

# Processing Tools for Greek and Other Languages of the Christian Middle East

Bastien Kindt

# ▶ To cite this version:

Bastien Kindt. Processing Tools for Greek and Other Languages of the Christian Middle East. Journal of Data Mining and Digital Humanities, 2018, Special Issue on Computer-Aided Processing of Intertextuality in Ancient Languages, 10.46298/jdmdh.4184. hal-01671592

# HAL Id: hal-01671592 https://hal.science/hal-01671592

Submitted on 22 Dec 2017

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers. L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.





# Processing Tools for Greek and Other Languages of the Christian Middle East

# **Bastien Kindt**

Research Assistant – Université catholique de Louvain, Belgium

bastien.kindt@uclouvain.be

### Abstract

This paper presents some *computer tools* and *linguistic resources* of the GREgORI project. These developments allow automated processing of texts written in the main languages of the Christian Middel East, such as Greek, Arabic, Syriac, Armenian and Georgian. The main goal is to provide scholars with tools (lemmatized indexes and concordances) making corpus-based linguistic information available. It focuses on the questions of text processing, lemmatization, information retrieval, and bitext alignment.

### Keywords

Lemmatization; Greek; Syriac; Arabic; Armenian; Georgian; lexical tagging; POS tagging; Unitex; concordances; indexes; finite state transducers; bitext; bilingual alignment; translation memories; Unitex; mkAlign.

# INTRODUCTION

The GREgORI project, carried out for many years at the Oriental Institute of the Université catholique de Louvain in Louvain-la-Neuve (Belgium) (https://uclouvain.be/fr/instituts-recherche/incal/ciol/gregori-project.html), brings together linguists, philologists and computer scientists around a common goal: developing *computer tools* and *linguistic resources* allowing automated processing of texts written in Greek [Coulie, 1996; Kindt, 2004; Kindt, 2010] and in the languages of the Christian Middle East, such as Middle Arabic [Tuerlinckx, 2004], Syriac, Armenian and Georgian [Coulie, Kindt and Pataridze, 2013]. From the point of view of computational approach all these languages still remain, especially in the case of the ancient texts, poorly resourced. These *tools* and *resources* make it possible to turn raw texts into tagged corpora enriched with lexical (lexical entry or lemma), morphosyntactical (noun, verb, adjective, etc.), and inflectional (case, gender, number, voice, tense, person, etc.) labels, defining each word-form of processed texts.

This work can lead to both monolingual and multilingual corpora. In the first case, tagged data is used in creating lemmatized concordances and indexes. It is also exploited with corpus

analysis software or data retrieval system. In the second case, the analysis is focused on translation methods. Lexical data of a *source text* is related to lexical data of a *target text* in order to highlight how lexical features from a source language (in relation with a specific cultural area) are spread and received in a target language (linked with another cultural area). In both cases, the main purpose is to provide researchers with linguistic information exclusively drawn from corpus-based evidences.

This paper focuses on the Unitex corpus processor, on the one hand, and on specific outputs such as monolingual or bilingual concordances, indexes and other lexicographical lists provided in the shape of PDF documents, on the other hand. Since two contributions included in the present issue, [Pataridze and Kindt, 2017] (*Text Alignment in Ancient Greek and Georgian: A Case-Study on the First Homily of Gregory of Nazianzus*) and [Van Elverdinghe, 2017] (*Recurrent Pattern Modelling in a Corpus of Armenian Manuscript Colophons*), are founded on the developments presented here, this paper should be regarded as a "general introduction" to these two more specific studies. Given that these contributions concern Armenian and Georgian languages, this paper deals with Greek and Syriac languages.

### I. TEXT PROCESSING AND LEMMATIZING

Unitex is an open source corpus processor [Paumier, 2016 and http://unitexgramlab.org/fr]. It supports the Unicode encoding standard and is language independent. This software suite works with specific resources, such as *dictionaries* [Paumier, 2106:43-74] and *grammars* [Paumier, 2106:93-160]. *Dictionaries* provide word-forms labeled with lemmata and grammatical tags (part-of-speech tags, inflectional tags). *Grammars* are user-friendly graphical representations of lexical or syntactic patterns based on the formalism of recursive transition networks.

### 1.1 Unitex

The following lines illustrate how Unitex allows the user to open, to process, and to explore a Greek corpus presently composed by two hundred and forty-nine letters of Gregory of Nazianzus (329-390) published by [Gallay, 1964-1967] and [Gallay and Jourjon, 1976]. This data includes 45,971 word-forms (11,318 different forms; 4,452 lemmas). This corpus is part of a bigger corpus entitled *Corpus Epistularum Patrum Cappadocum – Corpus épistolographique des Pères cappadociens*, bringing together the letters of Basilius of Caesarea, Gregory of Nyssa, Gregory of Nazianzus, and Firmus of Caesarea, henceforth available on the website of the GREgORI project.

When user opens a corpus, the interface proposes a three-step preprocessing: 1) splitting sentences, 2) normalization of the special words or expressions, and 3) lexical lookup.

Unitex deals with the sentence as a linguistic unit. The texts are split up into sentences based on punctuation. This starting process is followed by the normalization task aimed at improving analysis of some words or expressions. In Greek, this step deals with the *crasis* or *multiword expressions* containing one lexical unit corresponding (out of context) to more than one analysis, but being non ambiguous in the given sequence.

The graphical forms of the *crasis* combine two lexical units written as a single word. The process has to provide an analysis identifying clearly the presence of these two lemmata in the single form. For instance, the forms κἀγὼ "and I" and τοὕδαφος "the ground-floor" are processed offering a lexical description including, respectively, καί "and" and ἐγὼ "I" and ὑ "the" (form τὸ) and ἕδαφος "ground-floor". After preprocessing, the word-form κἀγὼ in the sentence Πλὴν ὥσπερ ὑμεῖς κὑριοι τῆς ὑμετέρας γνώμης, οὕτω κἀγὡ τῆς ἐμῆς (Letter 153, 2)

is rewritten {@κἀγώ@,2.K} {καί,καί.I+Part} {ἐγώ,ἐγώ.PRO+Per1s} and τοὕδαφος, in the sentence κόπτω μὲν οἶον **τοὕδαφος** τοῖς ποσί (Letter 153, 2), is rewritten {@τοὕδαφος@,2.K} {τό,ό.DET:Nns:Ans} {ἕδαφος,ἕδαφος.N+Com}. These changes are exclusively applied to the text handled by Unitex and do not modify the original version provided by the published edition. In any case, the concordances are displaying the original Greek texts.

