

Представление и анализ библейских,
святоотеческих и литургических
цитат в корпусе СКАТ:
ВОЗМОЖНОСТИ ТХМ

Alexei Lavrentiev

UMR IHRIM – ENS de Lyon / CNRS

Corpora 2019
СПбГУ, 28 июня 2019

Корпус SCAT2018

- <http://portal.textometrie.org/demo>
 - 24 текста, 186 542 с/у
 - свободный доступ онлайн
 - возможность скачивания «бинарного корпуса»

Корпус SCAT2018-CIT

- 2 текста

- Житие Дионисия Глушицкого, 9 606 с/у
- Житие Григория Пельшемского, 7 105 с/у

Визуалізація

■ CSS

<22>

м̃ ца іюна дн̃ ъ л̃ житіє и подвизи и Ѡчасти
чюдѣ исповѣда̃ є прї вна̃ Ѡнше̃ дїонїсіа глѣ цка̃
;
се во нн̃ ѣ о сѣ бл̃ женѣ свѣтлыи оуч̃ тль ,
славныи павѣ в̃пїє̃ . их̃ ж̃ жи̃ є на нв̃ сѣ̃ є̃ .
5 наслаж̃ аа̃ чюство̃ безвещны̃ нена
сытна̃ . и невїдима̃ добротою видѣ̃ .
акы разжмна̃ , разжм̃ и оум̃ точїю . р̃а ѱ
аса достоина̃ потѣ̃ прїє̃ , Ѡданїа трѣдѣ̃ ,
ѣ̃ трѣднѣю почѣ̃ . іа̃ тамо при̃ стїє̃ прїє̃ .
10 и пр̃ но с тѣми сы̃ , их̃ ж̃ нога ста на право̃̃ .
но оба̃ оуне Ѡ трѣ̃ вны̃ тѣ̃ потѣ̃ , на̃ ѡдобрити сло̃̃
.

Конкорданс

■ <q> []+ </q>

Import XML-TEI Zero + CSV de SCAT2018-CIT | DGlush - 2 | SCAT2018-CIT <q> []+ </q> ☒

Requête

| ref | Pivot |
|---------------|--|
| DGlush 22/3 | свѣтлыи оучѣтъ славныи павѣ въпїтѣ ихъ жиѣ на нвсѣ ѣ |
| DGlush 22/5 | наслаждаа чюствѣо безвещныи ненасытнаи невїдима добротѣю видѣа акы разжмна разжмѣо и оумѣо точїю раѣаса |
| DGlush 22/10 | нога ста на правоѣ |
| DGlush .22/9 | рава оногѣо погревшѣа в земли талантѣ |
| DGlush 23/3 | рѣ вѣ гѣ любви ма |
| DGlush 23/3 | в законѣ поучитѣса днѣ нѣ |
| DGlush 23/4 | влженѣ члкъ иже овраще премадрѣ |
| DGlush 23/9 | пытайте рѣ гѣ писанїа акв в тѣ имате оврѣсти жиснѣ вѣчнїю |
| DGlush 23/14 | разъшири оуста моа и наполни а |
| DGlush .23/17 | са во рѣ гѣ призовите ма и оуслышоѣ ва |
| DGlush .23/18 | ги са рекѣ еси оученикѣо таче и къ всѣм в та вѣроуѣщїи акв без мене не можете творити ничтѣо |
| DGlush .24/17 | в семь въпїетѣ вѣщїцѣ прѣкъ двѣдѣ срце съкржшено и смѣренно вѣ не оуничижї |
| DGlush .24/19 | и аплѣо глѣущѣоу творащїи заповѣди вса поспїютѣса вѣ блгоѣ |
| DGlush 25/5 | хоцѣ вѣ всѣа в разж прити и вели всѣа спастиса |
| DGlush 25/7 | добрѣдѣтелми възидѣ на горж и да оуслышїи асно что глѣть гѣ вѣ в на глѣть во мурѣ и просвѣщенїѣ и исцѣленїѣ дшевыи съкржшенїѣемѣ |
| DGlush 25/13 | |

