

HW5. Работа с НКРЯ.

Выполняла: студентка 1 курса НИУ ВШЭ, группы 1712-2
Добровольская Яна Владиславовна

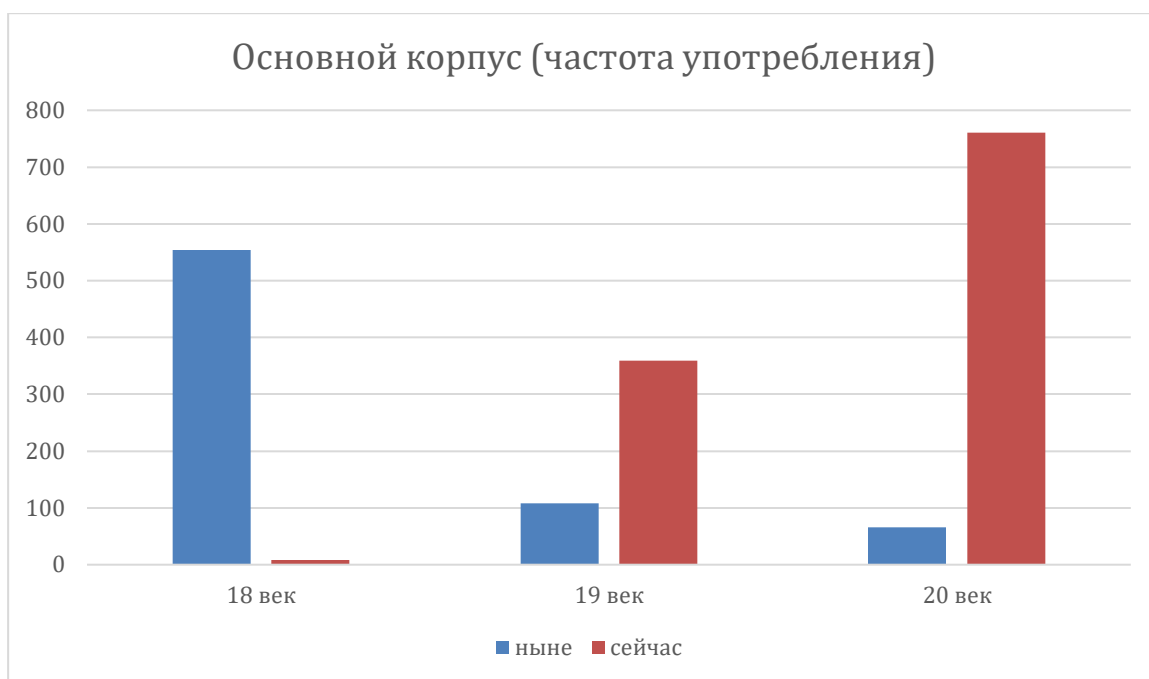
Задание 1.

1. Для того, чтобы сравнить частоту употребления слов «ныне» и «сейчас» в разных веках (18-20), каждый раз я задавала новый подкорпус (например, 1701-1800, 1801-1900 и 1901-2000). Там я смотрела на количество слов (объем подкорпуса) и количество вхождений. На основе данных я составила таблицу и построила диаграмму и сделала соответствующие выводы по ним.

Основной корпус (количество вхождений)			
	18 век	19 век	20 век
ныне	2950	5975	10346
сейчас	43	19787	120408
Основной корпус (количество слов)			
	18 век	19 век	20 век
	5325444	55166493	158277632

2. По формуле я высчитала частоту употреблений этих слов в основном корпусе. Я делила количество вхождений на объем корпуса за определенный век (количество слов) и умножала на 1 миллион.

