

Отчет по работе с Национальным корпусом русского языка

Задание номер 1:

Частота употребления слов «сейчас» и «ныне» в 18-20 веках.

Сравнение частоты употребления слов «сейчас» и «ныне» в разные века в основном подкорпусе НКРЯ показало, что второе слово встречается значительно чаще, чем первое в XVIII веке. Но начиная с XIX века, употребление данных слов изменяется в противоположную сторону. Таким образом мы можем наблюдать возрастающую тенденцию в использовании слова «сейчас» на протяжении XIX-XX веков. Что касается поэтической литературы, то слово «ныне» на протяжении 3 веков было более употребляемо, чем слово «сейчас», но к XX веку использование «ныне» становится реже практически в 2 раза, чем в XIX веке.

Ход работы:

Для анализа частоты употребления слов «сейчас» и «ныне» в основном корпусе НКРЯ, был задан подкорпус по критерию «дата создания текста». Для поиска частоты употребления слов были поставлены даты с 1701 по 1800 годы для XVIII века, с 1801 по 1900 годы для XIX века и период с 1901 по 2000 год для XX века.

Поиск в основном корпусе:

Результаты поиска в основном корпусе

Объем всего корпуса: 115 645 документов, 23 803 881 предложение, 283 431 966 слов.

Поиск ведётся по пользовательскому подкорпусу объемом 1 809 документов, 253 475 предложений, 4 670 457 слов.

ныне

Найдено 607 документов, 2 804 вхождения.

Результаты поиска в основном корпусе

Объем всего корпуса: 115 645 документов, 23 803 881 предложение, 283 431 966 слов.

Поиск ведётся по пользовательскому подкорпусу объемом 1 809 документов, 253 475 предложений, 4 670 457 слов.

сейчас

Найдено 28 документов, 37 вхождений.

Результаты поиска в основном корпусе

Объем всего корпуса: 115 645 документов, 23 803 881 предложение, 283 431 966 слов.

Поиск ведётся по пользовательскому подкорпусу объемом 5 443 документа, 3 987 015 предложений, 53 488 208 слов.

сейчас

Найдено 1 892 документа, 19 496 вхождений.

[Распределение по годам](#) [Статистика](#)

Результаты поиска в основном корпусе

Объем всего корпуса: 115 645 документов, 23 803 881 предложение, 283 431 966 слов.

Поиск ведётся по пользовательскому подкорпусу объемом 5 443 документа, 3 987 015 предложений, 53 488 208 слов.

ныне

Найдено 1 159 документов, 5 782 вхождения.

[Распределение по годам](#) [Статистика](#)

Результаты поиска в основном корпусе

Объем всего корпуса: 115 645 документов, 23 803 881 предложение, 283 431 966 слов.

Поиск ведётся по пользовательскому подкорпусу объемом 69 449 документов, 13 820 280 предложений, 154 876 648 слов.

ныне

Найдено 4 039 документов, 10 189 вхождений.

[Распределение по годам](#) [Статистика](#)

Результаты поиска в основном корпусе

Объем всего корпуса: 115 645 документов, 23 803 881 предложение, 283 431 966 слов.

Поиск ведётся по пользовательскому подкорпусу объемом 69 449 документов, 13 820 280 предложений, 154 876 648 слов.