1 - 100 / 183

Конкорданс → Текст

- `<q> []+ </q>`

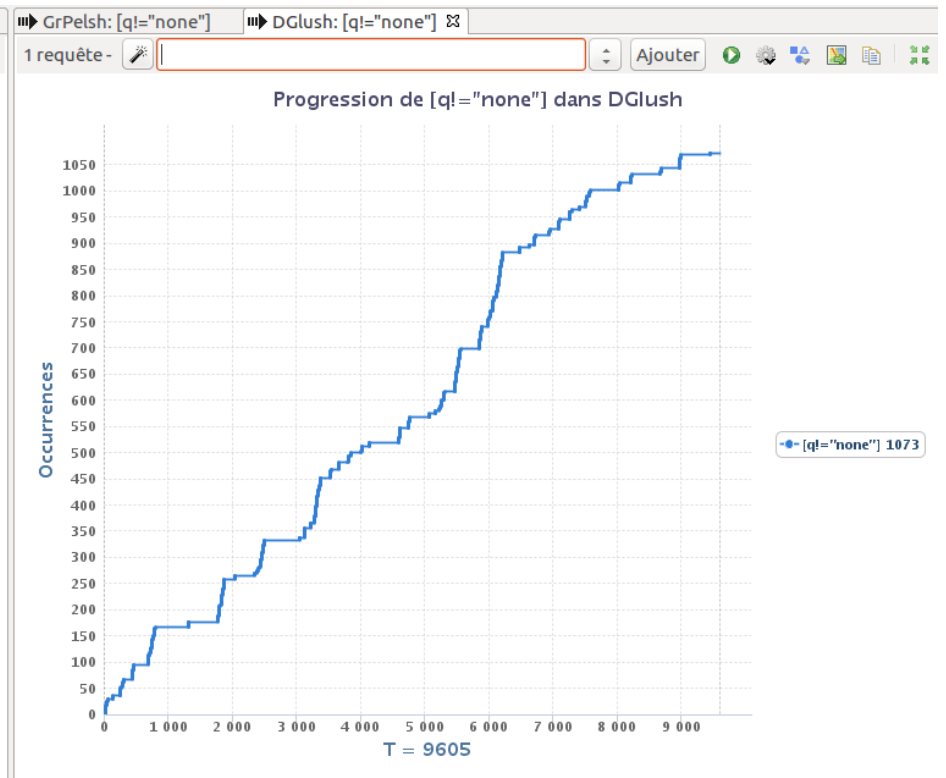
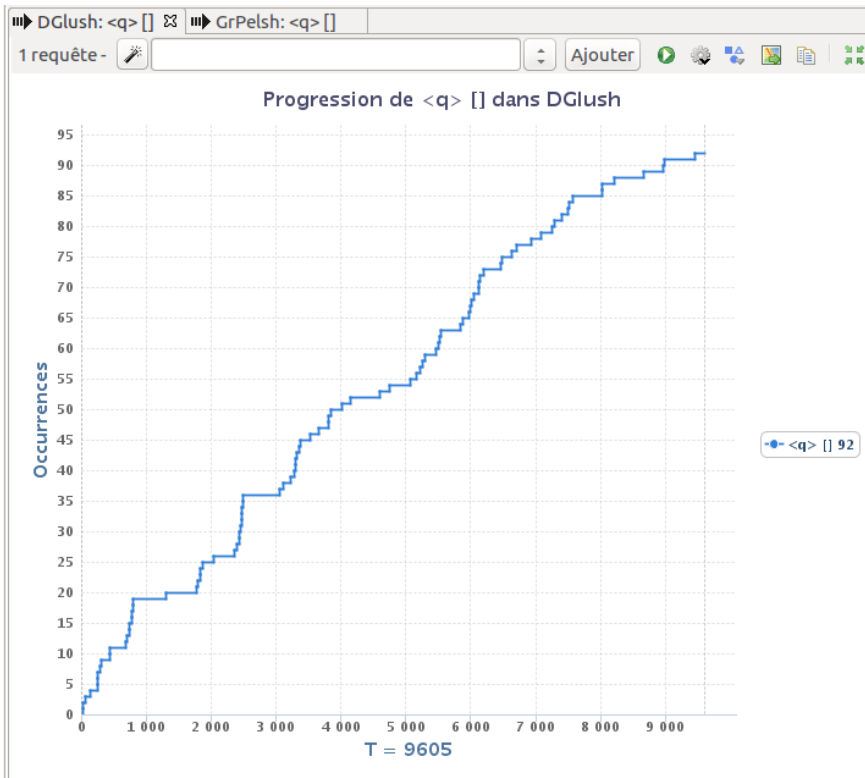
The screenshot shows a software interface for text analysis. The top window, titled "DGlush - 4", displays a list of text fragments with highlighted search results. The text is in Cyrillic and includes words like "жнѧго поути шествѣѧ . рекше по", "вжѣ твенѣ писанѣ окормити житѣ", "е свое . рѣ вѣ гѣ любви ма , в законѣ", "поучитьсѧ днѣ нѣ . блженъ члкъ", "5 иже овраще премадрть . семжѣ пж", "ти на ло стѣ оцѣ оученѣ . томоуѣ воу", "пслоуѣ дхѣ , стѣи . ѣи наоучивѣи пре".