Основной корпус (частота употребления)			
	18 век	19 век	20 век
ныне	553,94	108,31	65,37
сейчас	8,07	358,86	760,74

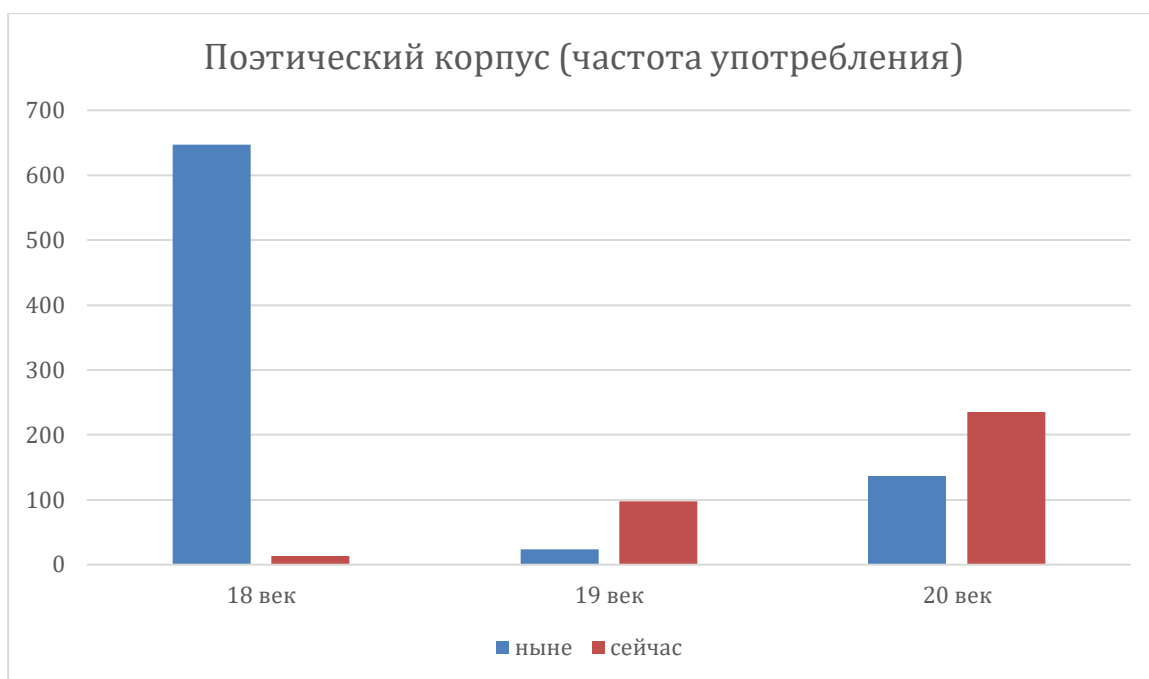


Вывод по диаграмме: в итоге наибольшая частотность «ныне» наблюдается в 18 веке, а наибольшая частотность «сейчас» в 20 веке, что вполне предсказуемо, учитывая то, что в настоящее время «ныне» можно встретить только в книгах.

3. Далее я сравнивала употребление «ныне» и «сейчас» в основном и поэтическом корпусах. Для этого я также составила таблицу и построила диаграмму.

Поэтический корпус (количество вхождений)			
	18 век	19 век	20 век
ныне	596	858	889
сейчас	12	347	1526
Поэтический корпус (слова в подкорпусе)			
	18 век	19 век	20 век
	921424	3572009	6501047

Поэтический корпус (частота употребления)			
	18 век	19 век	20 век
ныне	646,82	24,02	136,74
сейчас	13,02	97,14	234,73



Вывод: Если сравнивать с частотой употребления данных слов в основном подкорпусе, то можно выявить, что и в поэтическом и в основном корпусах самое частое употребление “ныне” было в 18 веке, а слова “сейчас” в 20 веке. И “ныне” и “сейчас” редко употреблялось в литературе 19го века.

Задание 2.

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

-Ох, ну просто немыслимо! Нет, как можно быть таким лопухом?!