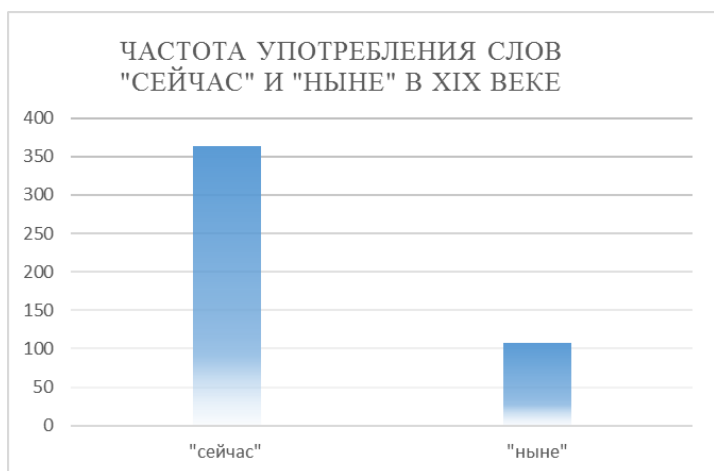
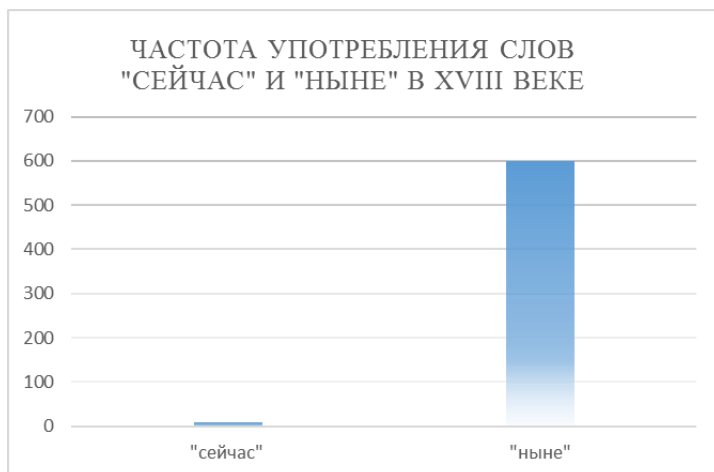
сейчас

Найдено 13 224 документа, 117 877 вхождений.

[Распределение по годам](#) [Статистика](#)

Частота употребления вычислялась по формуле:

$\text{количество вхождений} / \text{количество слов в подкорпусе} * 1000000$





Аналогичная работа с поиском частоты употребления слов «сейчас» и «ныне» была проделана в поэтическом подкорпусе НКРЯ.

XVIII век:

Результаты поиска в поэтическом корпусе

Объем всего корпуса: 78 868 документов, 1 045 898 предложений, 10 967 173 слова.
Поиск ведётся по пользовательскому подкорпусу объемом 3 159 документов, 62 855 предложений, 921 424 слова.
сейчас
Найдено 8 документов, 12 вхождений.

Результаты поиска в поэтическом корпусе

Объем всего корпуса: 78 868 документов, 1 045 898 предложений, 10 967 173 слова.
Поиск ведётся по пользовательскому подкорпусу объемом 3 159 документов, 62 855 предложений, 921 424 слова.
ныне
Найдено 354 документа, 596 вхождений.

XIX век:

Результаты поиска в поэтическом корпусе

Объем всего корпуса: 78 868 документов, 1 045 898 предложений, 10 967 173 слова.
Поиск ведётся по пользовательскому подкорпусу объемом 19 168 документов, 315 152 предложения, 3 572 009 слов.
сейчас
Найдено 222 документа, 347 вхождений.

Результаты поиска в поэтическом корпусе

Объем всего корпуса: 78 868 документов, 1 045 898 предложений, 10 967 173 слова.
Поиск ведётся по пользовательскому подкорпусу объемом 19 168 документов, 315 152 предложения, 3 572 009 слов.
ныне
Найдено 656 документов, 858 вхождений.

XX век:

Результаты поиска в поэтическом корпусе

Объем всего корпуса: 78 868 документов, 1 045 898 предложений, 10 967 173 слова.

Поиск ведётся по пользовательскому подкорпусу объемом 56 500 документов, 669 814 предложений, 6 501 047 слов.

ныне

Найдено 799 документов, 889 вхождений.

Результаты поиска в основном корпусе

Объем всего корпуса: 115 645 документов, 23 803 881 предложение, 283 431 966 слов.

