**Отчет о проекте: Шахова Елизавета Андреевна**

Для исследования выбраны корпуса  и .

Все уровни

**Общая текстометрия**

Частотный список корпуса 1: <https://1drv.ms/x/c/c5e14c0bf253d4bd/EXlgVRQa0KNJv_iqpxm1s-wBBWtTFe_J6nll4gBFKd0UYA?e=Z196bg> ссылка на таблицу (Приложение 1).

Частотный список корпуса 2: <https://1drv.ms/x/c/c5e14c0bf253d4bd/EbBPZMOgV-BAnfbV0dnTiQYB0dGKjXk2TuEG3F0t8PrEJg?e=IEQbtr> (Приложение 2).

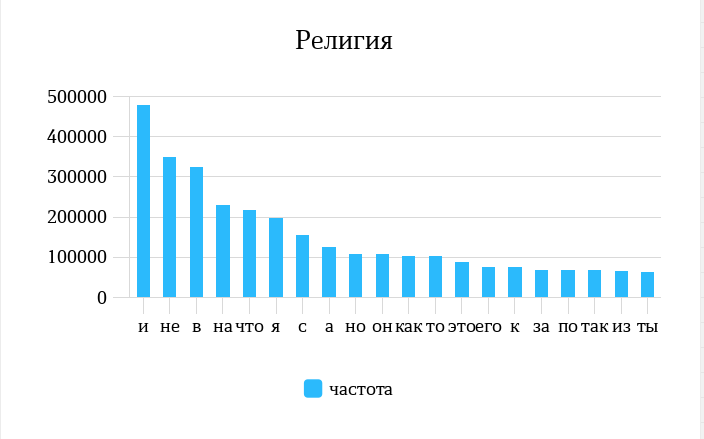
Без стоп-слов 1: <https://1drv.ms/x/c/c5e14c0bf253d4bd/ET46g2pQZuBCpoMJypmacV4BF1sKXeFO9VvOq4VJ6IgkEw?e=ri9402> (Приложение 3).

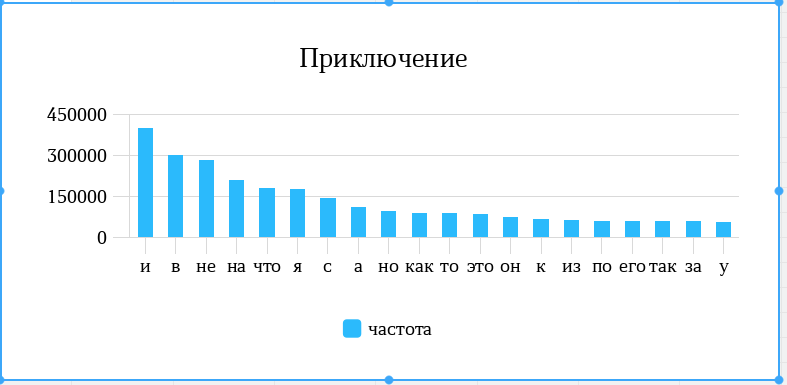
Без стоп-слов 2: <https://1drv.ms/x/c/c5e14c0bf253d4bd/EbuFNkoewAlDvPbojrByvQwBjFtenWzIhGHne0gsu20soQ?e=8QbcN7> (Приложение 4).

Ссылка на таблицу с диаграммой по корпусу 1. <https://1drv.ms/x/c/c5e14c0bf253d4bd/Eci9BHYX3D5CsAcprPFZ6GYB2qFPJ02xGqaYfOlZDAM6gg?e=6VzyM4>

Ссылка на таблицу с диаграммой по корпусу 2. https://1drv.ms/x/c/c5e14c0bf253d4bd/Eci9BHYX3D5CsAcprPFZ6GYB2qFPJ02xGqaYfOlZDAM6gg?e=6VzyM4

Диаграммы (рис. 1, рис. 2):





Вывод: Рассматривая эти корпусы, можно заметить, что списки частотных слов схожи.

