Отчет по анализу работы с НКРЯ

Луткова Анна 1710, ИЯИМКК

Задание 1

Дано: два слова с похожим семантическим значением "ныне" и "сейчас".

Задание: сравнить частоту употребления в различные века.

Ход работы:

- 1. создание пользовательского подкорпуса с необходимыми датами
- 2. поиск совпадения вхождений по каждому слова в каждый промежуток времени
- 3. вычисление частоты вхождения на миллион слова с помощью формулы:
 - ірт = (кол-во вхождений) / (кол-во слов всего) * 1000000

4. общий анализ данных и сравнение

Пояснение, почему использовалось именно расчет по формуле, а не данные таблицы из "Распределения по годам":

Причина:

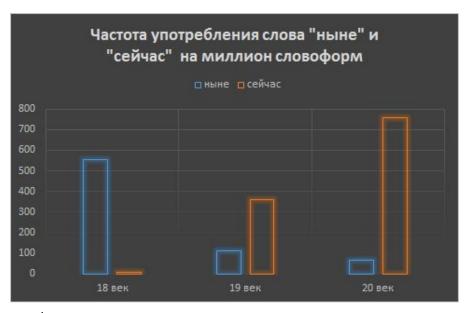
Ipm, рассчитанное в графиках употребления при нулевом сглаживании, не отражали данные таблицы под графиком.

Почему?

Ipm -- это *относительная частота определения*, а в данных таблиц было указано *абсолютная частота употребления*.

Главное -- не запутаться, а значит, нужно упрощать задачи. Формула -- есть упрощение.

Отчет по работе с первым заданием



puc.1

Временные рамки:

- 1700-1800 гг.
- 1800-1900 гг.
- 1900-2000 гг

Употребление слова "ныне":

Век	Количество вхождений	Количество слов в подкорпусе	lpm на миллион слов
18 век	2950	5 325 444	553.94
19 век	6 026	55 271 174	109.03
20 век	10 397	159 090 499	65.35

Употребление слова "сейчас":

Век	Количество вхождений	Количество слов в подкорпусе	lpm на миллион слов
18 век	43	5 325 444	08.07
19 век	19 787	55 271 174	358.00
20 век	120 781	159 090 499	759.2

Анализ данных:

С течением времени, архаичная форма "ныне", обозначающая временное обстоятельство постепенно вытеснялось более современной формой "сейчас". В восемнадцатом и двадцатом веках очерчивается огромный зазор в частоте употребления данных слов, но, по идеи, в промежуточном веке, девятнадцатом, ожидаемо было сравнение частоты употребления. Но, судя по графикам и данным из НКРЯ (см. рис.1), упоминания "сейчас" и его общее вхождение в речи было настолько стремительным, что дало возможность создать преимущество практически в два раза в области употребления над "ныне" (358.00 против 109.03)

Анализ данных в двух словах: в 20-м веке "ныне" ушло в архаику, "сейчас" стало общеупотребимым словом с высоким коэффициентом IPM. В 18 веке было с точностью наоборот.

Поэтический и основный корпуса

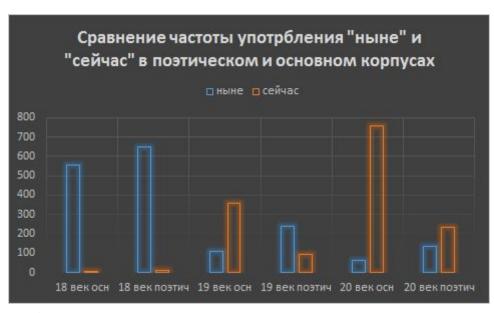
Задание 1.2

Дано: два слова с похожим семантическим значением "ныне" и "сейчас". **Задание**: сравнить частоту употребления в различные века между поэтическим и основным корпусами

Ход работы:

Аналогичен ходу работу с основным корпусом, отличие лишь в том, что создание пользовательского подкорпуса будет проходить в поэтическом.

Отчет по работе



puc.2

Употребление слова "ныне":

Век	Количество вхождений	Количество слов в подкорпусе	lpm на миллион слов
18 век	596	921 424	646.82
19 век	865	3 584 806	241.3
20 век	898	6 542 355	137.26

Употребление слова "сейчас":

Век	Количество вхождений	Количество слов в подкорпусе	lpm на миллион слов
18 век	12	921 424	13.2
19 век	347	3 584 806	96.7
20 век	1530	6 542 355	233.86

Анализ данных:

Как и в основном подкорпусе, частота употребления "сейчас" со временем стремительно росла (из 13.2 в 18 веке выросла до 233.86 в 20 веке), когда в то же время "ныне" испытывала постоянный спад (с 646.82 в 18 веке до 137.26). Но если сравнивать "ныне" хоть и потеряла позиции, но не настолько сильно как в основном корпусе. Более того, зазор между частотой употребления этих слов в поэтическом корпусе в 20-м веке намного ниже, чем в основном (137.26 и 233.86 в поэтическом против 65.35 и 759.2)

Анализ данных в двух словах: "ныне" в поэтическом корпусе тоже проседает с течением времени, как и в основном, но не настолько сильно. "Сейчас" держит уверенный рост во этих двух корпусах без скачком и проседаний.

Задание 2

Дано: реплики с различимыми словами-маркерами

Задание: выяснить, могли ли употребляться такие выражения при Екатерине второй.

Ход работы:

- 1. правление Екатерина Второй —1762-1796
- 2. выделение основных словесных маркеров в каждом диалоге
- 3. создание подкорпуса с выделенными параметрами
- 4. поиск совпадений слов-маркеров в произведениях той эпохи
- 5. анализ полученных данных

Отчеты работы по второму заданию

Цитата 1

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

Слова-маркеры:

Пиши пропало, пряжка, уперла, поганка, сволочь (прим.: "уперла" искать только в значении "украсть")

Совпадения в подкорпусе:

```
пиши пропало — 0 результатов пряжка — 0 результатов поганка — 0 результатов уперла — 0 результатов сволочь — 4 результата
```

Вывод:

Четыре употребления "сволочи", увы, недостаточно. чтобы закрыть глаза на отсутствие всех остальных маркеров, в особенности, наиболее яркого -- словосочетания "пиши пропало". Значит, такого диалога не случилось бы в правление Екатерины Второй. Ну хоть ругаться в то время вполне себе умели.

Цитата 2

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

Слова-маркеры:

немыслимо, лопух

(прим.: возможен поиск по "просто немыслимо", как устойчивое выражение; "лопух" обязателен в поиске по "др. знач." в корпусе)

Совпадения в подкорпусе:

просто немыслимо — 0 результатов немыслимо — 0 результатов лопух — 0 результатов

Вывод:

Могло ли быть такое при Екатерине Второй? - Совершенно точно нет.

Цитата 3

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

Слова-маркеры:

растолковать, напасть

(прим.: возможен поиск по "великая напасть", как по устойчивому выражению, "напасть" искать по "др.знач" с семантическими признаками **"непредметных с отрицательной оценкой"**)

Совпадения в подкорпусе:

растолковать — 5 результатов напасть -- 131 результат великая напасть -- 0 результатов

Вывод:

Несмотря на то, что устойчивое выражение нельзя найти, я уверена, что так могли говорить в то время на основе данных из НКРЯ.