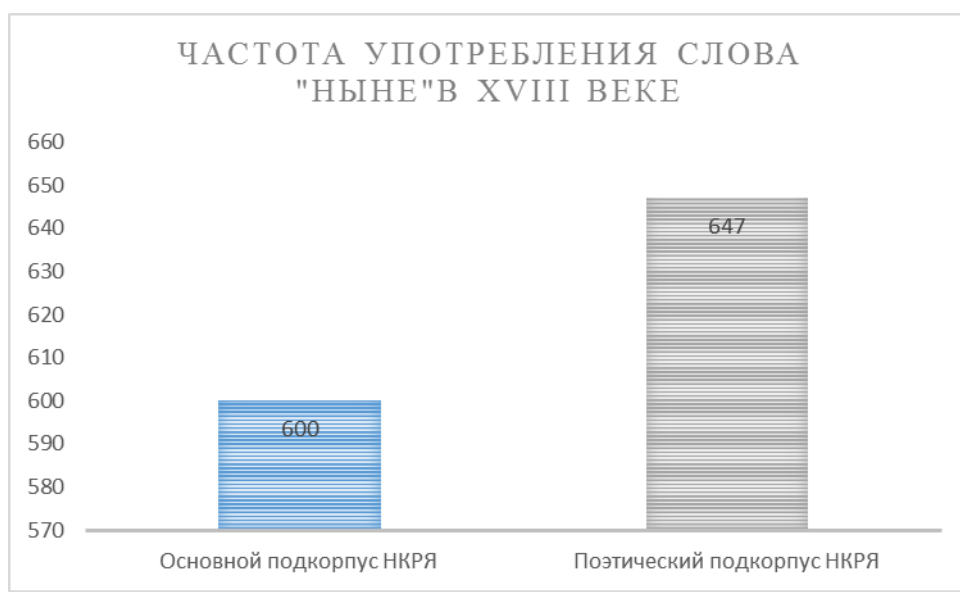
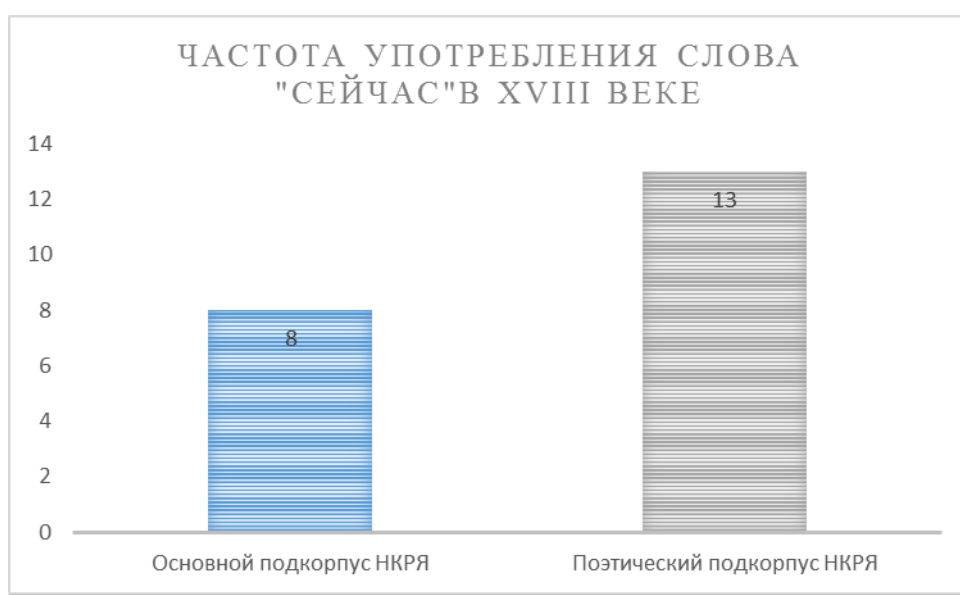
Поиск ведётся по пользовательскому подкорпусу объемом 69 449 документов, 13 820 280 предложений, 154 876 648 слов.

ныне

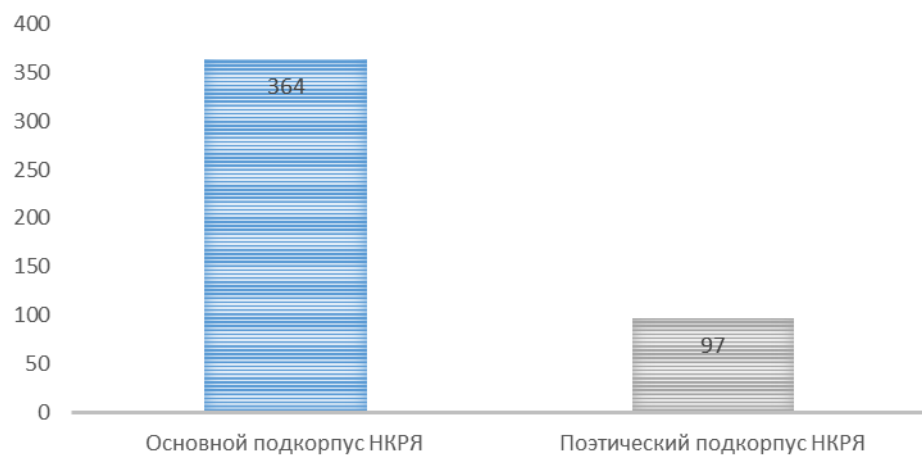
Найдено 4 039 документов, 10 189 вхождений.

