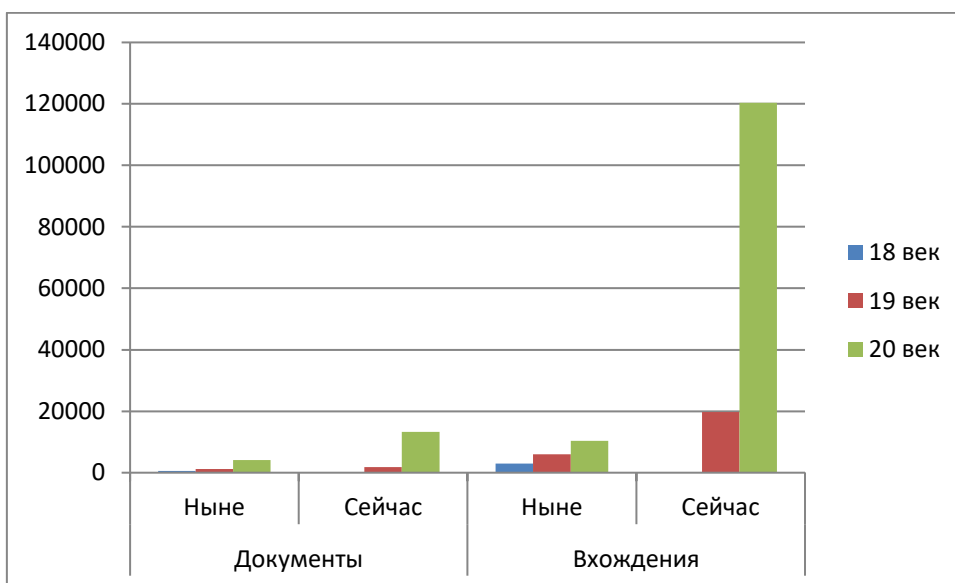


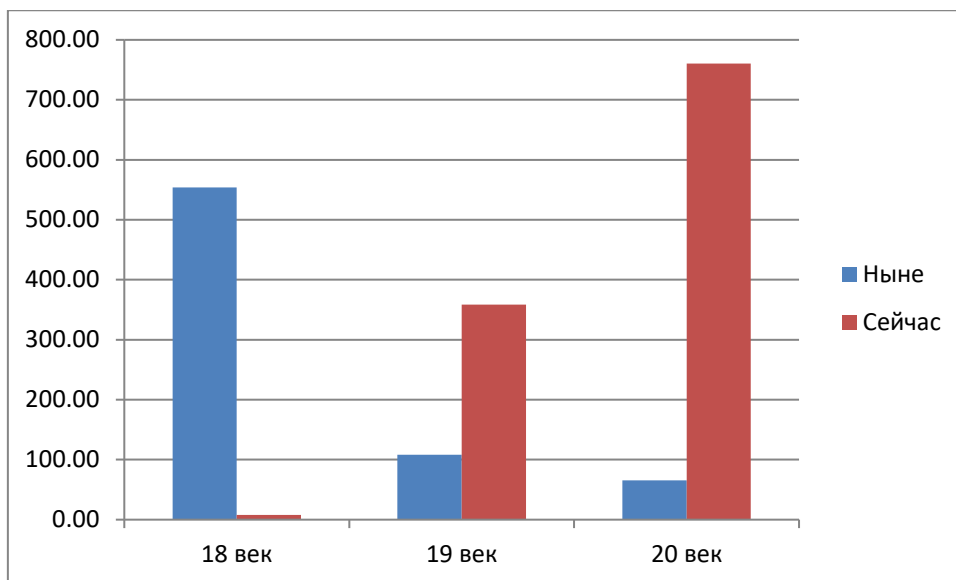
## Домашнее задание №5

### Задание №1

Мы сравнили частоту употреблений (ipm) слов «ныне» и «сейчас» в 18, 19, 20 веках на основе Национального корпуса русского языка. Сначала мы посчитали абсолютную частоту употреблений и количество документов, в которых встречаются эти слова (основной корпус).