The bottom window, titled "Import XML-TEI Zero + CSV de SCAT2018-CIT", shows a search interface. The search query is "`<q> []+ </q>`". The search results are displayed in a table with columns "ref" and "Pivot". The selected entry is highlighted in orange:

| ref | Pivot |
|--------------|-------------------------------------|
| DGlush 22/10 | нога ста на правоѣ |
| DGlush .22/9 | рава оногѣ погребшѣ в земли талантѣ |
| DGlush 23/3 | рѣ вѣ гѣ любви ма |
| DGlush 23/3 | в законѣ поучитьсѧ днѣ нѣ |
| DGlush 23/4 | блженъ члкъ иже овраще премадрть |

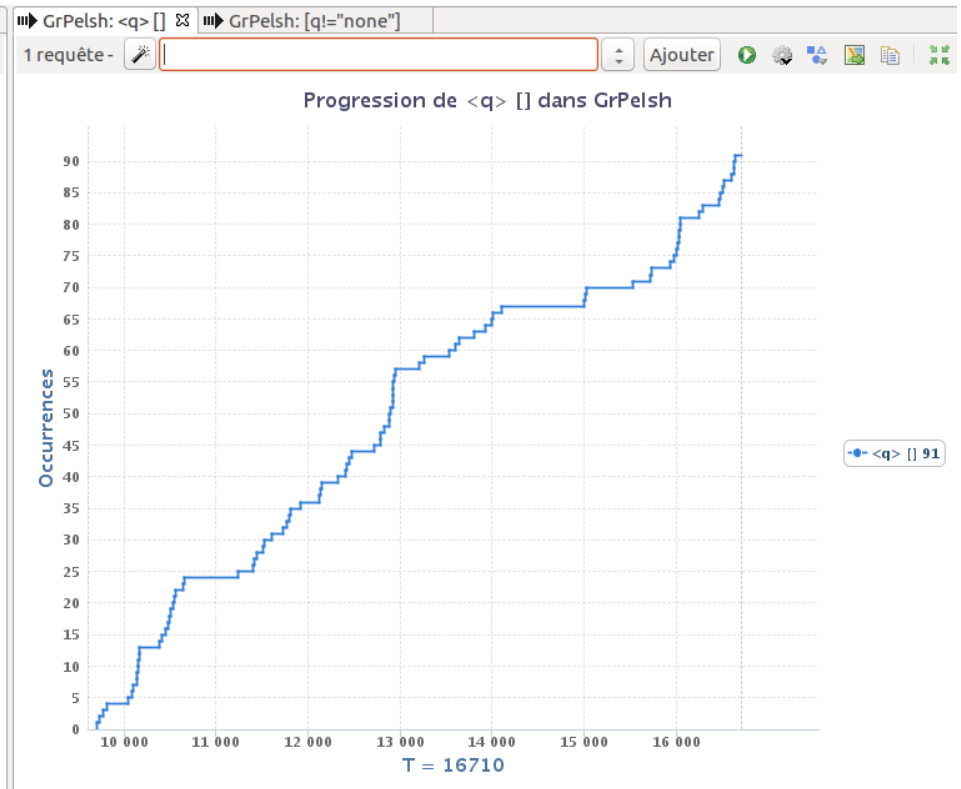
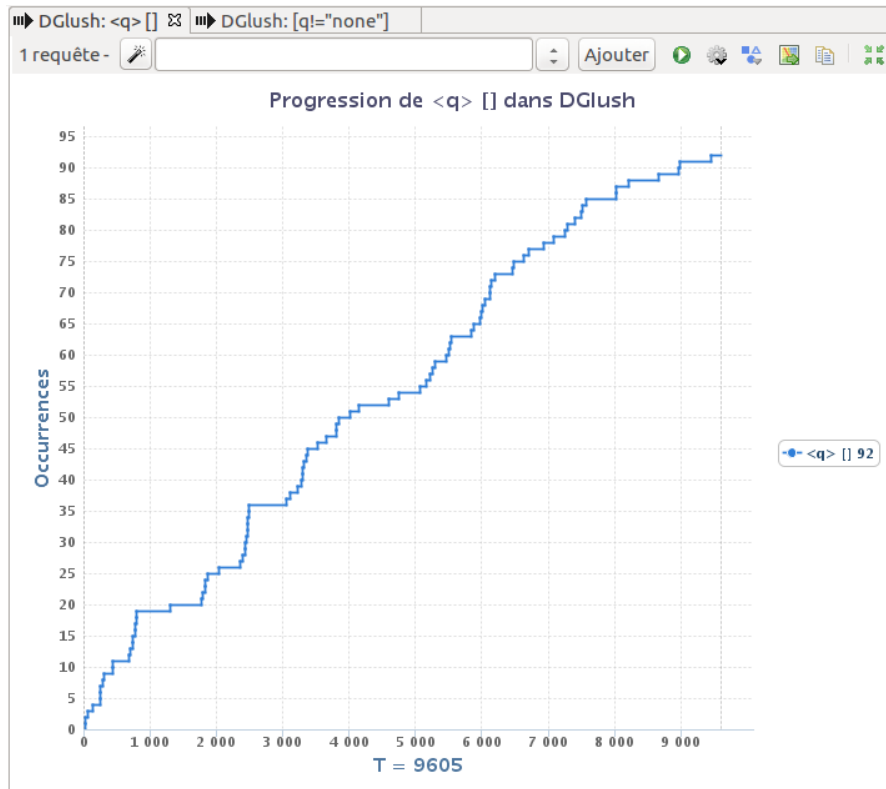
Прогрессия