[Распределение по годам](#) [Статистика](#)

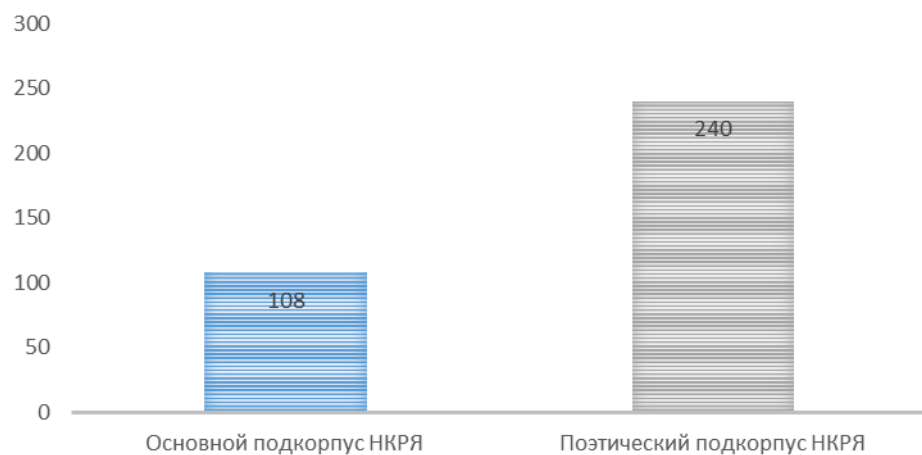
Сравнение частоты употреблений в основном и поэтическом подкорпусах :