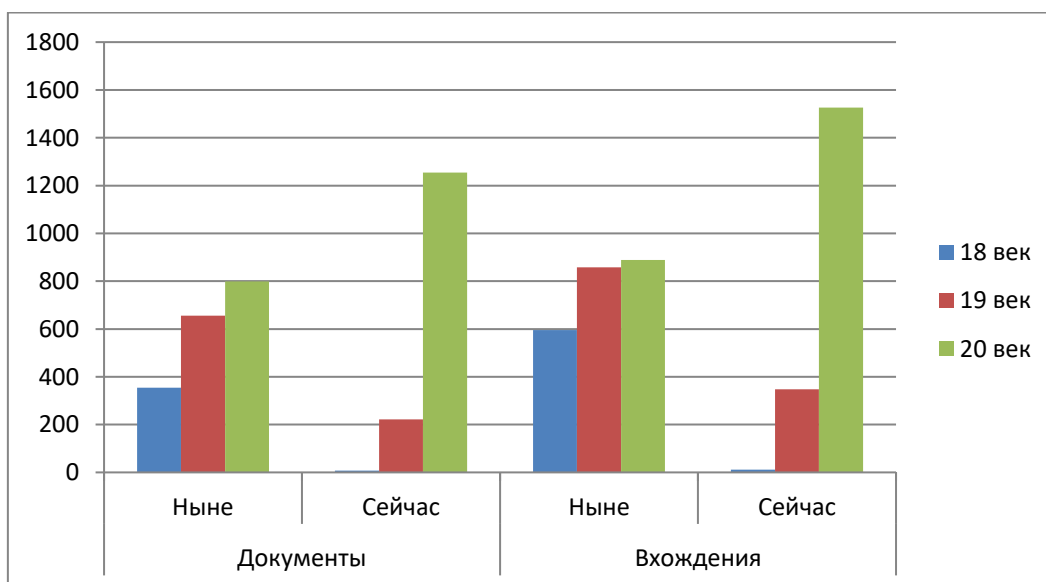


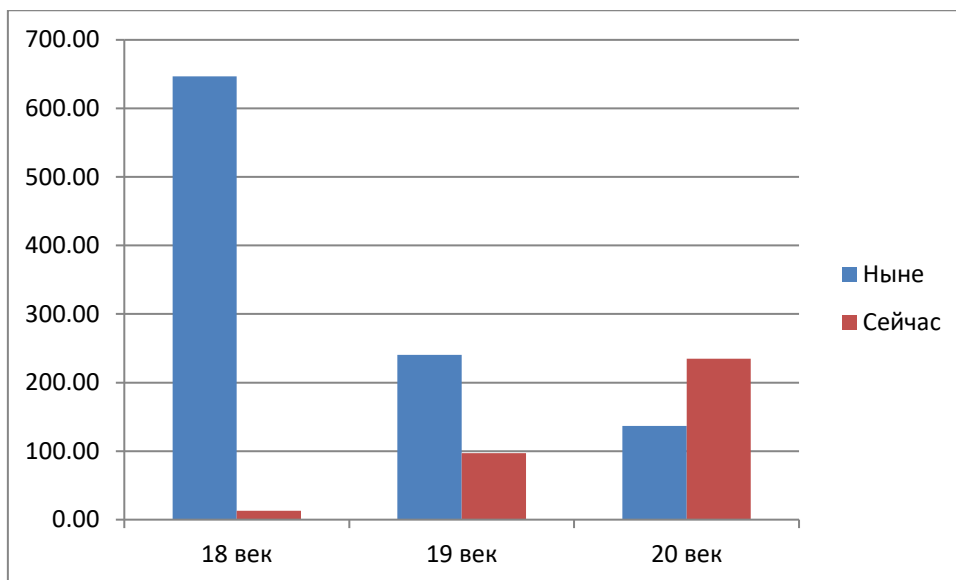
Визуально можно заметить, что слово «сейчас» в 18 веке почти не употреблялось, но в 20 веке его популярность превышала популярность слова «ныне» в несколько раз. Однако, выборки текстов, на которых измеряется абсолютная частотность, довольно сильно отличаются по своим размерам. Поэтому, чтобы получить объективную статистику о частоте употребления слова, необходимо рассчитать относительную частоту употребления. Относительная частота характеризует число употреблений на миллион слов корпуса, или ipm (instances per million words). Мы рассчитали относительную частоту для каждого из подкорпусов (18, 19, 20 век).



