The online corpus of the GRE*g*ORI Project – Instruction manual – May 2021

Table of contents

Table of figures

*

*   *

1 Introduction

The GRE*g*ORI Project – hosted at the Institut orientaliste of the UCLouvain, Louvain-la-Neuve, Belgium – provides scholars with lemmatized corpora in Classical and Byzantine Greek as well as in some of the main languages of the Christian East, such as Ancient Armenian, Old Georgian, and Syriac. These data are now available online through interfaces combining a concordancer, a text browser and an editions viewer.

These corpora put together digital versions of texts taken from critical editions and enriched by a complete lemmatization and part-of-speech tagging. Users can submit queries based on word-form, lemma, or part of speech tags. Results appear in the shape of a "Key Word In Context" concordance.

Each item of the concordance is also linked to a set of metadata, including not only information about the text (author, work, topic, date, etc.), wider context, lexical analysis, and bibliography, but also a snapshot of the text and its translation as they appear in the printed edition.

Text processing is made possible by the tools and linguistic resources developed by the <u>GREgORI Project</u>.

Interfaces initially developed by the <u>CENTAL - Centre de Traitement Automatique du Langage</u> (UCLouvain, Belgium) have been completed in collaboration with the company <u>WebFabrick</u> (Paris, France).

These interfaces will remain available free of charge.

In the near future, the GRE*g*ORI Project will regularly add new lemmatized data and develop interfaces featuring multilingual search capabilities.

<div style="border:1px solid">

We share tools and data with our collaborators.
Information, bug report and contact: info-gregori@uclouvain.be

</div>

2 Home page

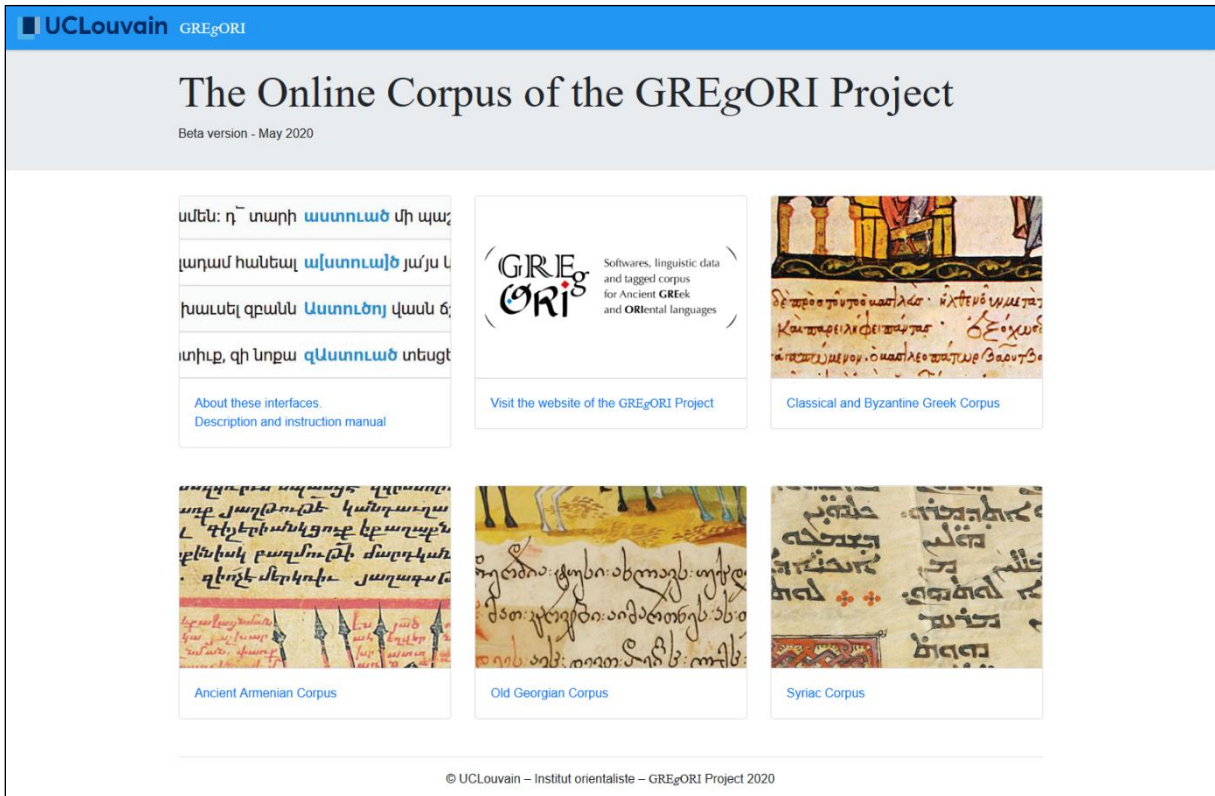The "Home page" of the online corpus of the GRE*g*ORI Project is available <u>here</u>.