■ Первые слова в цитатах / все слова



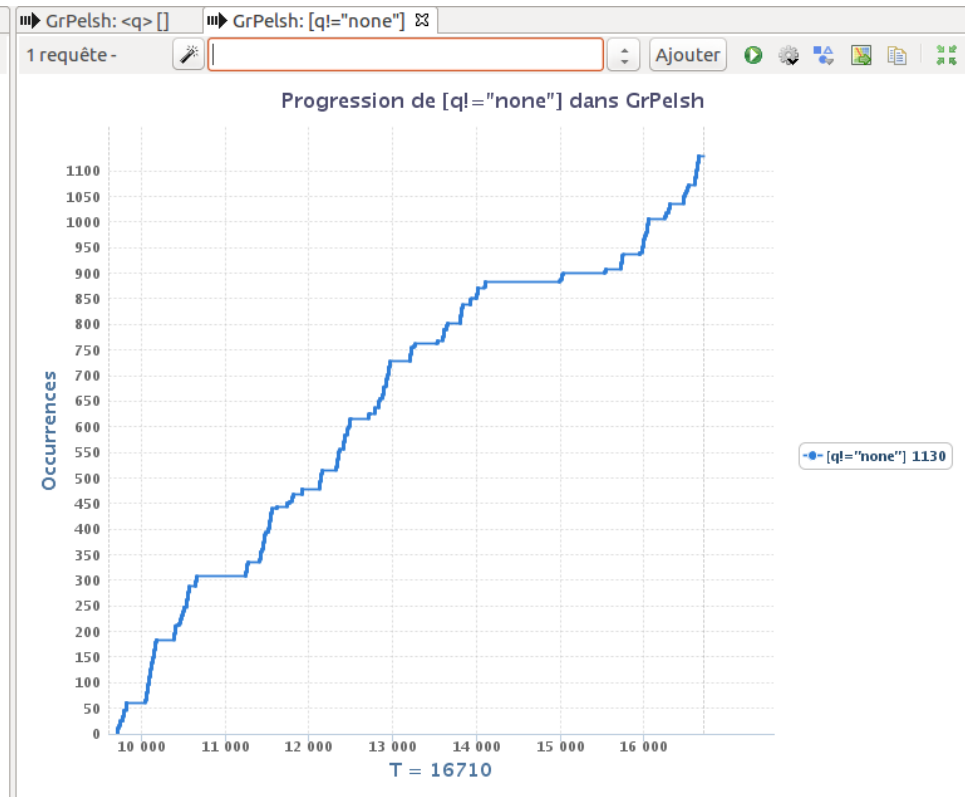
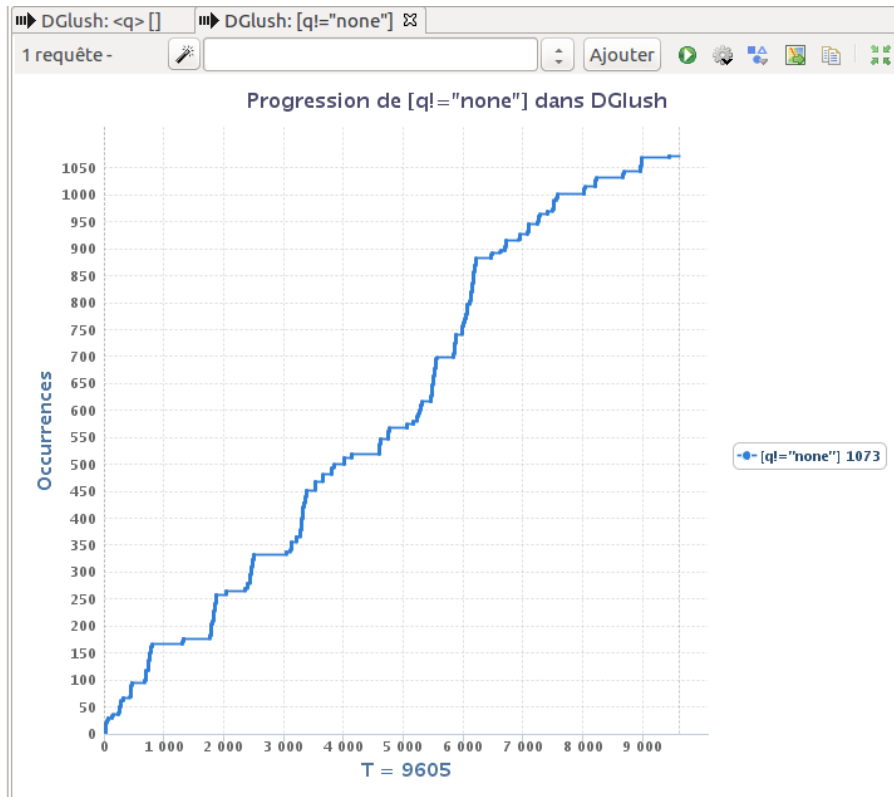
Прогрессия

■ Первые слова в цитатах



Прогрессия

■ Все слова в цитатах



Частотный словарь

■ Число текстоформ по типам цитаты

AZ SCAT2018-CIT: q

Propriétés : q Éditer

| q | Fréquence |
|----------------|-----------|
| no-cit | 14508 |
| standard | 935 |
| modified | 738 |
| author-speaker | 451 |
| allusion | 79 |

AZ SCAT2018-CIT: q AZ Index: texts: []: q


Requête

| q | Fréquence T=16711 | DGlush t=9606 | GrPelsh t=7105 |
|----------------|-------------------|---------------|----------------|
| no-cit | 14508 | 8533 | 5975 |
| standard | 935 | 430 | 505 |
| modified | 738 | 345 | 393 |
| author-speaker | 451 | 250 | 201 |
| allusion | 79 | 48 | 31 |