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

Анализируя результаты поиска в НКРЯ, я обнаружила, что ни один из диалогов не мог быть употреблен во времена правления Екатерины Второй (1762-1796), так как такие слова и словосочетания не были найдены в процессе поиска:

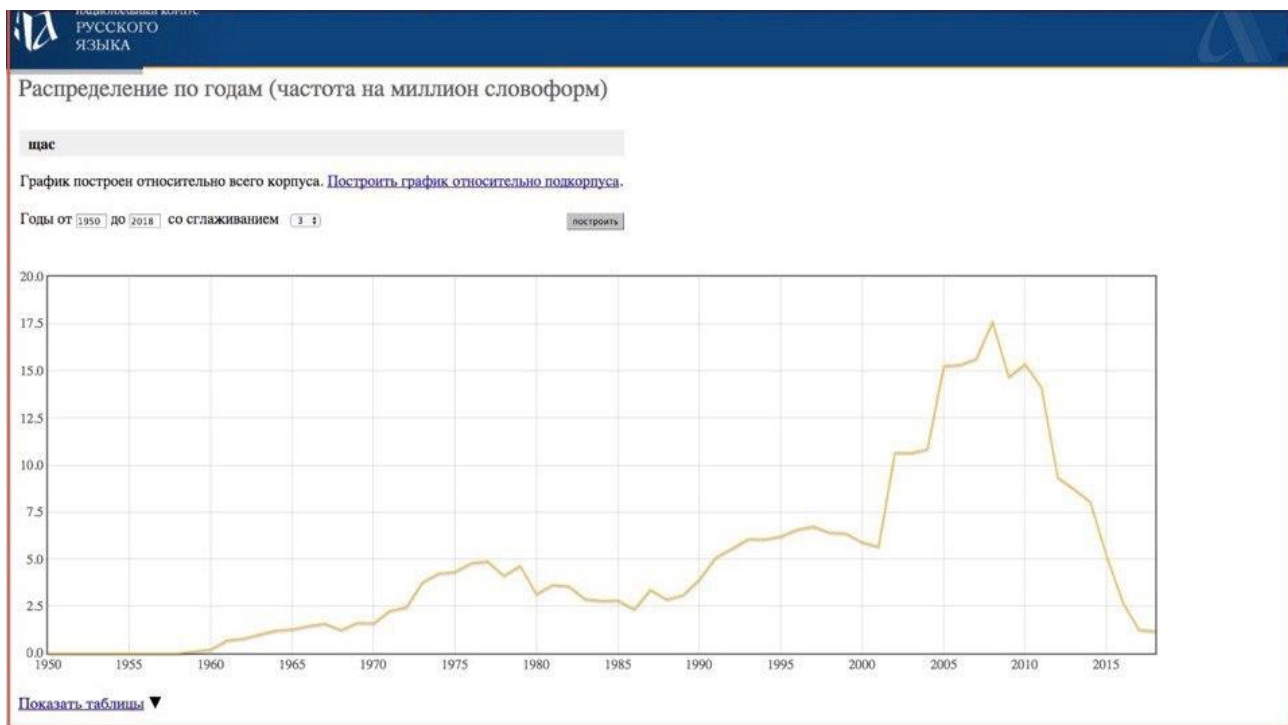
«пиши пропало», «поганка», «уперла», «немыслимо», «лопухом», «великая напасть».

Однако некоторые слова и словосочетания всё же встречались: «отдай» (сочетании с др. частями речи), «не растолкую»- вх.- 3, док.- 3, «можно быть»- вхождений 887, документов- 723, «здесь живет»- вх.-374, док.-310.

Задание 3.