Among the *expressions* containing one ambiguous word without context, we can quote  $\dot{\omega}_{\zeta} \circ \tilde{i} \mu \alpha i$ *"in my opinion"* or  $\ddot{\omega}_{\zeta} \phi \eta \sigma i$  *"as we say"*. The word-form  $\dot{\omega}_{\zeta}$  is either the conjunction  $\dot{\omega}_{\zeta}$  *"as"* or the relatively less frequent preposition  $\dot{\omega}_{\zeta}$  *"in, by"*, synonymous with the more common  $\varepsilon i_{\zeta}$ , but exclusively used with a name of person or place such as in the sentence  $\dot{\alpha}\phi$ ( $\kappa\varepsilon\tau\sigma$   $\dot{\omega}_{\zeta}$  Περδίκκαν καὶ ἐς τὴν Χαλκιδικήν *"he arrived at Perdiccas in Chalkidiki"* [Liddell-Scott, 1958, p. 2039]. Followed by oἶμαι or φησι the word-form  $\dot{\omega}_{\zeta}$  is unequivocally the conjunction. For these two expressions, the sentences are rewritten, after preprocessing, as  $\kappa\varepsilon\kappa\lambda$ ήκατε ἡμᾶς ἐπὶ τὴν μητρόπολιν, { $\dot{\omega}_{\zeta},\dot{\omega}_{\zeta}$  ( $\ddot{\omega}_{\zeta},\dot{\omega}_{\zeta}$  ( $\ddot{\omega}_{\zeta},\dot$ 

The third step of the preprocessing involves the lexicon lookup. The GREgORI project develops its own dictionaries for each processed language. The dictionaries are Unicode text files listing the word-forms. The Greek lexicon records 441,464 word-forms connected with 67,347 lemmata and, to be used with Unitex, is converted into DELAF format [Paumier, 2016:43-46]. Each word-form is provided with a one-line lexical entry using a precise syntax: inflected form, lemma corresponding to the form, part-of-speech tag of the lemma (noun, verb, adjective, adverb, pronoun, etc.) and inflectional tags describing the inflected-form (case, gender, number, voice, tense, mood, person, etc.). For instance :  $\tau \circ$ ,  $\delta$ .DET:Nns:Ans;  $\delta a \phi \circ \varsigma$ ,  $\delta a \phi \circ \varsigma$ .N+Com:Nns:Vns:Ans, etc. The DELAF formalism is also used for the other languages, such as Armenian npnpp,npnp.N+Com, Syriac  $\prec i \circ$ ,  $\prec i \circ$ .NOUN or Georgian  $\frac{1}{2}\delta \sigma_{130}, \frac{1}{2}\delta \sigma_{130}$ .N+Com, these three lemmata designating "the son".

It is important to precise that the lexical lookup does not take into account the context in which the words appear. Therefore, when a word corresponds to more than one lemma, all the possibilities are memorized by the process, such as, for example, the inflected form  $\varphi \delta \sigma \varepsilon_1$ , corresponding to both the verb  $\varphi \delta \omega$  "engender" and the noun  $\varphi \delta \sigma \varsigma_2$  "nature". This is a case of lexical ambiguity (see sections 1.3 and 1.5, below).

### **1.2 Information retrieval**

After the preprocessing, the user can search the text for words or expressions by using the "Locate Pattern" menu [Paumier, 2016:84-89]. The results are displayed in the shape of "Key Word In Context" concordances. Queries may concern inflected forms, lemmata, part-of-speech tags but also the combination of these arguments, as shown in table 1 below.

Locate Pattern	Result
εἴχομεν	The concordance of the word-from εἴχομεν
(Thurs)	The concordance of all the inflected form of the verb
<ex@></ex@>	έχω
	The concordance of all the inflected form of lemmata
< V >	tagged as verb
L Nee course	The concordance of all the lemmata tagged as negation
<1+INCg> <exw></exw>	followed by an inflected form of the verb $\xi_{\chi\omega}$

Table 1. Samples of queries.

Fig. 1 below gives a part of the concordance corresponding to the query  $\langle I+Neg \rangle \langle \tilde{\epsilon} \chi \omega \rangle$ . Note once again that, at this stage of the process, before lexical disambiguation, queries on both  $\langle \phi \dot{\omega} \omega \rangle$  and  $\langle \phi \dot{\omega} \sigma \iota \zeta \rangle$  provide concordances including the word-form  $\phi \dot{\upsilon} \sigma \iota$ .

[GrNa-29-9-26-25]	εἰ δόξειε μόνος τῶν πάντων Καισάριος	μή ἐσχηκέναι	φίλους, ὁ δὴ καὶ πολλοὺς ἔχειν
[GrNa-176-1-126-18]	ἐγκαλέσομεν, οὔτε ἀργίαν, οὔτε τὸ	μὴ ἔχειν	ὄ τι γράψεις·{S} οὐ πρὸς σοῦ τὸ ταῦτα
[GrNa-202-14-92-11]	ἐκεῖνον τὸν ἄνωθεν ἥκοντα τὸν νοῦν	μὴ ἔχειν,	ἀλλὰ τὴν θεότητα τοῦ Μονογενοῦς τὴν τοῦ
[GrNa-249-19-181-14]	ὑπεροχὴ τὰ χαυνότερα ἤθη διαφυσήσῃ,	ούκ είχον	ὅπως ἐμαυτὸν ἀτρεμεῖν νουθετήσω,
[GrNa-202-20-94-7]	τοῦ αὐτοῦ πράγματος ἀληθεῖς εἶναι φύσιν	οὐκ ἔχει.{S}	Πῶς οὖν ὑπέμεινέ σου ἡ μεγαλοφυὴς καὶ
[GrNa-170-1-123-5]	γραμμάτων ὑμῖν ὁμιλεῖν, ἐπειδή γε ἄλλως	ούκ ἔχομεν,	οὕτως οἰκονομήσαντος {τοῦ,ὁ.DET}
[GrNa-79-7-70-13]	καὶ τὴν πνευματικὴν ἐπιστασίαν	οὐκ ἔχοντος;{S}	Ἐκεῖνο δέ σοι καὶ πρὸ πάντων δῆλον, ὅτι
[GrNa-96-1-79-22]	Ύπατίῳ Πολὺν ἐζημιώμεθα χρόνον,	οὐκ ἐχούσης	τὸν πρῶτον ἐν ἀνδράσι τῆς πρώτης ἐν
[GrNa-82-2-72-17]	μαχομένοις αὐτοῖς ἦ τι πλέον, τὸ ἇθλον	ούκ ἔχουσι	τοῦ παλαίσματος.{S} Τῷ αὐτῷ Ἐπαινῶ ὅτι
[GrNa-31-7-28-13]	λέγω·{S} τὸ δὲ ἐζημιώθην·{S}	ούκ ἔχω	οἵου σὺ χρήζεις, τὴν Ἰλιάδα.{S} Μὴ γάρ μοι