Figure 1 : The "Home page" of the online corpus of the GREgORI Project

Users can select one of the available corpus (Classical and Byzantine Greek, Ancient Armenian, Old Georgian, and Syriac).

The "About these interfaces. Description and instruction manual" menu gives access to descriptions of the interfaces and to instructions for different ways to use them.

The "About the GRE*g*ORI Project" menu gives access to the GRE*g*ORI Project website.

There are two main ways of using these tools: a "Simple search" mode and an "Advanced search" mode.

3 Simple search

3.1 Buttons, icons, and research field description (1/2)
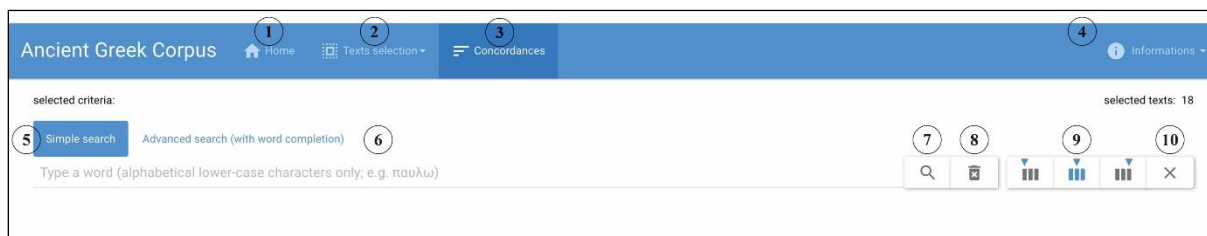

Figure 2 : Buttons, icons, and research field description (1/2)

1. "Home" button: back to the Home page
2. "Text selection" button: display the "Text selection" window (see below, 5)
3. "Concordances" button: display search modes and concordances setup menus
4. "Information" icon: display information about the corpus and search modes (part-of-speech tagset description, corpus description, including bibliography, etc.)
5. "Simple search" button: switch to the "Simple search" mode
6. "Advanced search" button: switch to the "Advanced search" panel
7. "Magnifier" icon: submit query
8. "Trash box" icon: delete query
9. "Sort on the left context", "Sort on the search result", and "Sort on the right context" buttons: allow alphabetical sorting of the concordance (text order by default)
10. "Reset contexts and keywords order settings" button: reset order settings of the concordance

3.2 Requests

The "Simple search" mode allows users to search word-form (alphabetical lower-case characters only).

Note that a "Simple search" based on the request "αγαπη" will provide the concordance of all the forms ἀγάπη, but also Ἀγάπη, ἀγάπῃ, or Ἀγάπῃ.

By default, requests cover the whole corpus. To select a sub-corpus, use the "Text selection" menu (see 5, below).

Samples of requests:

Language	Request	Result
Classical and Byzantine Greek	παυλω	*These requests display the concordance of all the word-forms matching with this characters string*
Ancient Armenian	պաւղոսի	
Old Georgian	პავლესსაცა	
Syriac	ܕܝܠܗ	

By default, requests cover the whole corpus. To select a sub-corpus, use the "Text selection" menu (see 5, below).