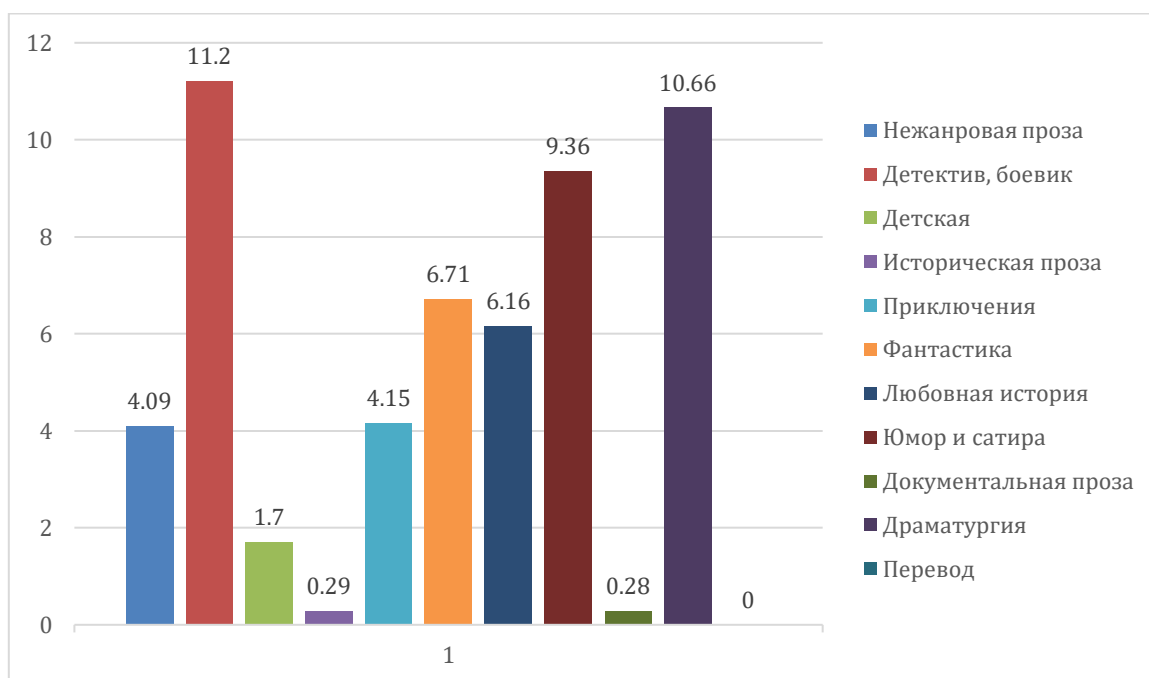
1. Для того, чтобы увидеть, как изменялось употребление слова «щас» с 1950 года по настоящее время, я построила график в основном корпусе.

Мой вывод по графику: здесь можно заметить, что употребление слова «щас» увеличивалось с каждым годом (были и исключения, когда оно уменьшалось). Самое частое употребление данного слова было зафиксировано в 2008 году.



2. После этого я смотрела употребление слова «щас» в различных жанрах (11 жанров). Для этого я создавала разные подкорпусы, записывая данные в таблицу. Сравнивая динамику употребления «щас» в разных жанрах, я опиралась на такие показатели как: количество вхождений, объем слов подкорпуса и частота употребления.

Жанр	Объем подкорпуса	Количество вхождений	Частота употребления
Нежанровая проза	69230465	283	4,09
Детектив, боевик	7587093	85	11,20
Детская	5868568	10	1,70
Историческая проза	10367857	3	0,29
Приключения	3610684	15	4,15
Фантастика	8199359	55	6,71
Любовная история	973982	6	6,16
Юмор и сатира	4595212	43	9,36
Документальная проза	10808635	3	0,28
Драматургия	3378029	36	10,66
Перевод	85093	0	0,00



Вывод по диаграмме: на диаграмме видно, что самое частое употребление слова «щас» замечено в детективах и боевиках (возможно из-за более частой разговорной речи в произведениях), меньше- в таких жанрах как юмор и сатира, приключения, фантастика, нежанровая проза и любовная история. Еще реже в детской, исторической и документальной прозах. Совсем не употребляется в переводе (вероятно из-за того, что это более официальный стиль, а «щас» употребляется больше в разговорном).

Задание 4.

Введя «Язык + сущ. в Родительном падеже (S, gen.), я получила результат в 9168 вхождений и 4306 документов. Я отсортировала таблицу по столбцу F от А до Я, в следствие чего первыми словами в столбце стали «аборигенов», «Аввакума» и «Августа». «Языков мира» стало самым популярным сочетанием и было замечено 63 раза. В биграммах же самым популярным сочетанием «языка» и существительного в Род. Падеже стало «язык народов»- оно было замечено в 46 документах и в 64 вхождениях.