Figure 1. Concordance of the lemmata tagged as negation and followed by an inflected form of the verb ἔχω.

### **1.3 Text Automaton**

Unitex provides a visual representation of each sentence of the processed text, called "automaton" [Paumier, 2106:161-198], as shown on fig. 2. It begins with an arrow and ends with a square inscribed in a circle. Each word form of the sentence appears in a box. Links between boxes symbolize the continuity of the sentence. Each word-form is accompanied with the corresponding lemma and part-of-speech tag, information coming from the lexical lookup.



Figure 2. Automaton of the sentence οὐ γὰρ ἔχω τι τούτων εἰπεῖν οἰκειότερον. (Letter 1, 1).

When a word corresponds to more than one lemma, two or more boxes appear on the same horizontal axis, making different interpretations of words explicit. The automaton in fig. 3 shows the lexical ambiguity of the word-form  $\pi \acute{\epsilon} \tau \rho \alpha \varsigma$ , inflected form of  $\Pi \acute{\epsilon} \tau \rho \alpha$  "Petra" (the city of Petra), or of  $\pi \acute{\epsilon} \tau \rho \alpha$  "stone".



Figure 3. Automaton of the sentence Ποῖον ἐκ πέτρας ὕδωρ τοῖς ἐν ἐρήμῷ πλημμυροῦν; (Letter 44, 2).

The processed text is fully lemmatized, without lexical ambiguities, when the automaton shows only one box for each word of the sentences of the text, i.e. only one interpretation. At this stage, it is possible to generate valid statistics and to edit lemmatized concordances or other outputs. Lexical ambiguities are processed automatically or manually. In the second case, the user can use the lemmatization interface of Unitex, as shown in section 1.5 below.

### 1.4 Information retrieval by grammars

Grammars are powerful tools able to describe sequences of words. The query  $\langle I+Neg \rangle \langle \tilde{\epsilon} \chi \omega \rangle$  used above can be graphically represented in the shape of a grammar, as shown by fig. 4. The graphical interface of Unitex allows users to draw grammars by themselves.



Figure 4. Grammar corresponding to the query  $\langle I+Neg \rangle \langle \tilde{\epsilon} \chi \omega \rangle$ .

By applying this grammar to the text with Unitex, the user will be provided, once again, with the concordance of fig. 1. However, the grammars allow to search more complex patterns corresponding to a wide range of possible searches.



Figure 5. Grammar predicting the optional presence of an adverb or a particle between the arguments <I+Neg> and < $\epsilon\chi\omega$ >.

The grammar of fig. 5 above will furnish the concordance of all the sequences constituted by a negation followed by the inflected forms of the verb  $\xi_{\chi\omega}$ , just like the previous one, but predicts, between these two arguments, the optional presence (the sign  $\langle E \rangle$  allows to recognize an empty sequence) of one or several adverbs or particles.



Figure 6. Grammar providing outputs and alphabetical sorting of inflected forms of the last verbal argument.

The grammars can also provide the outputs in the shape of the tags included in the results (fig. 6 above). These tags follow or surround the sequences displayed by the concordance. This grammar provides the same concordance as above, but inflected forms of  $\xi\chi\omega$  are classified according to the alphabetical order (due to the use of the asterisk before  $\langle\xi\chi\omega\rangle$ ). Here, the use of variables allows to repeat in the output the part of the sequence inscribed between the brackets, as shown on the concordance provided below (fig. 7).

[GrNa-249-19-181-14]	χαυνότερα ἤθη διαφυσήσῃ,	οὐκ εἶχον <Εχρι=οὐκ+ἔχω>	ὅπως ἐμαυτὸν ἀτρεμεῖν
[GrNa-29-9-26-25]	τῶν πάντων Καισάριος	μὴ ἐσχηκέναι <Εχρι=μὴ+ἔχω>	φίλους, ὁ δὴ καἰ πολλοὺς
[GrNa-81-1-72-4]	Μηδαμῶς, ὧ θαυμάσιε,	μὴ οὕτως ἔχε <Εχρι=μὴ οὕτως+ἔχω>.{S}	Τῶν μὲν γὰρ ἀκούσιος ἡ
[GrNa-202-20-94-7]	ἀληθεῖς εἶναι φύσιν	οὐκ ἔχει <Εχρι=οὐκ+ἔχω>.{S}	Πῶς οὖν ὑπέμεινέ σου ἡ
[GrNa-162-3-117-13]	Νῦν δὲ δέδοικα μὴ	οὐχ οὕτως ἔχει <expr=οὐχ οὕτως+ἔχω="">·{S}</expr=οὐχ>	μάλιστα μὲν γὰρ οὐδὲ μεμέ
[GrNa-176-1-126-18]	οὔτε ἀργίαν, οὔτε τὸ	μὴ ἔχειν <Εχρι=μὴ+ἔχω>	ὄ τι γράψεις·{S} οὐ πρὸς
[GrNa-202-14-92-11]	ἄνωθεν ἥκοντα τὸν νοῦν	μὴ ἔχειν <Εχρι=μὴ+ἔχω>,	ἀλλὰ τὴν θεότητα τοῦ
[GrNa-170-1-123-5]	όμιλεῖν, ἐπειδή γε ἄλλως	οὐκ ἔχομεν <Εχρι=οὐκ+ἔχω>,	οὕτως οἰκονομήσαντος
[GrNa-215-4-155-7]	ὑψηλότερον.{S} Ἡμεῖς δὲ	οὐχ οὕτως ἔχομεν <Εχρr=οὐχ οὕτως+ἔχω>,	ἀλλ΄ ὅτι τοῦ ἑνὸς διημάρτο
[GrNa-79-7-70-13]	πνευματικὴν ἐπιστασίαν	οὐκ ἔχοντος <Εχρι=οὐκ+ἔχω>;{S}	Ἐκεῖνο δέ σοι καὶ πρὸ

Figure 7. Concordance with outputs and inflected forms of  $\xi_{\chi\omega}$  in alphabetical order.