Click on the magnifier (or press ENTER) to display the concordance.

> Note that the request also takes into account word-forms included in crases (in Greek) or in compound word-forms combining attached prepositions and/or pronouns, or attached determiners (as in Armenian, Georgian and Syriac).

> If a text is not lemmatized yet, we advise users to search for word-forms only, by using the "Advanced search" mode, and by selecting the "contains" operator (in place of the "equals" operator).
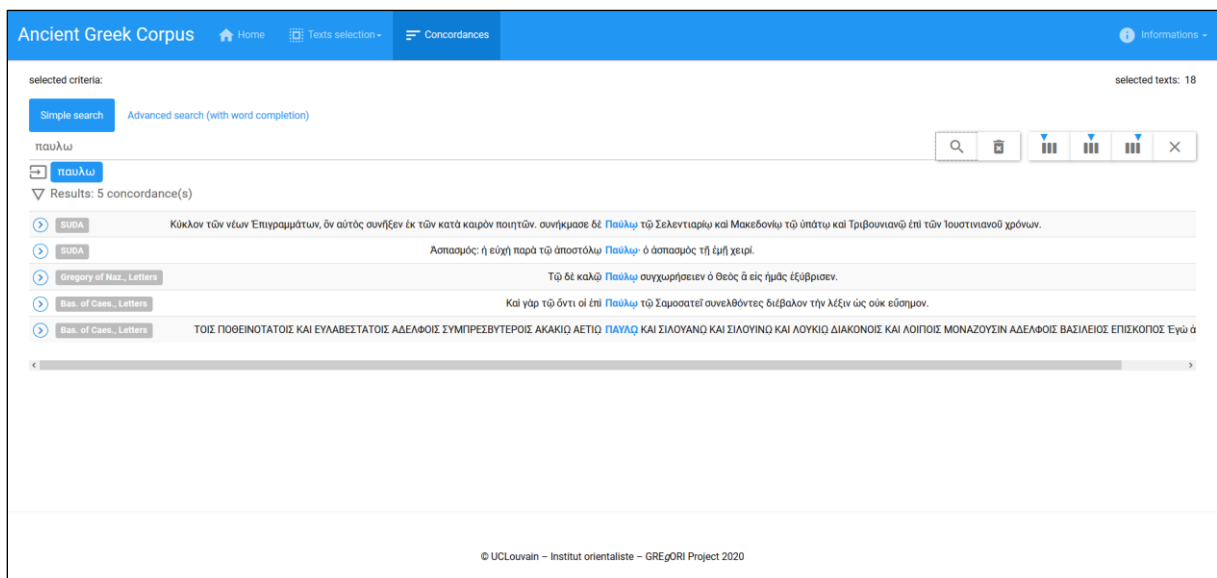
3.3 Concordance display

Figure 3 : Concordance display

Each line of a concordance starts with an arrow inscribed in a circle (⊙) followed by a tag. This tag identifies the author and/or work to which the quoted text belongs (e.g. Gregory of Naz., Letters for "Gregory of Naz., Letters" referring to the *Letters* of Gregory of Nazianzus).

Clicking on the arrow displays the "Concordance details", consisting of:
- Full references to the text edition
- Links to the text and translation in Adobe PDF format (as they appear in the printed edition), or to other online applications
- A wider context
- Lexical analysis of the nearest word-forms
- Information about the text and the edition