Частотный словарь

■ Словоформы в цитатах

AZ SCAT2018-CIT: q AZ Index: texts: []: q AZ Index: texts: [q!="no-cit|author-speaker"]

Requête  [q!="no-cit|author-speaker"]

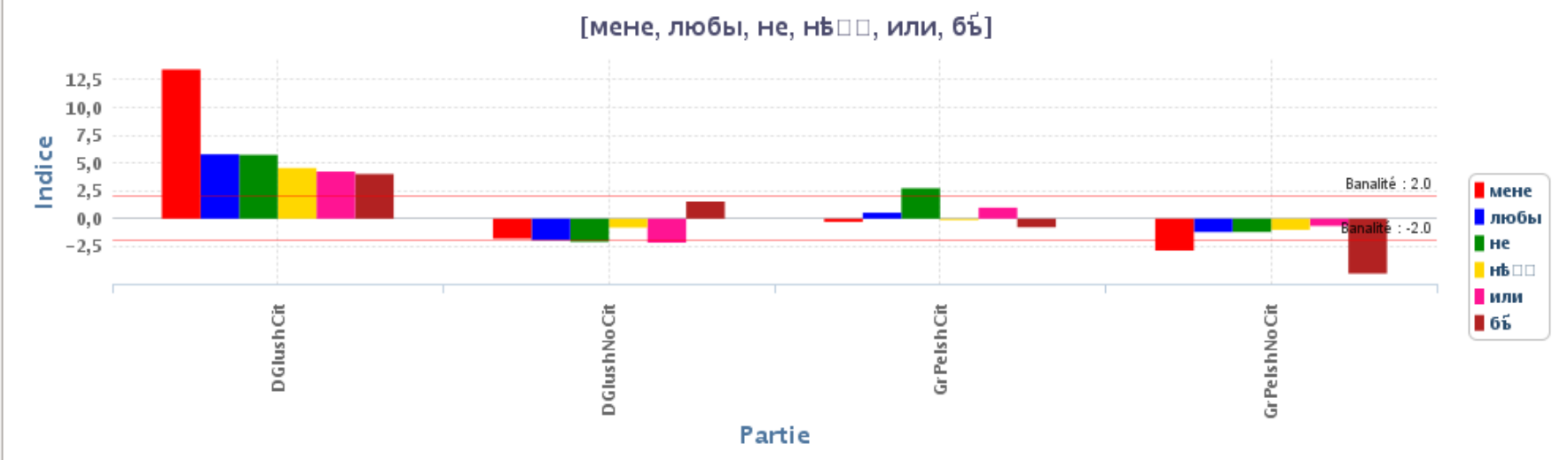
| word | Fréquence T=16711 | DGlush t=9606 | GrPelsh t=7105 |
|------|-------------------|---------------|----------------|
| и | 124 | 52 | 72 |
| не | 47 | 26 | 21 |
| в | 32 | 13 | 19 |
| на | 26 | 8 | 18 |
| ω | 20 | 10 | 10 |
| мене | 17 | 16 | 1 |
| въ | 16 | 7 | 9 |
| яко | 15 | 4 | 11 |
| есть | 14 | 7 | 7 |
| но | 14 | 6 | 8 |
| а | 13 | 6 | 7 |
| его | 13 | 4 | 9 |

