

Вторая часть задания № 6

Сравниваются тексты:

«Анна Каренина» Толстой Л.Н. и «Узник золотого лабиринта» Артамонова Е.В

а) Отсортируйте слова по частоте (Word List - Sort by Freq). Какие слова оказались в 5-ке наиболее частотных? Не только назовите конкретные слова, а подумайте, почему они там оказались.

The screenshot shows the 'Word List' window for the file 'anna_karenina.txt'. The window has tabs for 'Concordance', 'Concordance Plot', 'File View', 'Clusters/N-Grams', 'Collocates', 'Word List', and 'Keywords'. The 'Word List' tab is active, displaying a table of word frequencies. The table has columns for 'Rank', 'Freq', and 'Word'. The top 13 words are listed, with 'и' being the most frequent (12915 occurrences).

Rank	Freq	Word
1	12915	и
2	6538	не
3	6127	что
4	5723	в
5	5552	он
6	3596	на
7	3436	она
8	3358	с
9	3217	я
10	2688	как
11	2581	но
12	2578	его
13	2233	это

Below the table, there are search options: 'Search Term' (set to 'порядочные'), 'Words' (checked), 'Case' (unchecked), and 'Regex' (unchecked). There are also buttons for 'Start', 'Stop', and 'Sort'. On the right, there are options for 'Hit Location' (set to 0), 'Lemma List' (unchecked), and 'Word List' (unchecked).

The screenshot shows the 'Word List' window for the file 'uznik-zolotogo-labirin'. The window has tabs for 'Concordance', 'Concordance Plot', 'File View', 'Clusters/N-Grams', 'Collocates', 'Word List', and 'Keywords'. The 'Word List' tab is active, displaying a table of word frequencies. The table has columns for 'Rank', 'Freq', and 'Word'. The top 10 words are listed, with 'и' being the most frequent (942 occurrences).

Rank	Freq	Word
1	942	и
2	913	в
3	816	не
4	626	на
5	460	что
6	409	он
7	382	а
8	382	с
9	310	но
10	255	его

В обоих случаях самые популярные варианты почти идентичны и открывает их «и». Большинство из них частица, предлоги, личные местоимения, союзы. Это самые распространенные служебные части речи, они связывают предложения и без них невозможно обойтись в целом.

б) Какие слова оказались самыми редкими? Как быстро получить эти результаты? (Выпишите, какие настройки вы использовали)

le Global Settings Tool Preferences Help

Corpus Files

anna_karenina.txt

Concordance	Concordance Plot	File View	Clusters/N-Grams	Collocates	Word List	Keyword List
Word Types: 33299		Word Tokens: 271360		Search Hits: 0		
Rank	Freq	Word	Lemma			
1	1	aber				
2	1	ablutions				
3	1	admirable				
4	1	adore				
5	1	age				
6	1	aise				
7	1	allemand				
8	1	aller				
9	1	allons				
10	1	alphonse				
11	1	amour				
12	1	an				

Corpus Files		Concordance		Concordance Plot	File View	Clusters/N-Grams	Collocates	V
uznik-zolotogo-labirin		Word Types: 9684		Word Tokens: 35893		Search Hits: 0		
Rank		Freq		Word			Lemma Word Forr	
1		1		books				
2		1		html				
3		1		http				
4		1		labirinta				
5		1		uznik				
6		1		www				
7		1		zolotogo				
8		1		абсолютно				
9		1		августе				
10		1		авоськой				
11		1		автоинспектора				
12		1		автономной				
13		1		автор				

Я использовала клавишу invert order. Сложно сказать, какие слова стали самыми редкими, потому что самыми редкими во втором случае слова из оглавления, а в первом просто иностранные, Толстой использовал много иностранных слов, не часто их повторял.

в) выберите 3 слова и проанализируйте контексты, в которых они используются (Concordance)

The screenshot shows a software interface for analyzing text corpora. On the left, a list of 'Corpus Files' includes 'anna_karenina.txt'. The main window displays 'Concordance Hits' for the search term 'плакал'. The results are shown in a table with columns for 'Hit' (line number), 'KWIC' (the context with the search term highlighted), and 'File' (the source file). Five hits are listed, all from 'anna_karenina.txt'. At the bottom, there is a search control panel with a 'Search Term' field containing 'плакал', checkboxes for 'Words', 'Case', and 'Regex', and buttons for 'Start', 'Stop', 'Sort', and 'Advanced'. A 'Show Every Nth Row' option is set to 1, and a 'Search Window Size' is set to 50.