In fact, the preprocessing described above (section 1.1) works with this kind of grammar to analyze crasis (fig. 8 and 9) and expressions (fig. 10). In that case, the sequences are fully replaced by the content of the grammar.



Figure 8. Grammar for the crasis κἀγώ.



{@τοὔδαφος@,2.K} {τό,ὁ.DET:Nns:Ans} {ἔδαφος,ἔδαφος.N+Com}

Figure 9. Grammar for the crasis τοὕδαφος.



Figure 10. Grammar for the expressions ὡς οἶμαι and ὡς φησι.

Since Unitex uses the Unicode encoding and is language independent, as written above, this kind of grammars can be adapted to analyse similar linguistic facts in Armenian or Georgian. As regards Armenian, a case study is provided by [Van Elverdinghe, 2017], in this issue.

### **1.5 Lemmatization**

Unitex provides a lemmatization interface (fig. 11) allowing to resolve lexical ambiguities manually. A special window displays the concordance and the text automaton of sentences containing lexical ambiguities.

S Unitex/GramLab IDE 3.1 Rev. 4314 (March 2	7, 2016) - current language is GREgORI_GRC	- a ×
Text DELA +Sgraph Lexicon-Grammar Kalign i	ne Egittou Wandows Helb lato	
Select elements to lemmatize	Lemma selection	Navigation
Pattern: <φύω> GO	φύσις.N+Com	ematches     ended     ended
Show matching sequences in context Context lengths: Sort according to: Left: 40 chars Center, Right 💌	Lemmatize selected item	Sentence # 327
Right:         55         chars         Show all elements           Limit:         matches         Show unresolved only	Apply lemma to all items	Remove greyed states Apply modifications to TFST
[GrNa-11-1-13-14 ] τηρ [GrNa-249-16-180-17 ] [GrNa-29-5-26-11 ] πρ [GrNa-249-16-180-22 ] [GrNa-102-28-82-16 ] τ [GrNa-101-19-44-6 ] ἐτ	ίαν.{S} Γρηγορίῳ Ἐχω τι δεξ γὰρ οὐδὲν ἀνθρώπους ὄντα οσθήσεις τιμῶν αὐτὸν ἐπιτάα ν τῷ διὰ σαρκὸς ἐπιδημῆσα ος;{S} Ἐπειτα κατηγοροῦσιν πηγγελμένης τοῖς ὀρθῶς πισ	όν ἐν τῆ φύσει (μεγαλαυχήσομαι γὰρ ἔν τι τῶν πολλῶν καὶ αὐτός)·{ ς ἐν ὁμοτίμῳ τῆ φύσει μηδἐν ἐθέλειν πλέον ἀλλήλων ἔχειν, ἀλλὰ τῆ ταπειν να, καὶ ὅ φύσει πἀντες ἔχομεν ἄνθρωποι μᾶλλόν τι τῷ ἀπελθόντι χαρ τῆ ἀνθρωπίνῃ φύσει, ἀλλὰ καὶ τὸν προδότην ἑαυτοῦ Ἰούδαν τῷ ἰδίῳ στόμ ἡμῶν ὡς δύο φύσεις εἰσαγόντων ἀπηρτημένας ἢ μαχομένας καὶ μεριζόντω τεύουσι.{S} Φύσεις μὲν γὰρ δύο Θεὸς καὶ ἄνθρωπος, ἐπεὶ καὶ ψυχὴ καὶ
πμῶν         αὐτόν           πμή         PRO+Dem           Ν+Com         πμῶν           πμῶν         V	έππάφια έππάφιος	

Figure 11. Lemmatization interface of Unitex.

By clicking on superfluous boxes – for instance, boxes with lemma " $\dot{o}$ .DET" and lemma " $\phi\dot{\omega}$ .V", the user can remove inadequate interpretations and solve ambiguities in order to obtain a fully lemmatized and disambiguated text, as shown in fig. 12.



Figure 12. Lemmatization interface with elimination of useless boxes.

All linguistic resources, dictionaries and grammars, can be created, updated and corrected by the users, without any special knowledge of computer programming. The linguistic data is not encapsulated in a "black box" and remains accessible and readable.

At the end of the process, all lemmatized and disambiguated data are exported and gathered in a relational database system. Then, the publishing tools (see section III) allow to edit the special outputs in PDF format, such as concordances, indexes and other lexicographical lists.

### II. TEXT ALIGMENT

The GREgORI project processes texts in Greek, Arabic, Syriac, Armenian and Georgian using appropriate resources for each language. In this context, we are interested in comparing the lexical data of a *source text* with its ancient or modern translations, the *target text*, in order to provide scholars with corpus-based information about translation methods. This task requires alignment software such as mkAlign [Fleury, 2012 and http://www.tal.univ-paris3.fr/mkAlign/]. The mkAlign software allows users to pair words or expressions of a source text with words or expressions of a target text.