Специфичность

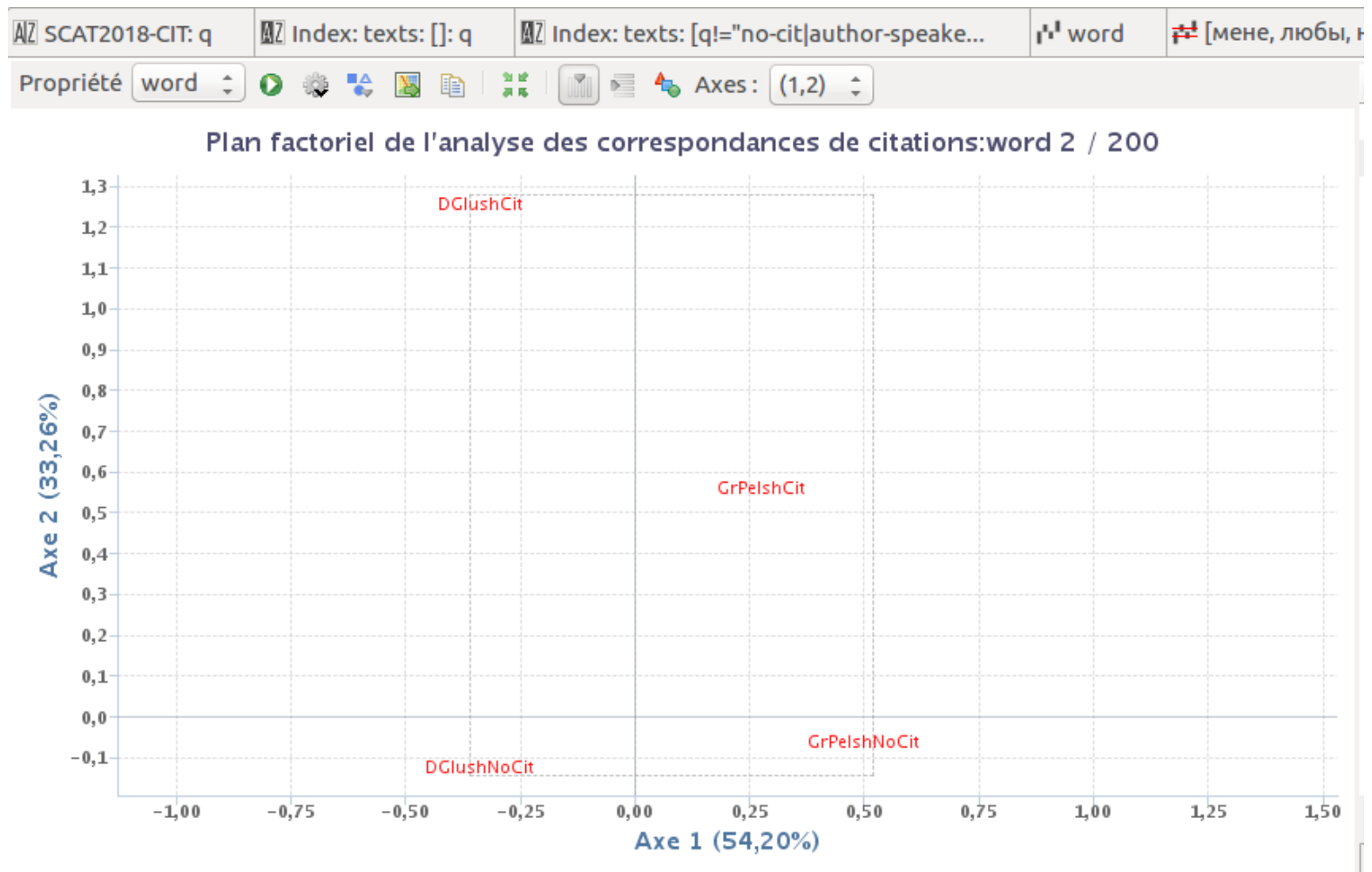
| Unités | Fréquence T 16711 | DGlushCit t=823 | ▲ indice | DGlushNoCit t=8783 | indice | GrPelshCit t=929 | indice | GrPelshNoCit t=6176 | indice |
|--------|-------------------|-----------------|----------|--------------------|--------|------------------|--------|---------------------|--------|
| мене | 29 | 16 | 13,4 | 9 | -1,8 | 1 | -0,3 | 3 | -2,9 |
| любы | 6 | 5 | 5,8 | 0 | -1,9 | 1 | 0,5 | 0 | -1,2 |
| не | 188 | 26 | 5,7 | 82 | -2,1 | 21 | 2,7 | 59 | -1,2 |
| нѣ | 5 | 4 | 4,6 | 1 | -0,8 | 0 | -0,1 | 0 | -1,0 |
| или | 10 | 5 | 4,2 | 1 | -2,2 | 2 | 1,0 | 2 | -0,7 |
| бѣ | 31 | 8 | 4,0 | 22 | 1,5 | 0 | -0,8 | 1 | -4,9 |