Hit	KWIC	File
1	онский и, закрыв лицо руками, плакал. Он вскочил на голос д	anna_karen
2	я, когда он приезжал ко мне и плакал, говоря о тебе, и какая	anna_karen
3	много общую радость. Гриша плакал, говоря, что и Николень	anna_karen
4	а так же слабо, по-детски, как плакал он. В это время дверь с	anna_karen
5	ласки. Алексей Александрович плакал, целуя ее руки, и говори	anna_karen

Search Term: ☒ Words ☐ Case ☐ Regex

Show Every Nth Row:

Search Window Size:

Total No. 1

Corpus Files
anna_karenina.txt

Concordance Concordance Plot File View Clusters/N-Grams Collocates Word List Keyword List

Concordance Hits 32

Hit	KWIC	File
1	мче и хлопотливее щебетали в чаще. Недалеко заухал филин,	anna_karenin
2	а, а тем выше ценил себя, чем чаще были примеры жен, изме	anna_karenin
3	лось, взглядом. Мальчик этот чаще всех других был помехой	anna_karenin
4	е они подходили к водам, тем чаще встречались большие, и в	anna_karenin
5	й рясе и шитом поясе, чаще и чаще выходил к боковым двер	anna_karenin
6	авал своего брата, тем чаще и чаще ему приходило в голову,	anna_karenin
7	рили вокруг. Но он все чаще и чаще и с бóльшим упорством	anna_karenin
8	лиловой рясе и шитом поясе, чаще и чаще выходил к боковы	anna_karenin
9	иже узнавал своего брата, тем чаще и чаще ему приходило в	anna_karenin
10	и заговорили вокруг. Но он все чаще и чаще и с бóльшим упо	anna_karenin
11	лушиваясь к пролетавшим все чаще и чаще мимо него пчела	anna_karenin
12	в Москве без дела, Левину все чаще и чаще, настоятельнее и	anna_karenin
13	. Чем долее Левин косил, тем чаще и чаще он чувствовал ми	anna_karenin
14	ому запозданию. Потом стали чаще и чаще поглядывать на д	anna_karenin

Search Term ☒ Words ☐ Case ☐ Regex Search Window Size 50

чаще

Start Stop Sort Show Every Nth Row 1

Total No.
1

File Global Settings Tool Preferences Help

Corpus Files
anna_karenina.txt

Concordance Concordance Plot File View Clusters/N-Grams Collocates Word List Keyword List

Concordance Hits 7

Hit	KWIC	File
1	, даже Лили нашла березовый гриб. Прежде бывало так, что	anna_karenin
2	то присевшая, чтобы поднять гриб, гибким движением подн	anna_karenin
3	епокорный, отскальзывающий гриб и встряхивая кружевami,	anna_karenin
4	сь косами. Но старик, встречая гриб, каждый раз сгибался, под	anna_karenin
5	и, стало утихать. – Березовый гриб – корень его напоминает	anna_karenin
6	нувшимися краями березовый гриб, он взглянул ей в глаза и, :	anna_karenin
7	на колени и защищая руками гриб от Гриши, она звала мале	anna_karenin

Search Term ☒ Words ☐ Case ☐ Regex Search Window Size 50

гриб

Start Stop Sort Show Every Nth Row 1

Total No.
1

Слово «плакал» используется в трех случаях не как отдельное действие, а как сопутствующее. Слово «чаще» наиболее употребляемо в словосочетании «чаще и чаще» то есть «больше и

больше», сегодня так говорят редко. В третьем случае речь идет в основном о березовом грибе и манипуляции с ними упоминаются и по отношению к женщинам, и мужчинам, и пожилым.

The screenshot shows a software interface for searching through a corpus. On the left, a list of corpus files includes 'uznik-zolotogo-labirin'. The main window displays 'Concordance Hits' for the search term 'крикнул'. There are 7 hits listed, each with a line number, a snippet of text (KWIC), and the source file name 'uznik-zoloto'.

Hit	KWIC	File
1	бавил ходу. - Давай, жми! – крикнул Денис, подбадривая н	uznik-zoloto
2	дет, - ответил Славка и громко крикнул: - Касьян! Выходи вст	uznik-zoloto
3	айся же, наконец! – с досадой крикнул Славка и с разберу то	uznik-zoloto
4	- Эй, Касьян, брось оружие! – крикнул ему вышедший из ма	uznik-zoloto
5	ьцов и ждите подкрепления, - крикнул на ходу отец Дениса, i	uznik-zoloto
6	ты здесь?! – на всякий случай крикнул он, проходя мимо неп	uznik-zoloto
7	Открывай, а то пожалеешь! – крикнул он, размахивая совер	uznik-zoloto

At the bottom, the search controls show the search term 'крикнул' with options for 'Words', 'Case', and 'Regex'. There are buttons for 'Start', 'Stop', 'Sort', and 'Advanced'. The 'Show Every Nth Row' is set to 1, and the 'Search Window Size' is set to 50.