The mkAlign interface is divided into two columns. The user loads the source text in the left column and the target text in the right one. The first alignment is automatically provided on the basis of the punctuation or textual references of the edition (chapters, paragraphs, etc.). The sample provided here concerns the Greek text of the Homily 13 by Gregory of Nazianzus, aligned with its Syriac (S2) version [Schmidt, 2002]. Then, the user can move the words or expressions to align the relevant translation units. Word by word alignment is not always possible, as shown on fig. 13, below, where the Greek form  $\alpha i \rho \dot{\rho} \mu \epsilon \nu \alpha i$  is aligned with three Syriac words. Aligned data are afterwards exported as bitext files in the Translation Memory eXchange format (.tmx) and gathered in the relational database system. The bitext alignment is currently done manually. In the near future, the bilingual data will be used as translation memories in order to automatize the alignment for recurring translation units.

mkAlign 2.00 (2.0b156)		- a ×
Home Param Align 1	Hap Graphe Dic Specif Coocs Liste Segment Concordance Variation Export-XML Ex	cport Export-L3 Rapport
HODE GENERAL : Chargement Source et Câle	<b>ίερέως_{186-0}</b>	באיד 187-0 גבאיד 187-0
Prétraitement Prease Paragraphe MODE ALIGNEMENT AUTO. : Alignement Source/Clible (Recherche de cognats)	αἰρόμεναι_{191-0}	ג 191-0_בי 0-192_סאוד שר 192-0_סאוד שר 192-0_סא ייער 195_מאוד שר 191-0 מייער 192-0 ייער ג ג ג ג ג ג ג ג ג ג ג ג ג ג ג ג ג ג ג
HODE Import/Export THX	έπi_{188-0}	<u></u> 189-0
Recherche Source/Cible	Τοῦ_{189-0} ὄρους_{190-0}	
Aides regexp	καὶ_{192-0} εἰς_{193-0} εὐχὴν_{194-0}	[سرگام] ملي 193-0 روز الماني الماني ملي 193-0
	34 / 140	0 699 <b>0 ∢∢ ∢ 34 </b>

Figure 13. The mkAlign interface.

# **III. CONCORDANCES, INDEXES AND LEXICOGRAPHICAL LISTS**

Specific tools are developed by the Centre de traitement automatique du langage (Cental) at the UCL (https://uclouvain.be/fr/instituts-recherche/ilc/cental) in order to edit concordances, indexes and other types of lexicographical lists. The purpose of these developments is to extract

the lexical data out of the databases and inject them in the frameworks corresponding to the output required by the user.

Fig. 14 shows the lemmatized concordance of the verbs included in the letters of Gregory of Nazianzus. In this PDF document, lemmata are displayed in green (accompanied with a frequency rate), POS tags in red and word-forms in blue, with downstream and upstream contexts. The references to the edition appear on the left of the downstream context. The bookmarks list lemmata and allow browsing the document.

Fig. 15 gives a sample of the alphabetical list of these verbs and Fig. 16 lists these verbs according to their frequency in the corpus.

Fig. 17 offers a sample of a bilingual Greek-Syriac concordance of Homily 13 by Gregory of Nazianzus. This document displays all the lexical data both for the source text and for the target text, including roots (in red) for Syriac. Such a bilingual concordance is the so called "scholar concordance". Fig. 18 provides the lexical data (lemmata and POS tags) only for the source text. Such a bilingual concordance is a so-called "light concordance", being more easily readable for the user.

<ul> <li>The state of the s</li></ul>	× * 🔩	re	<	
<ul> <li>A local chance have a service in a service of the ser</li></ul>		Commentai	äyaµaı {V	ν να καί ναν καί δ6 δ6 βθείας, βθείας, καί περί καί περί έν ούν καί περί γ, ούτω σοφοί, ή
Dipdic - Adobe Acrobat Poo         fiftingse Fergite Ajde         200       Imata :		Outils	Corpus Epistularum Patrum Cappadocum - Bastien Kindt - Véronique Somers	καί κύφραίκεν μάλατα πίφωεν. Έγω δε διδαφρόνως τών αών διφωά πε καί άσπάζομαι, πούτο έρώ όπ τής τού καρού δυστολίας ερείπονα σταυτά καί κύφραίκεν μάλατα πίφωεν. Έγω δε διδαφρόνως τών αύον διφωία τής υπομονής λογαιμούς δυπτιπούσης. Ότα καί μάλατα τον θάκοι κ τό πότο το πόχι πόλου της πόττως. Τς θε όκό το πύους φόσσιση τής ποθούσκως, οι σοφός αύτοι τόν συνηθοτημένων έτμέν έντοι πόρι μάγου θα μός και η λύπη, καί μάλατα όταν τό δικαλος διατοστολου σύπως έκαργιάς τον έφρώκαν τήμο διαγοράσνα πόλεμον. λόρ το πορόλομαν, πότο ότων πρότηρους λωλ ότη φιλοσσφούμεν, όγαταστολου τόπως έκαργιάς τον τόρα το πόρι σύρο πουν πόρι πόρο συνηθοτημένων τημό κατόν. Ο Εινόμας, όκαι όγαται ότα το δικάλο διατό το πορόλο διατολιάς κρείπον από το πορόλωμαν. προί τοι κατήστο ποις διάτον όρατας όταν τό δικαλιας φαταστολούτας. Δος επτικό πού δινατολιάς της άπορού δυστολιάς το φρόρ ποι διά διαδολη το το πορότη το το το κράτης καί ποτολούτας. Δος επτικό, πού το διοδι το πόρός στο πόρο διοστολιάς ερομαίο συν παρόλωμαν. προί τοι πός ήδικησόστα πού όρατη ξια το άπο το δικαλικό το παι τό διοδι της πολισμότικον. Το αύτο αποι σύλομας το το πορδωμάτων προίτος το το σύρους, οικάτη όρατη διατό το που διό το δια τός ποιδιούστας διά απότε σύλομας. Το διά τρισμού κατόν ο Εινόμας, οικάτη όρατη διατο κών τα πόρ όδη της πολισμότους. Το αύτο από το πορδωλιατικά το το πορδου το διατησύους σύκαι όρατης το διαπολούστας διά πός που διοδιατης τής άματού συν παιο σύλομας. Το διά της το πότον το διαστου, ώτα τό δια της τατό πόσο το δια της παιδιούστας. Τό διά τός ποτο το δια τότο το το τάτο πατό μόλομας το διατου, όνα τα τόπον, ότα τη διαστό το στο όρατας το διατότος σύστα το δια τός το στο ότας το πορδούτας. Τό από το παιο συλιστός το το το το διαστου, ών τα ήπον, ότα το όρατας το δια τός τατολου ποράρου στο το το το το το στο το τ
Construct         Construct <thconstruct< th=""> <thconstruct< th=""> <th< td=""><td>1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1</td><td>o4,5% ↓ H 🗄</td><td></td><td>ŭγαμαι (V) 3 Griua 10.12.12.44 Griua 10.12.12.44 Griua 10.2.24.24 Griua 10.2.24.24 Griua 781 681 (V) 6 Griua 781 681 (V) 6 Griua 785 6827 Griua 785 6827 Griua 785 6827 Griua 785 6827 Griua 782 583 15 Griua 782 583 15 Griua 782 583 15 Griua 273 258 15 Griua 274 274 172 15 Griua 274 172 15 Griua 274 172 15 Griua 274</td></th<></thconstruct<></thconstruct<>	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	o4,5% ↓ H 🗄		ŭγαμαι (V) 3 Griua 10.12.12.44 Griua 10.12.12.44 Griua 10.2.24.24 Griua 10.2.24.24 Griua 781 681 (V) 6 Griua 781 681 (V) 6 Griua 785 6827 Griua 785 6827 Griua 785 6827 Griua 785 6827 Griua 782 583 15 Griua 782 583 15 Griua 782 583 15 Griua 273 258 15 Griua 274 274 172 15 Griua 274 172 15 Griua 274 172 15 Griua 274
	pdf - Adobe Acrobat Pro ichage Fepêtre Ajde		Ŧ	nata:     nata:       nata:     μαι {V}       πτάω {V}     κλαρχέω {V}       κάζω {V}     κάζω {V}       λαρχέω {V}     κάζω {V}       νοτάω {V}     νοτάω {V}       νοτάω {V}     νοτάω {V}       νοτάω {V}     κάζω {V}       νοτάω {V}     κάζω {V}       νοτάω {V}     κάζω {V}       νοτόμαι {V}     νοτάμα {V}       νοτόμαι {V}     κάω {V}       νοτόμαι {V}     κάω {V}       δξάω {V}     γομάω {V}       κάω {V}     γομάω {V}

