1) характеристики специфичного корпуса:

(1) признаки специфичного корпуса - тексты отзывов

(2) объем корпусов: корпус отзывов: 13559

(3) параметры предобработки: сырые тексты

(4) для вычисления метрик делала лемматизацию

2) характеристики контрастного корпуса:

(1) признаки специфичного корпуса - тексты новостей Ленты.ру

(2) объем корпусов: корпус Ленты: 202567 токенов

(3) параметры предобработки: сырые тексты

(4) для вычисления метрик делала лемматизацию

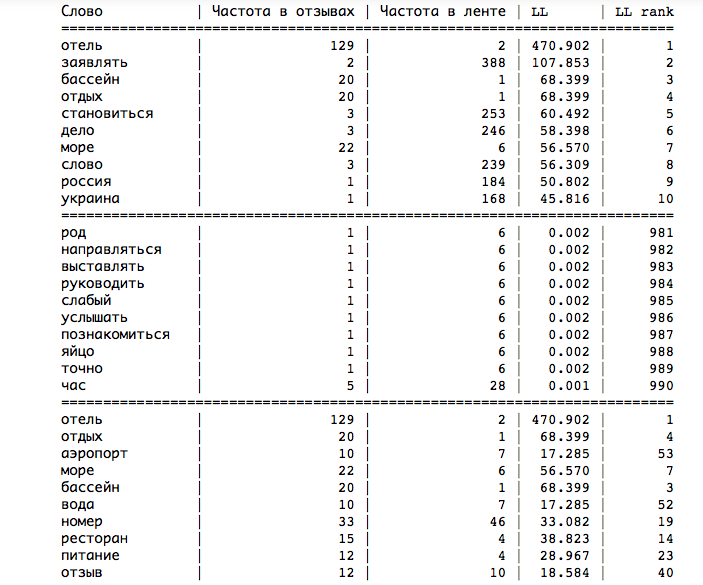
Оба корпуса нам когда-то присылали (но я не вспомнила, когда, поэтому не смогла найти подробностей)

Я выбрала следующие 10 слов: 'отель', 'отдых', 'аэропорт', 'море', 'бассейн', 'вода', 'номер', 'ресторан', 'питание', 'отзыв'

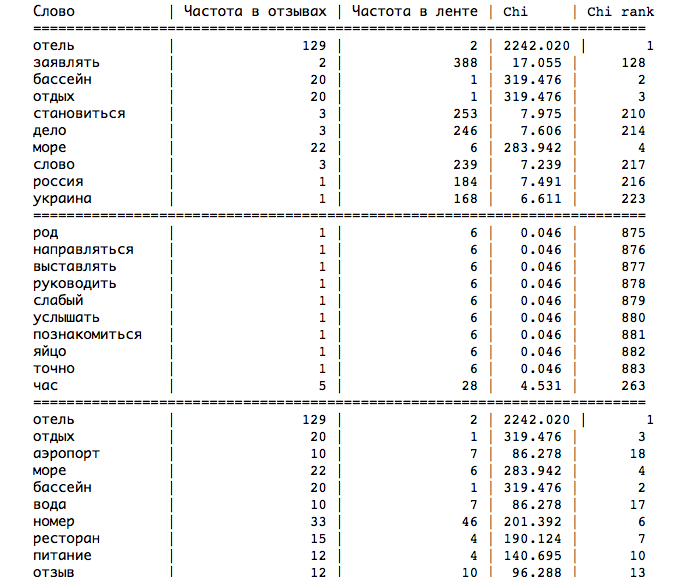
С топ-10 словами совпало 40%, не совпало 60% этих слов.

Ниже таблицы по двум весам (сначала LL, потом Chi-2). В первой части таблиц: топ-10 слов, во второй – 10 последних слов в упорядоченном по LL/Chi-2 списке, в третьей части таблиц – выбранные мною слова.

1. **LogL**



1. **Chi-2**

****

В целом все тематические слова из выбранных мной 10 слов действительно получили высокие ранги в списке (в пределах от 1 до 18).