Concordance Concordance Plot File View Clusters/N-Grams Collocates Word List Keyword List

Corpus Files
uznik-zolotogo-labirin

Concordance Hits 7

Hit	KWIC	File
1	обрыва. - Вряд ли. Смотри, чуть дальше – еще одна . Это –	uznik-zolot
2	ьно посмотрел на маму, и она чуть кивнула головой , давая ра	uznik-zolot
3	ну для замка, расположенную чуть ниже первой . И вновь Де	uznik-zolot
4	иснимым – он всего лишь чуть- чуть отклонился в сторону при	uznik-zolot
5	ечал, что остальные придут чуть позже, и, вообще , особо в	uznik-zolot
6	оехал на велосипеде мальчик, чуть притормозил у Славкиной	uznik-zolot
7	объяснимым – он всего лишь чуть-чуть отклонился в сторон	uznik-zolot

Search Term ☒ Words ☐ Case ☐ Regex Search Window Size 50

чуть Advanced

Start Stop Sort Show Every Nth Row 1

Total No.

Concordance Concordance Plot File View Clusters/N-Grams Collocates Word List Keyword List

Corpus Files
uznik-zolotogo-labirin

Concordance Hits 7

Hit	KWIC	File
1	думал, что ты пропал. Как твоя нога ? - Болит, но, это пустяк	uznik-zolot
2	его снова сильно разболелась нога . Оставаться на размытом	uznik-zolot
3	и. У Дениса все больше болела нога, а волосы на макушке дав	uznik-zolot
4	просто тряся от страха, да и нога болела все сильнее , делая	uznik-zolot
5	нувшись в щель? - Еще бы! Нога до сих пор болит. Кстати,	uznik-zolot
6	Я не могу спуститься. Больная нога не позволяет . Наверное , с	uznik-zolot
7	е чувствую и тебе не советую. Нога-то еще болит? - Болит	uznik-zolot

Search Term ☒ Words ☐ Case ☐ Regex Search Window Size 50

нога Advanced

Start Stop Sort Show Every Nth Row 1

Total No.

Слово «крикнул» употребляется в достаточно агрессивном контексте при упоминании оружия, угрозы и тд. И имена, тех, кто кричит или кому кричат достаточно простые и мальчишеские. Слово «чуть» используется в контексте «немного». Слово «нога» и все его упоминания связаны с болью в данном тексте.

г) постройте для них Concordance Plot для каждого слова

The screenshot shows the Concordance Plot software interface. On the left, under 'Corpus Files', the file 'anna_karenina.txt' is listed. The main window has several tabs: 'Concordance', 'Concordance Plot' (which is active), 'File View', 'Clusters/N-Grams', 'Collocates', 'Word List', and 'Keyword List'. Below the tabs, it shows 'Concordance Hits 5' and 'Total Plots (with hits) 1'. A plot area is visible with a grid of 5 columns and 1 row, labeled 'Plot: 1 FILE: anna_karenina.txt'. To the right of the plot, it says 'Hits: 5' and 'Chars: 1710408'. At the bottom, there is a search bar with the text 'плакал', checkboxes for 'Words' (checked), 'Case', and 'Regex', and a 'Plot Zoom' dropdown set to 'x1'. There are also buttons for 'Start', 'Stop', and 'Advanced', and a 'Show Every Nth Row' dropdown set to '1'.

Corpus Files

anna_karenina.txt

Concordance Hits 32

Total Plots (with hits) 1

Plot: 1 FILE: anna_karenina.txt

2	3	4	5	7	8	10	15	17	180	28	26	28	2932

Hits: 32
Chars: 1710408

Total No.
1

Files Processed

Search Term ☒ Words ☐ Case ☐ Regex

чаще

Advanced

Plot Zoom
x1

Start

Stop

Show Every Nth Row 1

Corpus Files

anna_karenina.txt

Concordance

Concordance Plot

File View

Clusters/N-Grams

Collocates

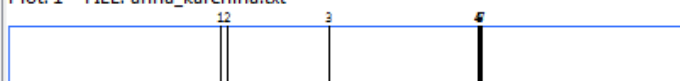
Word List

Keyword List

Concordance Hits 7

Total Plots (with hits) 1

Plot: 1 FILE: anna_karenina.txt



Hits: 7

Chars: 1710408

Total No.