UTEGOTY OF INAZIALIZUS letters by Шe 5 corpus III IIIe /eros DI 5 orualice rig. 14. Lemmanzeu

		Corpus Epistularum Patrum Cappado	cum - Bastien Kindt - Véronique Somers			
άγαμαι {V}	m	άληθεύω {V}	άνακεράννυμι {V}	1 ave	εγείρω {V}	-
άγανακτέω {V}	9	άλιεύω {V}	άνακομίζω {V}	1 ávé	EAKW {V}	2
άγαπάω {V}	00	άλίσκομαι {V}	άνακόπτω {V}	3 áve	ερωτάω {V}	-
άγγέλλω {V}	-	άλληγορέω {V}	άνακύπτω {V}	1 ave	еиріаки {V}	ŝ
άγελαρχέω {V}	-	άλλοτριόω {V}	άναλαμβάνω {V}	5 ávé	έχω {V}	11
àyiả(w {V}	4	άμαρτάνω {V}	άναλίσκω {V}	1 ávr	ηβάσκω {V}	L
άγνοέω {V}	19	άμάω (θερίζω) {V}	άναλύω {V}	3 ávr	jκω {V}	-
άγοράζω {V}	-	àµɛ(βw {V}	ávaltévu {V}	6 ávê	θέω {V}	-
άγρεύω {V}	2	άμελέω {V}	άναμύγνυμι {V}	1 ávu	áw {V}	7
àyptaivw {V}	-	άμνημονέω {V}	άναμιμινήσκω {V}	3 áví	יוויר {ע}	5
ăyu {V}	29	àµứvω {V}	άνανεόομαι (-όω) {V}	1 áví	מנוווו {V}	8
άγωνιάω {V}	2	άμφιέννυμι {V}	άνανεύω {V}	1 áví	(V) wyo	1
άγωνίζομαι {V}	14	άμφισβητέω {V}	άναπαύω {V}	7 àve	ontaivu {V}	1
άγωνοθετέω {V}	-	άναβαίνω {V}	άναπείθω {V}	1 ávo	oiyu {V}	-
άδικέω {V}	34	άναβάλλω {V}	άναπλάσσω {V}	1 avo	{\} wjmtho	1
άδοξέω {V}	-	άναγιγνώσκω {V}	άναπληρόω {V}	5 ávr	teyeipu {V}	-
άείδω {V}	11	άναγκάζω {V}	άναπνέω {V}	4 avr	τεισέρχομαι {V}	-
άθυμέω {V}	-	άναγράφω {V}	άνάπτω {V}	4 ávr	τεισφέρω {\	1
αίδέομαι {V}	27	àváyu {V}	άνασείω {V}	1 ávr	τεξάγω {V}	1
aipéw {V}	2	άναδείκνυμι {V}	άναστατόω {V}	1 ávn	τεπιδείκνυμι {V}	1
aipu {V}	2	άναδέχομαι {V}	άνατείνω {V}	1 ávr	τεπιτίθημι {V}	1
αίσθάνομαι {V}	6	άναδιδάσκω {V}	άνατίθημι {V}	2 ávr	cepũ {V}	1
αίσχύνω {V}	22	άναδύομαι {V}	άνατλήναι {V}	1 ávr	téχω {V}	ß
aitéw {V}	25	άναζητέω {V}	άνατρέπω {V}	1 avr	τιβαίνω {V}	-
αίτιάομαι {V}	6	άναθάλπω {V}	άνατρέφω {V}	1 ávr	τιβλέπω {V}	1
άκηδιάω {V}		άναίνομαι {V}	άναφέρω {V}	2 ávr	αδίδωμι {V}	00
άκολουθέω {V}	e	άναιρέω {V}	άναφύω {V}	1 ávr	ειλαμβάνω {V}	2
άκούω {V}	32	άνακαθαίρω {V}	άναχέω {V}	1 ávr	τιλέγω {V}	9
άκροάομαι {V}	-	άνακαινίζω {V}	άναχωρέω {V}	3 ávr	ແξύω {V}	-
άλγέω {V}	15	άνακαλέω {V}	άναψύχω {V}	2 ávī	tισόω {V}	1
άλείφω {V}	2	άνακαλύπτω {V}	άνδρίζω {V}	3 ávr	τισταθμέω {V}	-
21-11-17		Liste 'Lemr	nes - Analyses'			Page 1 sur 13

Fig. 15. Alphabetial index of the verbs in the corpus of the letters by Gregory of Nazianzus