4 Advanced search

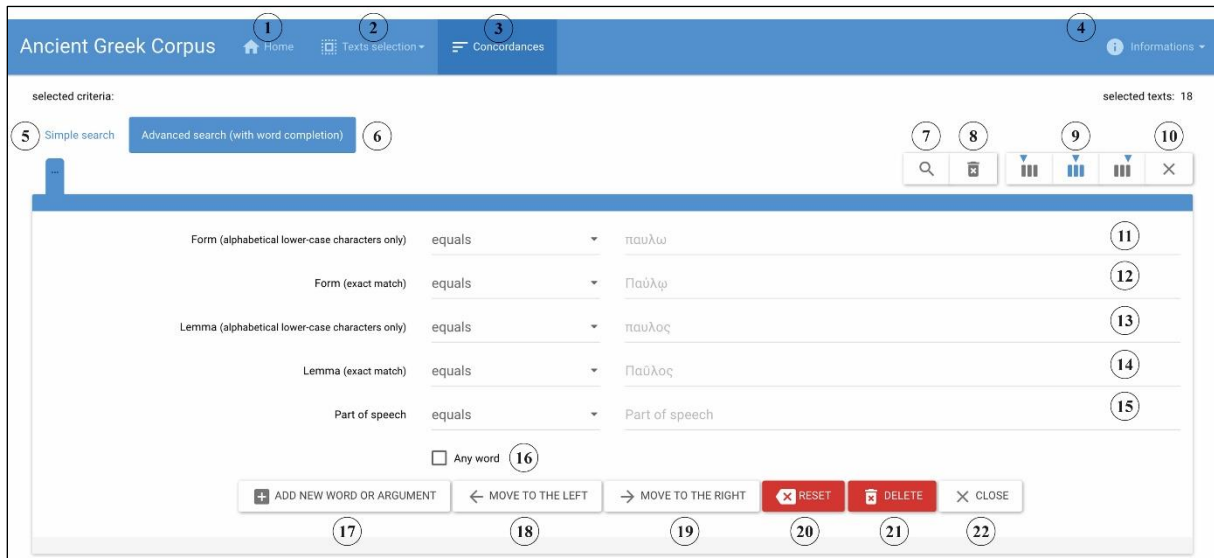## 4.1 Buttons, icons, and search fields (2/2)

Figure 4 : Buttons, icons, and research field description (2/2)

1. "Home" button: back to the Home page
2. "Text selection" button: display the "Text selection" window (see below, 5)
3. "Concordances" button: display search modes and concordances setup menus
4. "Information" icon: display information about the corpus and search modes (part-of-speech tagset description, corpus description, including bibliography, etc.)
5. "Simple search" button: switch to the "Simple search" mode
6. "Advanced search" button: switch to the "Advanced search" panel
7. "Magnifier" icon: submit query
8. "Trash box" icon: delete query
9. "Sort on the left context", "Sort on the search result", and "Sort on the right context" buttons: allow alphabetical sorting of the concordance (text order by default)
10. "Reset contexts and keywords order settings" button: reset order settings of the concordance
11. "Form (alphabetical lower-case characters only)" search field
12. "Form (exact match)" search field
13. "Lemma (alphabetical lower-case characters only)" search field
14. "Lemma (exact match)" search field
15. "Part of speech" search field
16. Checkbox "Any word" (see below)
17. "+ Add new word or argument": allow adding another search term
18. "Move to the left" button: move the current word or argument to the left
19. "Move to the right" button: move the current word or argument to the right
20. "Reset" button: reset all fields
21. "Delete" button: delete the current word or argument
22. "Close" button: close this panel

4.2 One word search

The "Advanced search" mode allows users to type a form, a lemma, or a part of speech tag in the corresponding search fields. Clicking on the magnifier generates a concordance based on that search.

Note that the concordance appears below the "Advanced search" panel. According to your screen resolution or display size, it could be useful to change the display settings on your computer or to use the "close" button in order to hide the "Advanced search panel" and to see better the concordance.

Samples of requests:

