

Домашнее задание №5.

Часть 1.

1. Частотность в основном корпусе.

Формула:

частота употребления (ipm) = кол-во вхождений/кол-во слов необходимого корпуса*1 000 000

С помощью данной формулы вычисляем частоты слов "ныне" и "сейчас" в разное время (для 18, 19, 20 веков) на основе данных из основного подкорпуса.

18 век:

ныне = $2\,914 / 5\,202\,929 * 1\,000\,000 = 560$

сейчас = $43 / 5\,202\,929 * 1\,000\,000 = 8,2$

19 век:

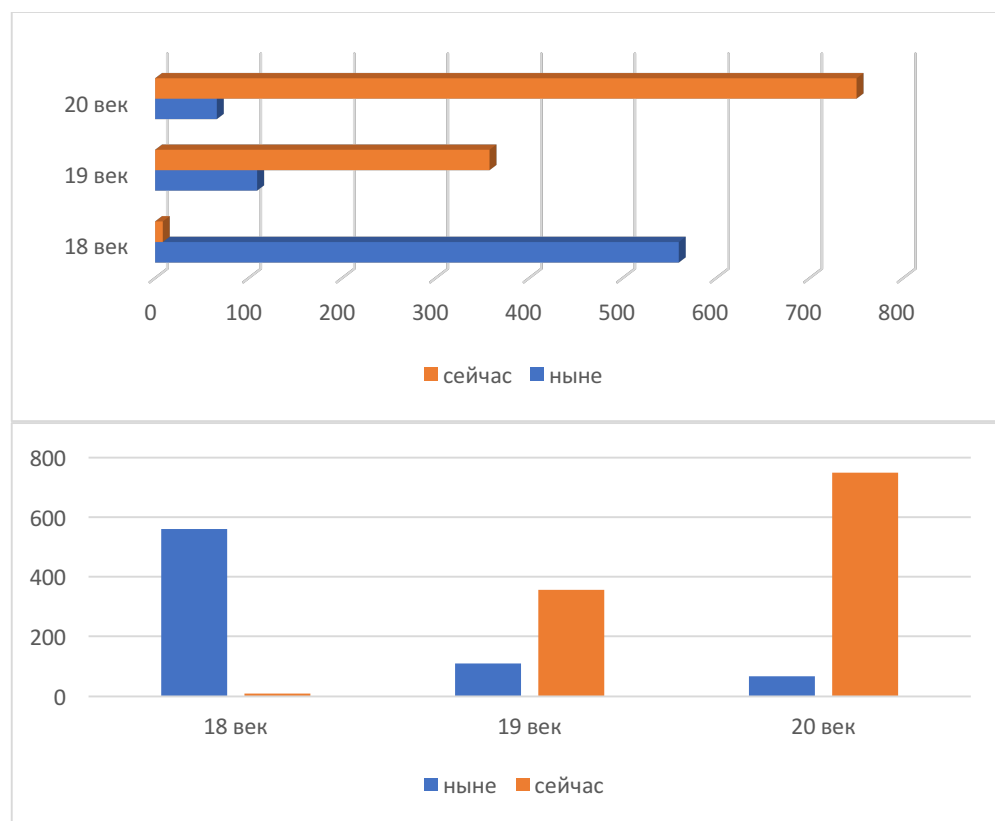
ныне = $5\,922 / 54\,274\,217 * 1\,000\,000 = 109,1$

сейчас = $19\,419 / 54\,274\,217 * 1\,000\,000 = 357,6$

20 век:

ныне = $10\,071 / 152\,456\,773 * 1\,000\,000 = 66$

сейчас = $114\,368 / 152\,456\,773 * 1\,000\,000 = 750,1$



Вывод: уже в 19 веке по сравнению с 18 прослеживается спад употребления слова «ныне», тогда как употребление «сейчас» резко вырастает в десятки раз.

2. Частотность в поэтическом корпусе.

Формула остается та же: частота употребления (ipm) = кол-во вхождений/кол-во слов необходимого корпуса*1 000 000

С помощью формулы вычисляем частоты слов «ныне» и «сейчас» для 18, 19, 20 веков в поэтическом корпусе.