		Corpus Epistularum Patrum Cappadoc	um - Bastien Kindt - Véronique Somers			
eiµí {V} 46	62	εύρίσκω {V} 35	{V} wpiζw {V}	22	φθέγγομαι {V}	16
ἔχω {V} 28	86	άδικέω {V} 34	μιμνήσκω {V}	22	ppovéu {V}	16
λέγω {V} 21	19	πράσσω {V} 34	παρίστημι {V}	22	άλγέω {V}	15
γίγνομαι {V}	80	έπιστέλλω {V}	Xpń {V}	22	άπολαύω {V}	15
ποιέω {V}	26	άκούω {V}	θεραπεύω {V}	21	εύφραίνω {V}	15
οίδα {V}	59	κρίνω {V} 32	ύπολαμβάνω {V}	21	μέλλω {V}	15
δέω (δεήσω) {V}	28	θαυμάζω {V} 31	KIVÉW {V}	20	όμολογέω {V}	15
γράφω {V}	4	λυπέω {V} 31	κρατέω {V}	20	σιωπάω { <b>V</b> }	15
δίδωμι {V}	85	άξιόω {V} 30	ποθέω {V}	20	φεύγω {V}	15
òpáw {V}	78	ăyu {V} 29	φυλάσσω {V}	20	άγωνίζομαι {V}	14
πάσχω {V} 7.	73	ăpχω {V} 29	άγνοέω {V}	19	έάω {V}	14
βούλομαι {V}	88	χράομαι (-άω) {V} 29	άπαιτέω {V}	19	κελεύω {V}	14
πάρειμι (είμί) {V}	67	ήκω {V} 28	ἔρχομαι { <b>V</b> }	19	<pre>µévw {V}</pre>	14
πείθω {V}	65	παρακαλέω {V} 28	ζάω {V}	19	παρέχω {V}	14
δοκέω {V}	62	πιστεύω {V} 28	kápvu {V}	19	ດພຸ່ໄພ {V}	14
φημί {V}	60	αἰδέομαι {V}	λύω {V}	19	άπατάω {V}	13
τιμάω {V} 5	22	έθέλω {V} 27	εŭχομαι {V}	18	àpiŋµı {V}	13
τυγχάνω {V}	22	έpῶ {V} 27	{\} moint(\)	18	διαλέγω {V}	13
φέρω {V}	52	διδάσκω {V} 26	μανθάνω {V}	18	καταγιγνώσκω {V}	13
δέχομαι {V}	21	αίτέω {V} 25	BonBéw {V}	17	καταλείπω {V}	13
έπαινέω {V}	20	έγκαλέω {V} 25	δηλόω {V}	17	κατέχω {V}	13
δύναμαι {V}	47	σπουδάζω {V} 25	έλπίζω {\}	17	μέλω {V}	13
voµíζω {V} 4	47	χαίρω {V} 25	ήσσάομαι (-άω) {V}	17	μετέχω {V}	13
χαρίζομαι (-ω) {V}	47	βουλεύω {V} 24	נסמואו {ע}	17	μιμέομαι {V}	13
үгүνώσкω {V}	46	žоιка {V} 24	καταλύω {V}	17	όνομάζω {V}	13
φαίνω {V} 4	46	προσάγω {V} 24	καταξιόω {V}	17	ópµáw {V}	13
φιλοσοφέω {V}	46	καλέω {V} 23	τίθημι {V}	17	πληρόω {V}	13
λαμβάνω {V}	4	vướu {V} 23	φροντίζω {V}	17	προσαγορεύω {V}	13
θαρσέω {V}	43	συγχωρέω {V} 23	παίζω {V}	16	προστρέχω {V}	13
οῖομαι (-ω) {V}	42	αίσχύνω {V}	πρεσβεύω {V}	16	συγγιγνώσκω {V}	13
προστίθημι {V}	40	βλέπω {V} 22	πυνθάνομαι {V}	16	συνάγω {V}	13
21-11-17		Liste 'Lemm	es - Analyses'		Pa	age 1 sur 13

Fig. 16. Frequency index of the verbs in the corpus of the letters by Gregory of Nazianzus

Conc	c_gma_13_grc-s1_verb_scholar.pdf - Adobe Acrobat Pro				×
Fichier	· Edition Affichage Fenëtre Aide				×
P	Créer ←   🎒 🗒 🖶 💟   🏶 💬 🐶 🔊	1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1			**
۲	<ul> <li>▲</li> <li>1 / 28  </li> <li>▲</li> <li>▼ 76,7% ↓</li> <li>⊨</li> </ul>	æ		Outils	Commentaire
Ç	Signets		Gregory of Nazianzus. Oratio 13. GRC-SYC		<
Ŋ		<ol> <li>άδικέω {V}</li> </ol>			
Ŀ		<ol> <li>kai •</li></ol>	(1-hurd-1)		
U.	Lemmata :	1 1 kai âônor	covinevot 1 xui* dôncia (P-Pur-V)		
)	🖶 🕷 άδικέω {V}	1	וסב 1 מען 1 מנד 1808 מים בני באל 5 מען 1 מבד מלבהלעון 1 מבד מלבהלעון	0 8 ct • -	
R.	eur áteíðu {V} eur arpu {V}	GRNA,GRC,Or,13,2,35,853,C,8 GRNA,SYC,Or,13,S2,2,9,21	אסטאר און אראיט אין אראיט אין אראיט אין איט אין איטער איז אין	ينه التابع فتسالي	
	🖶 🕡 άλλοτριόω 🧧 V }	1 aciõo {V}			
	🗄 🕡 άνατείνω 🧜 γ	1 1 ğoate	1 ἀείδω (V)		
	🕂 🕡 άπαιτέω { V }	1	N عحم 1 عحمه 1 م	30.1	
	🖽 ἀποδίδωμι {V}	GRNA,GRC,Or,13,1,35,853,A,10 GRNA,SYC,Or,13,S2,1,5,16	ήμεξς δε την κατά δύναμιν είσούσομεν σήμερον. Τοιγορούν άσατε τῷ Κυρέφ άσμα καινόν δτι θαυμαστά έπούησεν. Έκσμαινόμεθα για το	مي مل etea. mus	
	🖶 🕕 άπολαμβάνω 🧧 V }				
	🕀 🕡 ἀπόλλυμι { V }	1 appa {V}			
	🕀 🕕 άπορρίπτω {V}	1 1 αἰρόμεναι	ναι 1 αῦρο (V)		
	🗄 🕕 αὐλίζομαι (-ω) { V }	1	iom ک ۲۰۰۵٬۰۰۰ دد ۲۵۰٬۰۰۰ ۲۰ ۲۰ ۲۰۰۰ ۲۰۰۰ ۲۰۰۰ ۲۰۰۰ ۲۰۰۰	ct · iod ·	
	🕀 🕕 βάλλω { V }	GRNA,GRC,Or.,13,2,35,853,B,8 GRNA,SYC,Or.,13,S2,2,7,5	קרון אינדער אין אין אינדער אין	ראמבוילא לא מ	
	🗉 🕕 γιγνώσκω {V}		,		
	Β Π δέχομαι { V }	1 ἀκούοο {V}			
	🖽 🕕 δέω (δεήσω) { V }	1 1 ήκουσας	1 άκούα (V)		
	🖶 🕕 διανοέομαι (-έω) { V }	1	1 and and 1 and and 1	am & we	
	🕀 🕕 διδάσκω {V}	GRUA,GRC,Or,13,4,35,856,B,6 GRUA,SYC,Or,13,S2,4,13,9	دەنچ بىلغار تەرقىقار تەرۋىۋەرەر دەنچ قە قەقىدىتارەن. "لىدەنەمىر دەن قىتەندەن كەت قىتا بايىق قىق تەكىقە، قابىرەمە قاتەداقلانە قېتاپ سالى خىر تەخى قىلىمەر بىلىما، يەنىك قىلىيەن يەنىلامىتە. يىكەنىتەن بىلىغان يەن يەنى بىلىما، يىكەنمىيە يەنىك	Lui cutrul la	
	🖶 🕕 διέρχομαι { V }	Ι αλλοτριοώ {V}			
	🖶 🖉 διίωσισι { V }		© UCL - GREGORI Project - 2017	Page 1 sur 28	>
F					