Classical and Byzantine Greek	Request	Result
Form (alphabetical lower-case characters only)	παυλω	*These requests display the concordance of all word-forms matching this character string*
Form (exact match)	Παύλω	
Lemma (alphabetical lower-case characters only)	παυλος	*These requests display the concordance of all occurrences of the lemma matching this character string*
Lemma (exact match)	Παῦλος	
Ancient Armenian		
Form (alphabetical lower-case characters only)	պաւղոսի	*These requests display the concordance of all word-forms matching this character string*
Form (exact match)	Պաւղոսի	
Lemma (alphabetical lower-case characters only)	պաւղոս	*These requests display the concordance of all occurrences of the lemma matching this character string*
Lemma (exact match)	Պաւղոս	
Old Georgian		
Form (alphabetical lower-case characters only)	პავლესსაცა	*These requests display the concordance of all word-forms matching this character string*
Form (exact match)	პავლესსაცა	
Lemma (alphabetical lower-case characters only)	პავლე	*These requests display the concordance of all occurrences of the lemma matching this character string*
Lemma (exact match)	პავლე	
Syriac		
Form (alphabetical lower-case characters only)	ܦܘܠܘܣ	*These requests display the concordance of all word-forms matching this character string*
Form (exact match)	ܦܘܠܘܣ	
Lemma (alphabetical lower-case characters only)	ܦܘܠܘܣ	*These requests display the concordance of all occurrences of the lemma matching this character string*
Lemma (exact match)	ܦܘܠܘܣ	

4.3 Multi-word search

The "+ Add new word or argument" button allows adding new terms in order to search for an expression combining several elements.
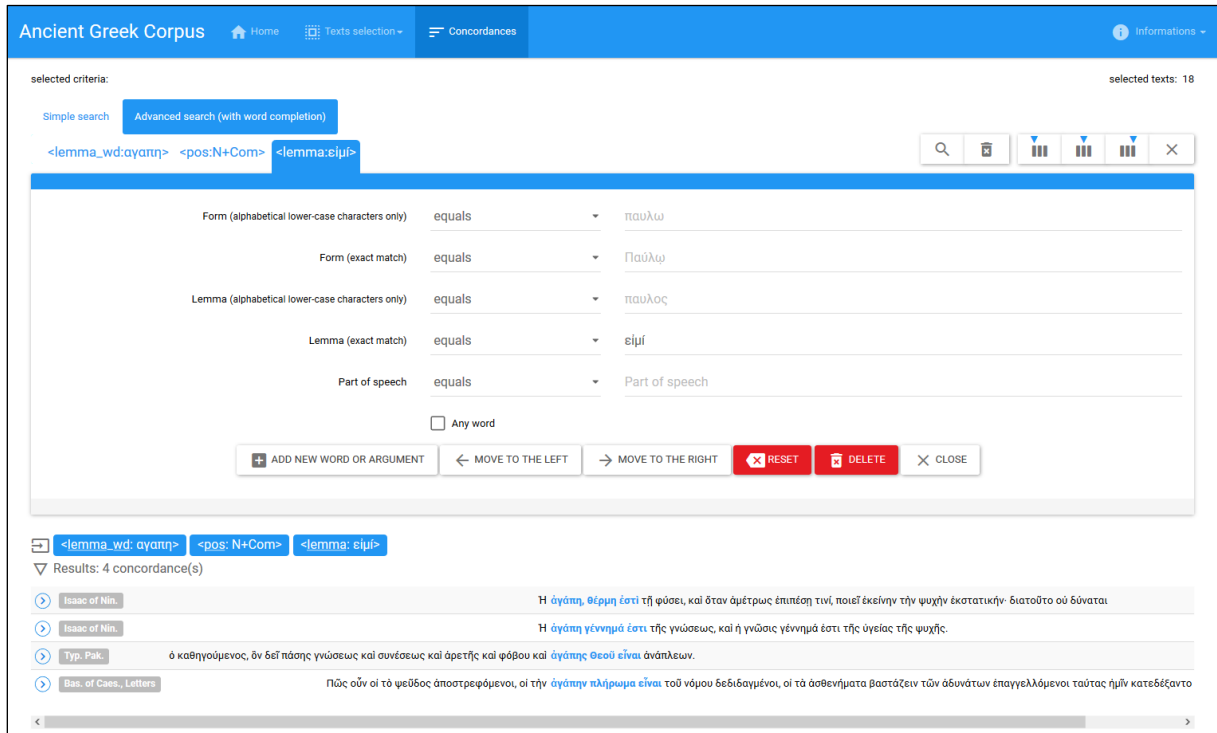
Next buttons allow users to modify words or arguments order in the expression, to delete a word, to reset search fields, or to close the "Advanced search" panel (see 4.1. above). The "Any word" button allows inserting any word without specifying its form, lemma, or part of speech.