18 век:

$$\text{ныне} = 595/909\,456 * 1\,000\,000 = 555,2$$

$$\text{сейчас} = 12/909\,456 * 1\,000\,000 = 13,1$$

19 век:

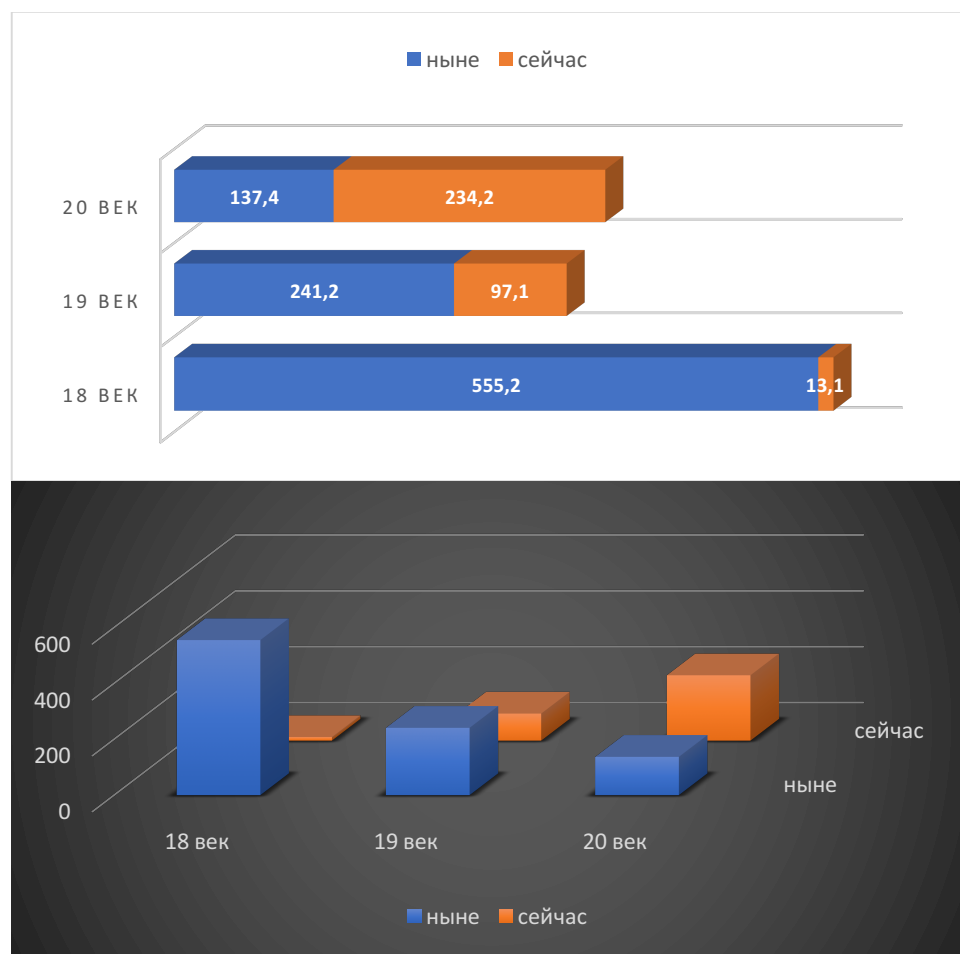
$$\text{ныне} = 852/3\,531\,055 * 1\,000\,000 = 241,2$$

$$\text{сейчас} = 343/3\,531\,055 * 1\,000\,000 = 97,1$$

20 век:

$$\text{ныне} = 898/6\,533\,720 * 1\,000\,000 = 137,4$$

$$\text{сейчас} = 1\,530/6\,533\,720 * 1\,000\,000 = 234,2$$



Вывод: в поэтическом корпусе, так же, как и в основном, явно прослеживается спад употребления слова «ныне», однако в данном случае он не является таким уж резким. Также, разница в частотности употребления слов «сейчас» и «ныне» в 20 меньше, чем в основном корпусе.

Часть 2.

Времена правления Екатерины II (1762-1796 гг.)

Мы предполагаем существование данных фраз и задаем поиск слов, из которых они состоят, и обозначаем временные рамки 1762-1796 гг.

1.

– *Отдай!*

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

«Пиши пропало» – 0 вхождений

«пряжка» – 0 вхождений

«поганка» – 0 вхождений

«сволочь» – 1 вхождение

«Александр Невский» – 0 вхождений

Вывод: вероятность существования подобного диалога маловероятна.

2.

– *Ох, ну просто невысказано! Нет, как можно быть таким лопухом?!*

«лопух» – 0 вхождений

Вывод: ключевое слово данной фразы не нашло вхождений, поэтому, вероятность существования подобного диалога крайне мала.

3.