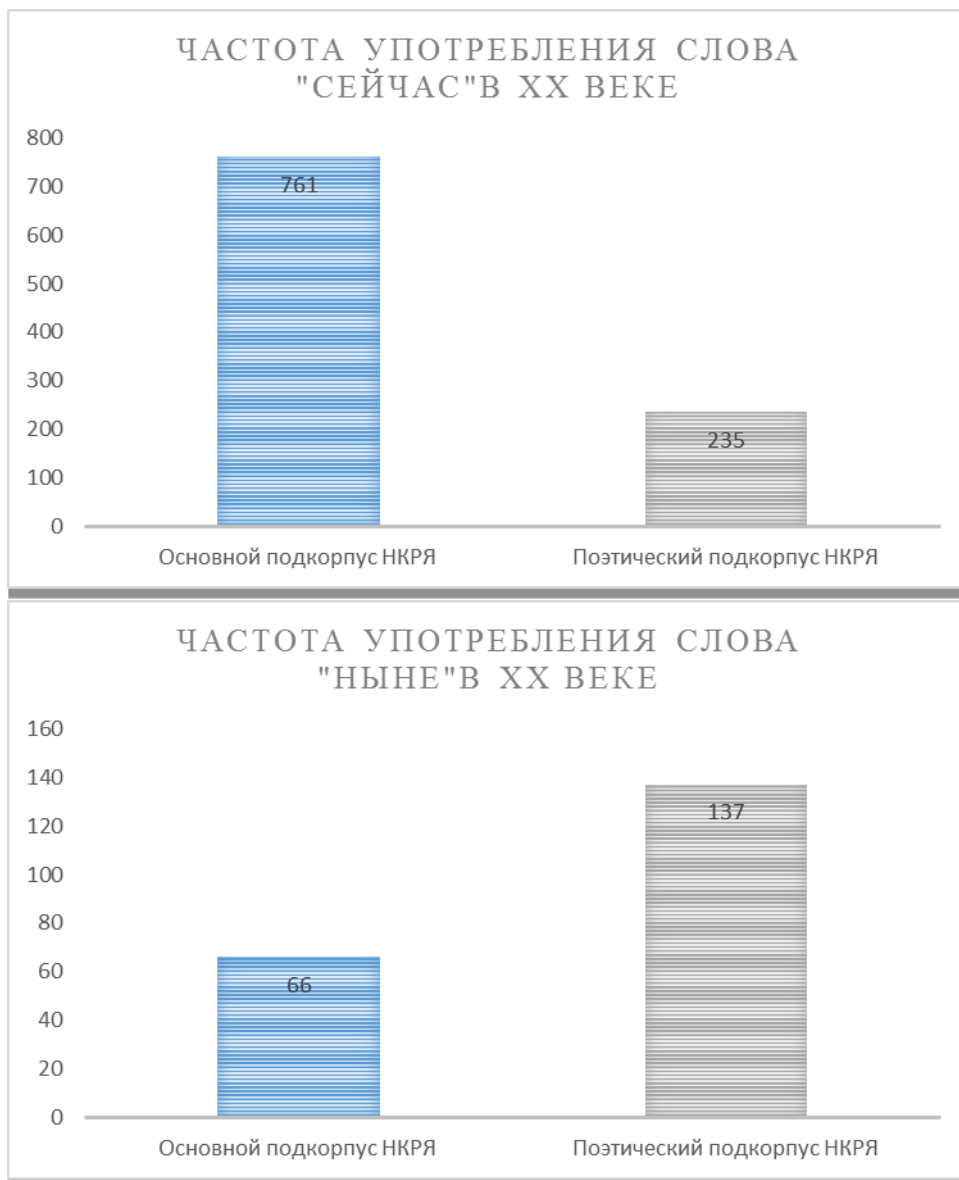


ЧАСТОТА УПОТРЕБЛЕНИЯ СЛОВА "СЕЙЧАС" В XIX ВЕКЕ



ЧАСТОТА УПОТРЕБЛЕНИЯ СЛОВА "НЫНЕ" В XIX ВЕКЕ





Задание номер 2:

Екатерина II являлась российской императрицей в период с 28 июня 1762 года по 6 ноября 1796 года.

Для определения возможно ли существование следующих 3 диалогов в период правления императрицы в основном корпусе НКРЯ был в поиске задан подкорпус по критерию «дата создания текста» и установлен период правления.

Диалоги:

– *Отдай!*
– *Пиши пропало. Что пряжка, третьего дня эта позанка у меня с груди звезды 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!*

На основе анализа употребления слов из диалогов, можно сделать вывод, что только 3 диалог мог существовать в период с 1762 года по 1796 год.

1 диалог:

Фразеологизм «пиши пропало», с которого начинается вторая реплика в диалоге, согласно поиску, в НКРЯ, не могло существовать в указанный период времени.

Результаты поиска в основном корпусе

Поиск ведётся по пользовательскому подкорпусу объемом 1 218 документов, 206 085 предложений, 3 804 608 слов.

пиши
на расстоянии 1 от пропало

По этому запросу ничего не найдено.

2 диалог:

Во втором предложении присутствует слово «лопухом». При поиске данного слова и другой его словоформы «лопух» было выяснено, что данное слово не употреблялось в указанный период. Анализ показал невозможность существования данного диалога в период правления Екатерины II.

Результаты поиска в основном корпусе

Поиск ведётся по пользовательскому подкорпусу объемом 1 218 документов, 206 085 предложений, 3 804 608 слов.

лопух

По этому запросу ничего не найдено.

Результаты поиска в основном корпусе

Поиск ведётся по пользовательскому подкорпусу объемом 1 218 документов, 206 085 предложений, 3 804 608 слов.

лопухом

По этому запросу ничего не найдено.

3 диалог:

В данном диалоге были проверены два, возможно, не специфичных слова для того периода «растолковать» и «напасть». Но поиск в НКРЯ убеждает нас в обратном, а следовательно существование 3 диалога возможно в указанный период.

Результаты поиска в основном корпусе

Объем всего корпуса: 115 645 документов, 23 803 881 предложение, 283 431 966 слов.
Поиск ведётся по пользовательскому подкорпусу объемом 1 218 документов, 206 085 предложений, 3 804 608 слов.
напасть
Найдено 30 документов, 44 вхождения.

Результаты поиска в основном корпусе

Объем всего корпуса: 115 645 документов, 23 803 881 предложение, 283 431 966 слов.
Поиск ведётся по пользовательскому подкорпусу объемом 1 218 документов, 206 085 предложений, 3 804 608 слов.
растолковать
Найдено 6 документов, 12 вхождений.