Clicking on the magnifier generates the concordance corresponding to the search.

Note that the concordance appears below the "Advanced search" panel. According to your screen resolution or display size, it could be useful to change the display settings on your computer or to use the "close" button in order to hide the "Advanced search panel" and to see better the concordance.

Samples of requests:

In this way, it is possible to search for an expression combining all the word-forms of the lemma ἀγάπη that are followed both by a common noun and a word-form of the lemma εἰμί. The result displays the concordance below.

Figure 5 : Sample of multi-word search

5 Text selection

The "Text selection" menu opens a page displaying various criteria.

By default, a search covers the whole corpus. The "Text selection" page makes it possible to limit the search to a given sub-corpus. Available criteria are authors, works, centuries, and transmission types (manuscripts, inscriptions, etc.).
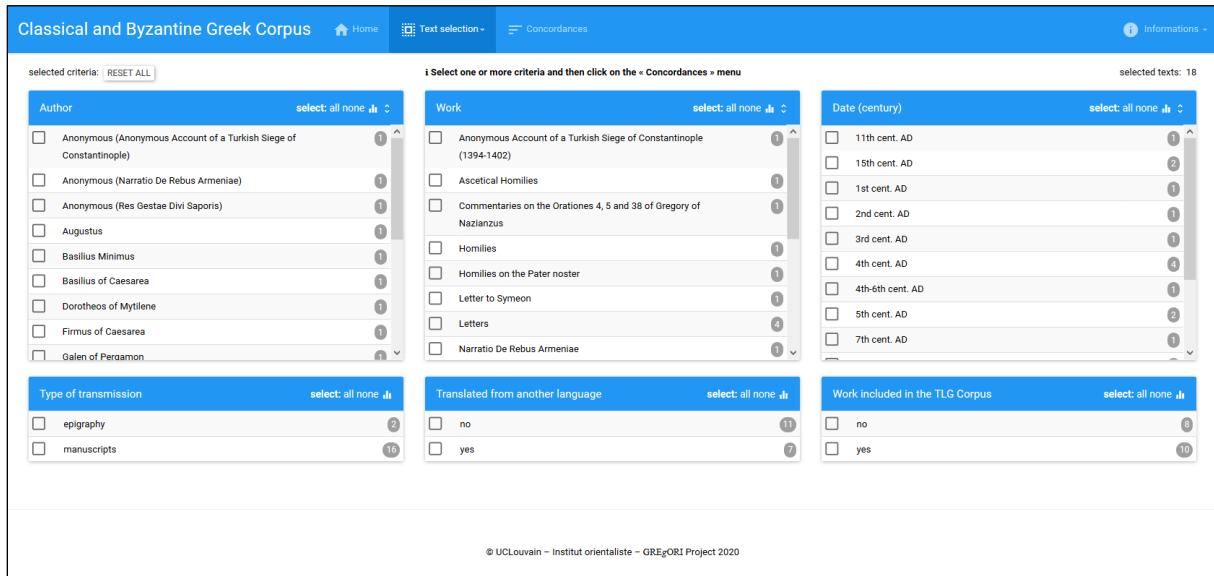
Figure 6 : The "Text selection" menu

It is also possible to select texts initially written in another language ("Translated form another language" – yes/no) and, in Greek, texts included or not in the Thesaurus Linguae Graecae corpus ("Work included in the TLG Corpus" – yes/no).

Users are free to choose one or more criteria.

Clicking again on the "Concordances" menu displays the simple or advanced search menus to search this specific sub-corpus.

Clicking on the "Reset all" button removes all criteria and performs the search in the entire corpus.

If a text is not lemmatized yet, we advise users to search for word-forms only, by using the "Advanced search" mode and by selecting the "contains" operator (in place of the "equals" operator).