Эта диаграмма наглядно демонстрирует, что частота употреблений слова «ныне» на миллион словоформ между 18 и 20 веком упала в несколько раз, а частота употреблений слова «сейчас» выросла в десятки раз. Если говорить точнее, слово «ныне» стало в 8,5 раз менее популярным, а слово «сейчас» стало употребляться в 95 раз чаще.

Мы осуществили те же операции в Поэтическом корпусе.





Можно заметить, что частота употреблений слова «ныне» в Поэтическом корпусе падала существенно медленнее, чем в массиве текстов Основного корпуса, а рост частоты употребления слова «сейчас» был не настолько резким. Из сравнения двух графиков можно сделать вывод, что слово «ныне» фактически вышло из общего употребления и приобрело книжный оттенок, а слово «сейчас» стало общеупотребимым.

## Задание №2

На основании данных Национального корпуса русского языка мы рассмотрели несколько диалогов и попытались определить, могли ли они встретиться в таком виде во времена правления Екатерины II (1762–1796). Мы вели поиск по подкорпусу объёмом 1218 документов, 206 085 предложений, 3 804 608 слов.

— Отдай!

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

Мы можем посчитать частоту употребления отдельных слов и выражений. В некоторых случаях поиск точных форм нас не удовлетворяет, и мы должны искать слово с изменяющейся формой. В этом случае пользуемся лексико-грамматическим поиском.

«Отдай» – 25 документов, 35 вхождений.

«Пиши пропало» – по запросу ничего не найдено.

«Пряжка» – 13 документов, 32 вхождения.

«Третьего дня» – 16 документов, 29 вхождений.

«Поганка» – по запросу ничего не найдено.

«Грудь» – 124 документа, 409 вхождения.

«Звезда» – 85 документов, 264 вхождения.

«Упереть» – 2 документа, 2 вхождения.

«Любить» – 346 документов, 2 387 вхождений.

«Сволочь» – 10 документов, 12 вхождений.

«Блестящий» – 50 документов, 148 вхождений.

Можно сделать вывод, что этот диалог в подобном виде во времена правления Екатерины II встретиться не мог. Слово «поганка» и словосочетание «пиши пропало» совсем не встречаются в документах, а слова «сволочь» и «упереть» употребляются не в таком значении, которое им придаётся в данном диалоге. Остальные слова и выражения достаточно употребимы.

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

«Ох» – 29 документов, 49 вхождений.

«Ну» – 75 документов, 337 вхождений.

«Просто» – 83 документа, 167 вхождений.

«Немыслимо» – по запросу ничего не найдено.

«Лопух» – по запросу ничего не найдено.

Можно сделать вывод, что эта реплика в подобном виде во времена правления Екатерины II встретиться не могла, так как слова «немыслимо» и «лопух» совсем не встречаются в документах. Остальные слова достаточно употребимы.

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

«Гнев» – 154 документа, 415 вхождений.

«Растолковать» – 24 документа, 40 вхождений.

«Дом» – 484 документа, 4 503 вхождения.

«Здесь» – 437 документов, 3 147 вхождений.

«Жить» – 496 документов, 3 611 вхождений.

«Великий» – 693 документа, 6 863 вхождения.

«Напасть» (как существительное) – 73 документа, 131 вхождение.

Все слова в этой реплике общеупотребимы для выборки текстов за 1762–1796 гг., поэтому она в подобном виде могла встретиться во времена правления Екатерины II.