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

«гнев» – 121 вхождение

«растолковать» – 11 вхождений

«напасть» – 35 вхождений

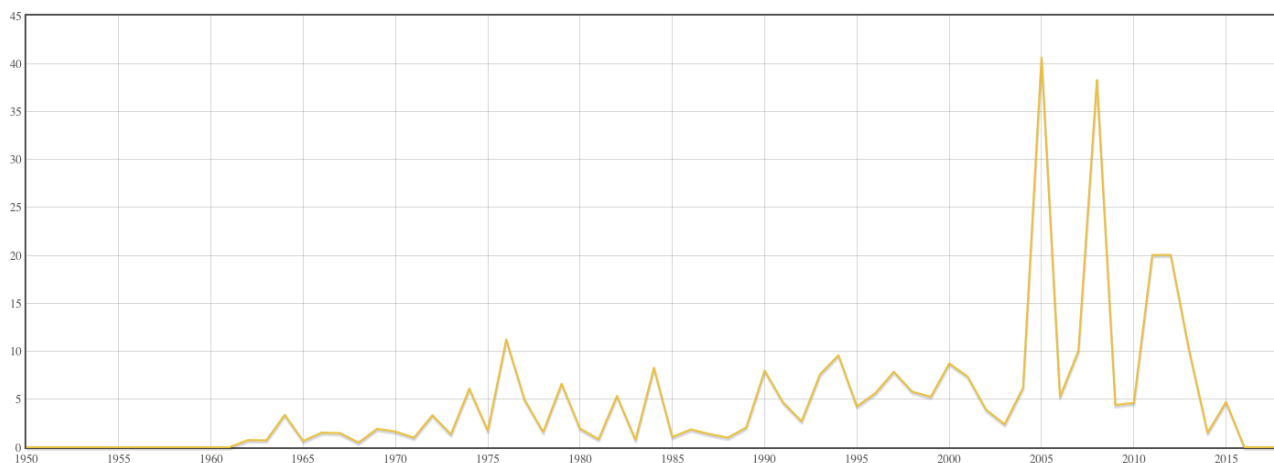
Вывод: данная фраза вполне может принадлежать времени правления Екатерины II.

Часть 3.

Распределение по годам (частота на миллион словоформ)

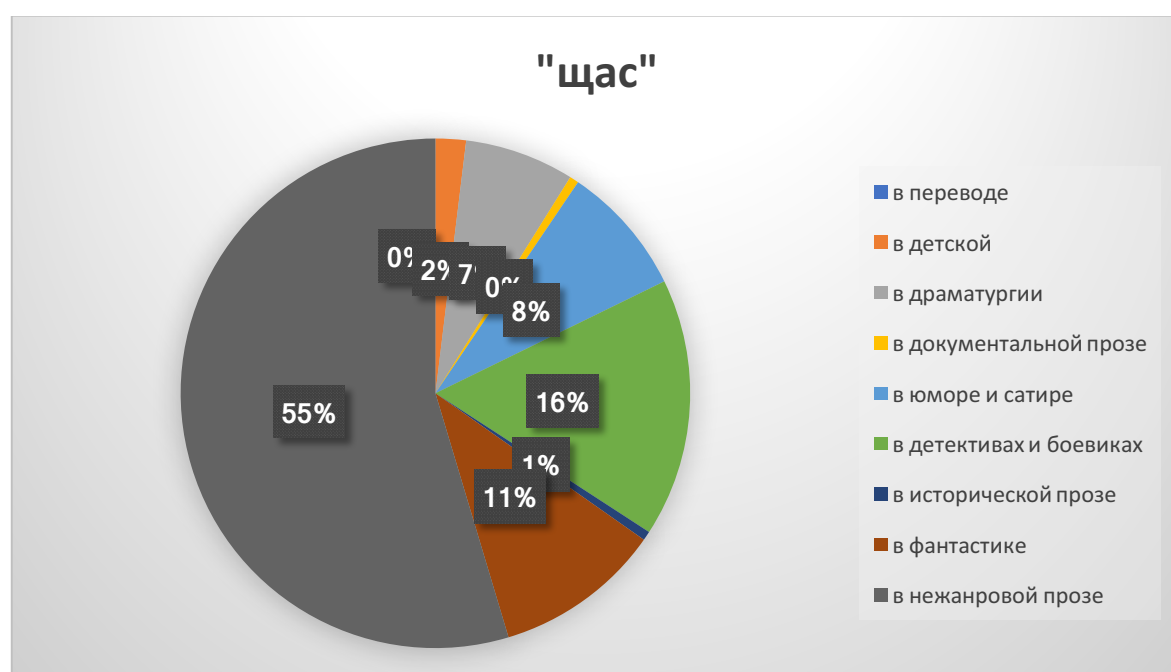
щас

Годы от 1950 до 2018 со сглаживанием 0 построить



Всего вхождений «щас» – 516

в переводе – 0
в детском – 10
в драматургии – 36
в документальной прозе – 3
в юморе и сатире – 43
в детективах и боевиках – 85
в исторической прозе – 3
в фантастике – 55
в нежанровой прозе – 283



Вывод: самое распространенное вхождение слова «щас» оказалось в внежанровой прозе, что, в принципе, неудивительно. На втором месте – детективы или боевики, а на третьем – фантастика, чего тоже следовало ожидать. Удивительно, что в юморе и сатире по сравнению с другими жанрами так мало вхождений, хотя, казалось бы, эти жанры должны быть хотя бы в тройке ведущих. Также, странно, что нет ни одного вхождения в переводе, ведь существует множество адаптированных переводов.