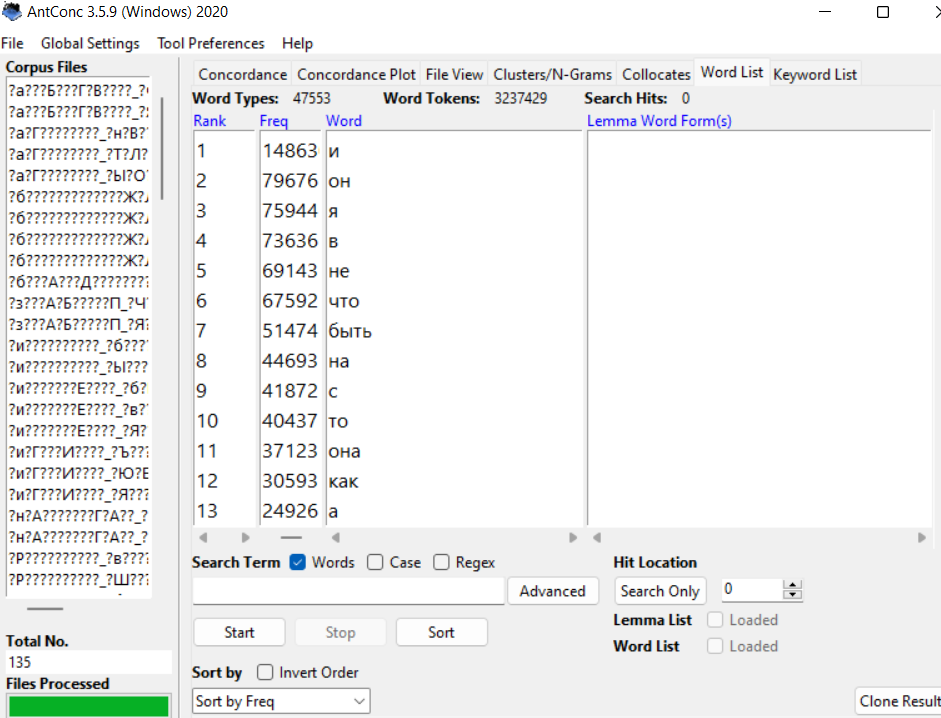
**Лексическое разнообразие**

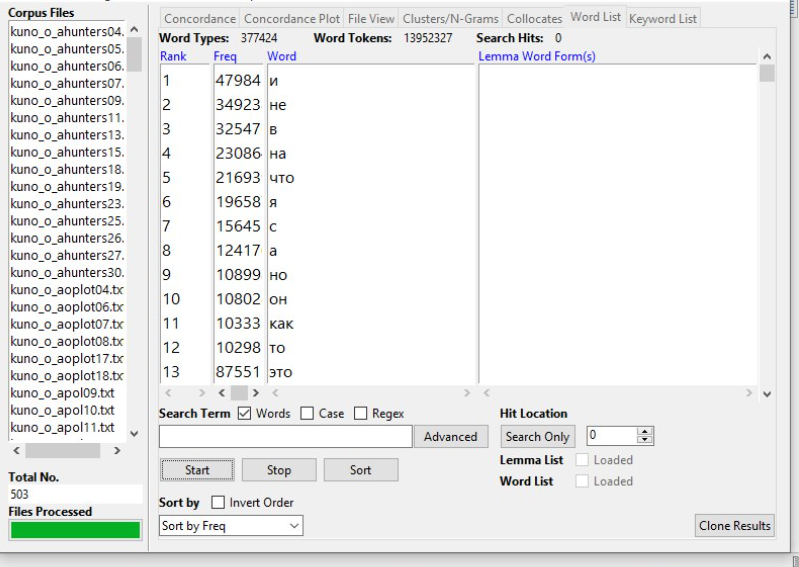
Показатель type/token ratio для корпуса 1:47553/3237429 = 0,015

Показатель type/token ratio для корпуса 2:377424/ 13952327 = 0,027

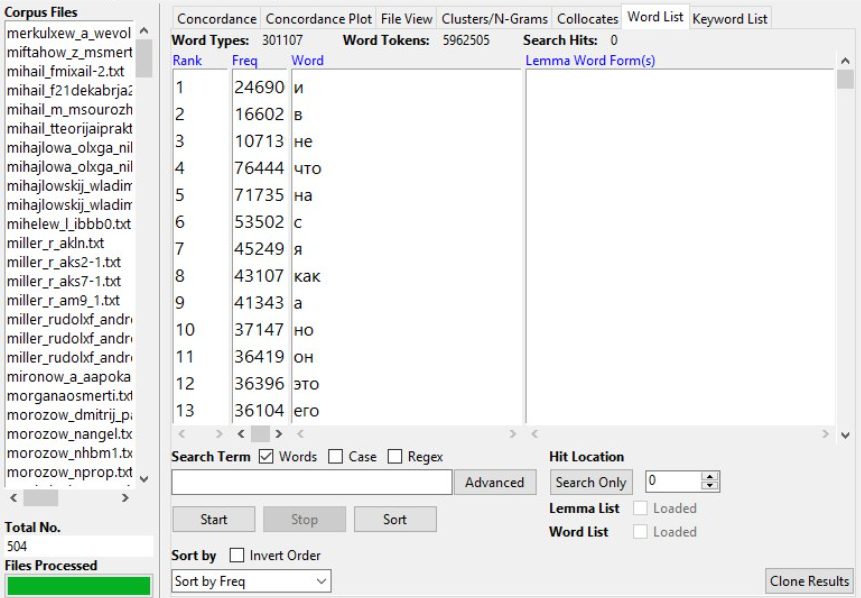
Показатель type/token ratio для референтного корпуса прозы: 47553/3237429 = 0,015

**Референтный корпус**





**Религия**



Вывод: сравнивания показатели этих корпусов можно заметить, что в корпусе «Религия» больше уникальных слов, а показатели рефернтного корпуса и корпуса «Приключения» одинаковы.