1

Files Processed

Search Term

☒ Words☐ Case☐ Regex

Plot Zoom

гриб

Advanced

x1

Corpus Files

uznik-zolotogo-labirin

Concordance

Concordance Plot

File View

Clusters/N-Grams

Collocates

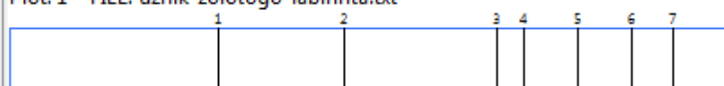
Word List

Keyword List

Concordance Hits 7

Total Plots (with hits) 1

Plot: 1 FILE: uznik-zolotogo-labirinta.txt



Hits: 7

Chars: 250560

Total No.

1

Files Processed

Search Term

☒ Words☐ Case☐ Regex

Plot Zoom

крикнул

Advanced

x1

Start

Stop

Show Every Nth Row 1


File Global Settings Tool Preferences Help

Corpus Files
uznik-zolotogo-labirin

Concordance Concordance Plot File View Clusters/N-Grams Collocates Word List Keyword List

Concordance Hits 7 **Total Plots (with hits) 1**

Plot: 1 FILE: uznik-zolotogo-labirinta.txt



Hits: 7
Chars: 250560

Total No.
1

Files Processed

Search Term ☒ Words ☐ Case ☐ Regex

чуть

Advanced

Plot Zoom
x1

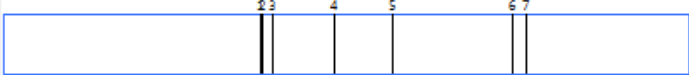
File Global Settings Tool Preferences Help

Corpus Files
uznik-zolotogo-labirin

Concordance Concordance Plot File View Clusters/N-Grams Collocates Word List Keyword List

Concordance Hits 7 **Total Plots (with hits) 1**

Plot: 1 FILE: uznik-zolotogo-labirinta.txt



Hits: 7
Chars: 250560

Total No.
1

Files Processed

Search Term ☒ Words ☐ Case ☐ Regex

нога

Advanced

Plot Zoom
x1

д) выберите по одному слову из каждого текста и составьте для них список биграмм - N-грамм (Clusters/N-gram → min (2) - max (5)?)

Corpus Files
anna_karenina.txt

Concordance | Concordance Plot | File View | Clusters/N-Grams | Collocates | Word List | Keyword List

Total No. of Cluster Types 8 **Total No. of Cluster Tokens 8**

Rank	Freq	Range	Cluster
1	1	1	порядочные женщины
2	1	1	порядочные женщины, по
3	1	1	порядочные женщины, по понятиям
4	1	1	порядочные женщины, по понятиям графини
5	1	1	порядочные люди
6	1	1	порядочные люди будут
7	1	1	порядочные люди будут удаляться
8	1	1	порядочные люди будут удаляться, разумеется

Search Term ☒ Words ☐ Case ☐ Regex ☐ N-Grams **Cluster Size**
порядочные Min. 2 Max. 5
 Min. Freq. Min. Range

File Global Settings Tool Preferences Help

Corpus Files

uznik-zolotogo-labirin

Concordance Concordance Plot File View Clusters/N-Grams Collocates Word List Keyword Lis

Total No. of Cluster Types 12 Total No. of Cluster Tokens 12

Rank	Freq	Range	Cluster
1	1	1	маме помогаешь
2	1	1	маме помогаешь? – отвлек
3	1	1	маме помогаешь? – отвлек его
4	1	1	маме помогаешь? – отвлек его голос
5	1	1	маме помочь
6	1	1	маме помочь убрать
7	1	1	маме помочь убрать приусадебный
8	1	1	маме помочь убрать приусадебный участок
9	1	1	маме скажем
10	1	1	маме скажем? - правду
11	1	1	маме скажем? - правду, - грустно
12	1	1	маме скажем? - правду, - грустно вздохнул

Search Term ☒ Words ☐ Case ☐ Regex ☐ N-Grams Cluster Size Min. 2 Max. 5

NAME Advanced

Start Stop Sort

Sort by ☐ Invert Order Search Term Position Min. Freq. 1 Min. Range 1

Total No. 1

е) сделайте вывод об индивидуальных особенностях авторских стилей

В некоторых местах видно, что стиль Толстого более старый, даже, если посмотреть на то, что слово «порядочный» относится к женщине, а со словом люди рядом соседствует слово «даляться», которые сейчас реже используется. В случае со вторым текстом видно, что он написан более простым, современным языком. В нем используются сокращения имен вроде «Славка», а маму не называют «матерью». Также Толстой использует иностранные слова, так как это было модно в его время. Если углубиться в анализ, то можно найти значительную разницу в употреблении и наличии многих слов.