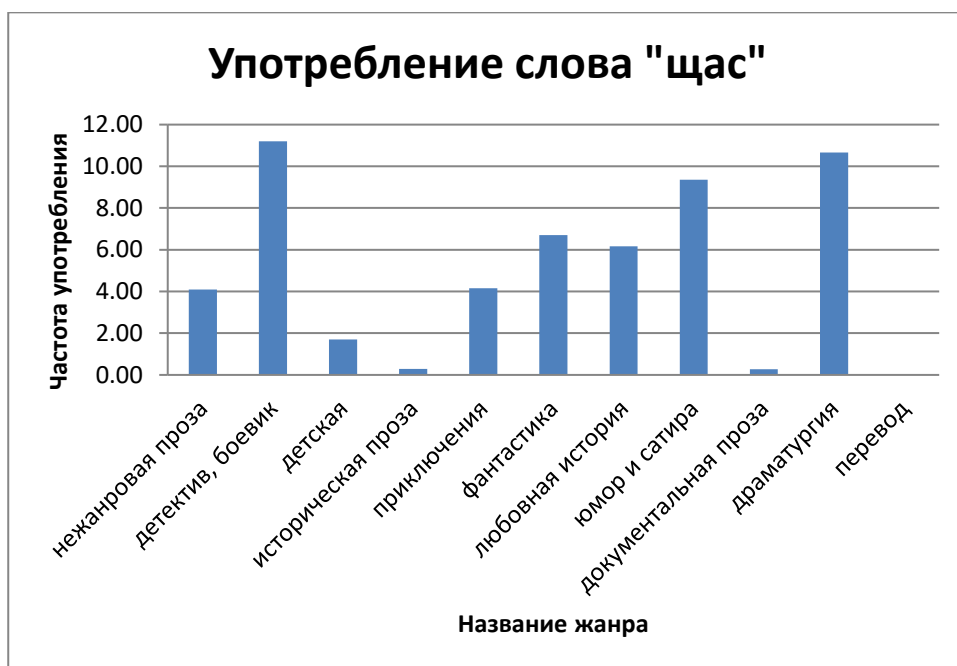
Вывод по графику: по нему мы можем заметить, что употребление слова «щас» росло с каждым годом, хотя в некоторые периоды времени употребление снижалось, но незначительно. Пика своего употребления это слово, согласно графику, достигло в 2008 году.

Однако, после 2008 года, начинается спад в употреблении данного слова, хотя и есть незначительные увеличения в частоте употребления заданного слова «щас».

2. Далее я посмотрела употребление слова «щас» в различных жанрах, создавая новые подкорпусы, добавила все результаты в таблицу. В данной ниже таблице я внесла значения объёма подкорпусов различных жанров (11 жанров), вставила значения количества вхождений слова «щас» в каждом из подкорпусов, а также, посчитала по формуле частоты употребления слова «щас» в различных жанрах.

Вхождения			
Жанр	Объём подкорпуса	Количество вхождений	Частота употребления
нежанровая проза	69 230 465	283	4,09
детектив, боевик	7 587 093	85	11,20
детская	5 868 568	10	1,70
историческая проза	10 367 857	3	0,29
приключения	3 610 684	15	4,15
фантастика	8 199 359	55	6,71
любовная история	973 982	6	6,16
юмор и сатира	4 595 212	43	9,36
документальная проза	10 808 635	3	0,28
драматургия	3 378 029	36	10,66
перевод	85 093	0	0,00

3. Построила столбчатую диаграмму на основе полученной таблицы и полученных частот употребления.



4. **Вывод:** по полученной диаграмме наглядно видно, что употребление слова «щас» наиболее распространено в детективах и боевиках, чуть менее употребляемо в драматургии, далее по мере снижения распространённости слова «щас» идут соответственно юмор и сатира, фантастика, любовная история, приключения, нежанровая проза. Совсем редко употребляемым слово является в таких жанрах, как детская, историческая проза и документальная проза. Слово «щас» совсем не встречается в таком жанре произведений, как перевод. Возможно, это связано с тем,

что в переводе используют часто официальный стиль, в котором слово «щас» не встречается. То же самое относится к документальной и исторической прозе, в которых частота употребления данного слова очень мала, так как в таких жанрах обычно разговорные слова не употребляются.

Задание 4.

1. В основном подкорпусе я нашла сочетания слова «язык» с существительными в родительном падеже (S, gen).
2. Далее скачала выдачу в формате Excel.
3. Отсортировала строки по столбцу F, где находились существительные в Р.п.
4. На основе выдачи можно сделать такой **вывод**, что больше всего упоминаются такие сочетания слова «язык» с существительными в Р.п., как: язык мира (встречается 63 раза), язык программирования (62 раза), язык описания (39 раз).

Также я заметила, что много сочетаний, встречающихся около 20 раз, например, язык жестов (24), язык математики (19), язык народа (народов) (20), язык науки (20), язык общения (20) и др.

5. Однако, если искать сочетания слова «язык» с существительными в родительном падеже в биграммах, то мы можем обнаружить, что наиболее употребляемым является такое сочетание, как «язык народов», оно встречается в 46 документах, имеется 64 вхождения.