люб
 Ѡид
 вā

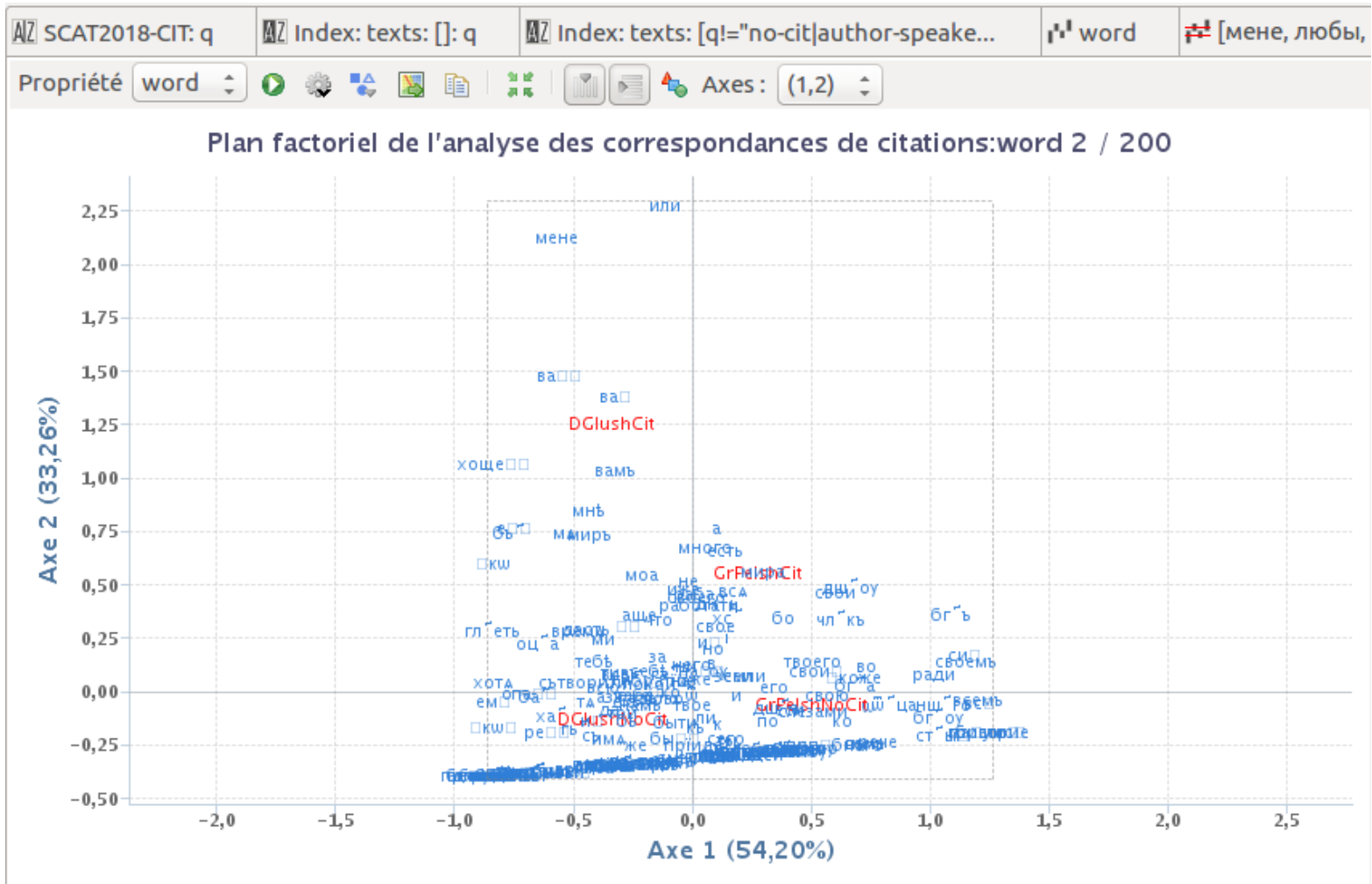
SCAT2018-CIT: q Index: texts: []: q Index: texts: [q!="no-cit|autho... word [мене, любы, не, нѣ, или, бѣ]



Факторный анализ



Факторный анализ



Перспективы

- Гипертекстовая навигация
- Анализ по леммам, граммемам, буквосочетаниям...