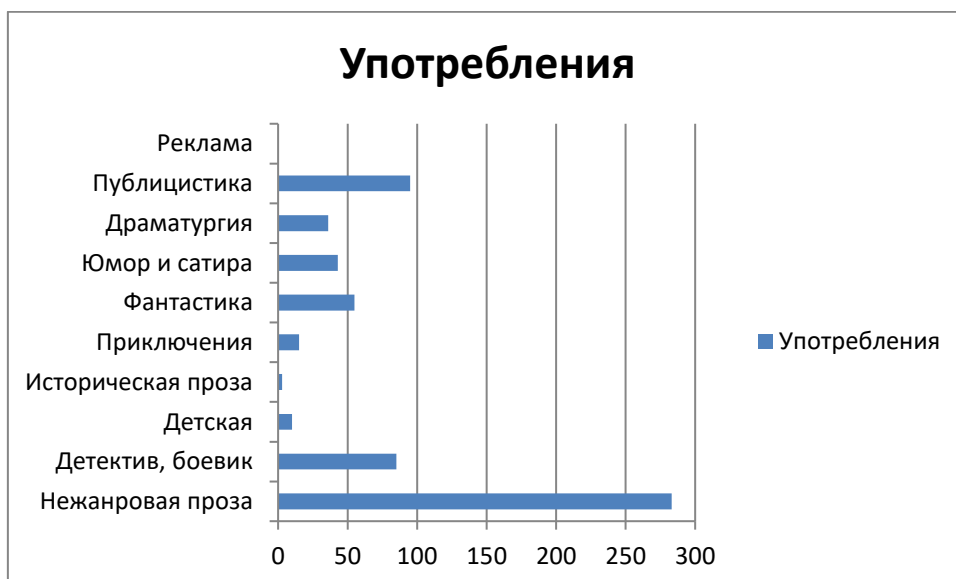
### Задание №3

Мы провели поиск по подкорпусу объёмом 57 807 документов, 13 592 620 предложений, 156 226 189 слов. Было найдено 896 употреблений слова «щас» в 354 документах. Чтобы проследить динамику употребления слова «щас» в 1950–2018 гг., необходимо построить график.

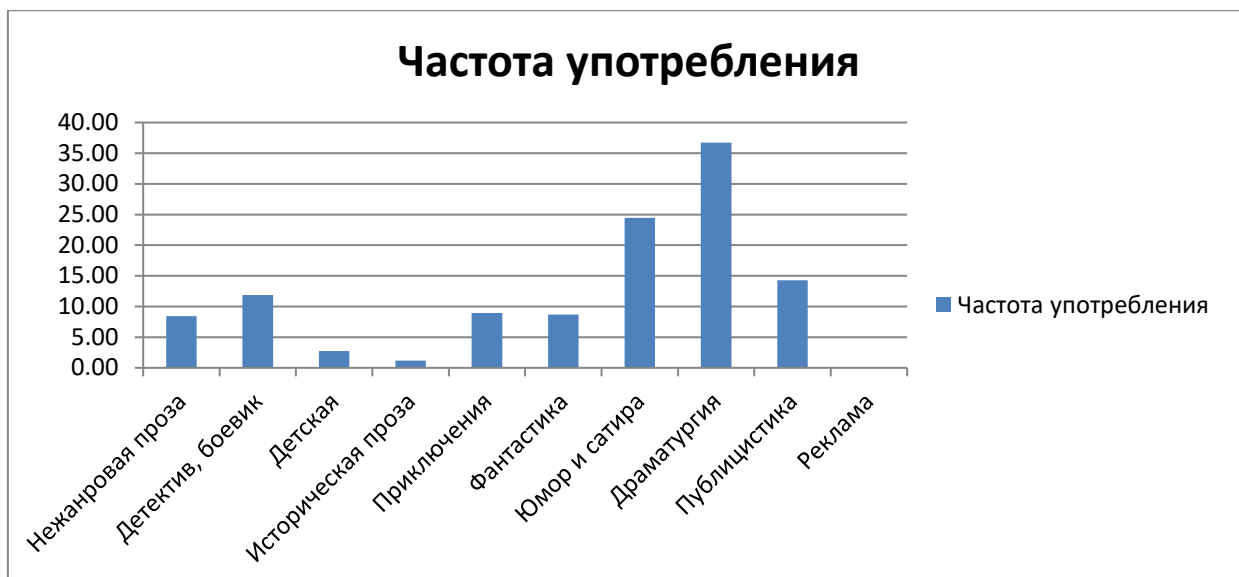


Мы видим, что слово «щас» было особенно популярно в 2000-е гг., а пик употребления пришёлся на 2008 год.

Мы сравнили частоту употребления слова «щас» в 1950–2018 гг. в разных жанрах художественных и нехудожественных текстов. Сначала приведём статистику абсолютной частоты употреблений.



Чтобы получить объективное представление о частоте употребления слова, необходимо рассчитать относительную частоту употребления ( $ipm$ ).



Из графика видим, что слово «щас» чаще всего употребляется в текстах жанров «драматургия» и «сатира», реже всего в детской и исторической прозе. В рекламных текстах слово «щас» не встретилось вовсе.