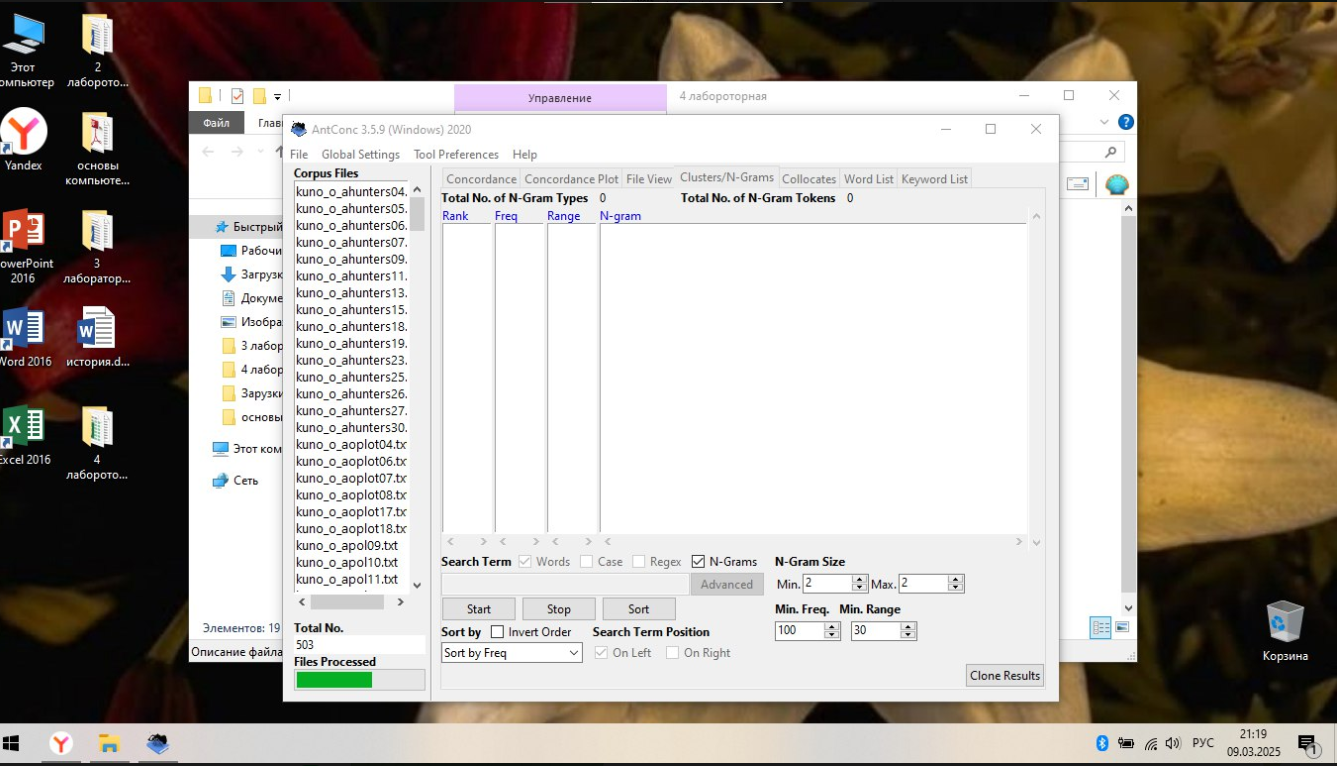
**Триграммы**

Ссылка на таблицу триграмм – корпус 1 (Приложение 7).

Ссылка на таблицу триграмм – корпус 2 (Приложение 8).

Самая частотная триграмма – корпус 1: <https://1drv.ms/x/c/c5e14c0bf253d4bd/EYd5f0DWyCFMo5N4tPvGymoB4dxCBMfNNmvKZvOsOITswQ?e=vpcTzR>

Самая частотная триграмма – корпус 2: и не (двуграмма)



Приложение не смогло посчитать триграммы для корпусов, удалось найти только двуграмму для корпуса «Религия», она прикреплена выше.