E A	onc_grna_13_grc-s1_verb_light.pdf - Adobe Acrobat Pro vier Edition Affichage Fenêtre Aide			× *
Ĩ	👌 Créer 🗸 📔 🖶 🖂 🛛 🎲 ラ 🐶 👪	2 ^		76-51
	● 1 / 17   IN &   = + 76,7% →   H		Outils	Commentaire
Ľ	Signets		Grepory of Nazianzus. Oratio 13, GRC-SYC	¢
		1 dôticên {V}		
	E Lemmata :	GRUM,GRC,Or,13,2,35,853,C,8 GRUM,SYC,Or,13,52,2,9,21	πόλεως τίμιον ίσμεν καρολήν γινώσκομεν άγιον καλούμεν και άδιωσύμενοι: μόνον έστα φιλότεκνος και πάσης προνούν Έκλλησίας. σάλαι 4 ω.ε.ο. «τάμα» πων γενολωτό νόσαι πουλα» ποι ωτά τάσικο μικαθάν τότι, μικων ση τότασο καιδιατίο τόται σό	
(I)	📲 άδικέω {V}		1	
Left.	αείδω { V }	1 aziõo {V}		
2	🐨 aĩpu {V}	GRNA,GRC,Or,13,1,35,853,A,10 GRNA,SYC,Or,13,S2,1,5,16	ήμεζ, δε τήν κατά δύναμιν είσοισομεν σήμερον. Τσιγαρούν δρατε το Κυρίο φομα καινόν- ότι θαυμαστά έτοιησεν. Εκυμαινόμεθα	
	📲 άκούω {V}		وم خد قدمه، مسم بدامد محتخديا، سا بم، منه بمرس ديمير، سنا بدد سمخر، حد له، عدسه التعديم المحتمالة مدارية، وليد بالمحتمار	
	🐨 άλλοτριόω {V}	1 appo {V}		
	🔐 άνατείνω {V}	GRNA,GRC,Or.,13,2,35,853,B,8	σχήματι. Τούτο γάρ Γοχνσαν ίερέως χέφος έπί τού δρους αίρόμεναι και είς εύχήν τυπούμεναι δ πολλαί μυριάδες ούκ Γοχνσαν. Ούτος τείχη	
	🔐 άπαιτέω {V}	GRNA,SYC,Or,13,S2,2,7,5	ڪسمينئ لے وداودلليه ويسيه يعنديه. سي حيد سولادين منده ديدهاه در وداوارندي سور ولولموله وداولهم سور شو وي وزده ه	
	🔐 ἀποδίδωμι {V}	1 ἀκούω {V}		
	🗗 άπολαμβάνω {V}	GRMA,GRC,Or,13,4,35,856,B,6	τοίς μέν ταπεινοίς τό βόδιον τοίς δέ ύγηλοϊς τό δύοκτητον. "Ηκουσας" τοῦ εἰπόντος Ότι δεῖ ήμᾶς διὰ πολλῶν θλίγεων είσελθείν εἰς τήν	
	μ ἀπόλλυμι { V }	GRMA,SYC,Or,13,S2,4,13,9	يديد عنكسة لسلم فخ يقح هعسمالات لسلم يم يتقح عقممة وعامسمالات عجناسمرا لخه يتمعنا بيبط لم تحد تصلوبك ميخكتك بيمل	
	🔐 άπορρίπτω {V}	1 άλλοτριόω {		
	🕡 αὐλίζομαι (-ω) -{ V }	GRNA,GRC,Or,13,3,35,856,A,10	τῷ σῷ μέγα τι έχαρίσω και σαυτόν τὰ μέγιστα έζημίωσας άλλοτριώσας σεαυτόν τής τοῦ Θεοῦ χώριτος.	
	📲 βάλλω { V }	GRNA,SYC,Or,13,S2,3,13,11	and orcell of that the same in and in act to act to act to a the set of the set of the set of the set of the set	
	🔐 γιγνώσκω {V}	1 Austria IV		
	υ δέχομαι { V }	T MULTING T	նառուսումու սուն մես են ուսուս Մնա մես մեստեսու համանում է Ծան 10 մես առուննեսու Մնումաստեննել և ունե Թոնենեսու ում առու	
	🔮 δέω (δεήσω) { V }	GRNA,SYC,Or,13,S2,3,11,6	יוסט אין איז איז איז אין איז אין איז	
	- διανοέομαι (-έω) { V }			
	🖑 διδάσκω {V}	1 ἀπαιτέω {V}		
		GRNA,GRC,Or,13,4,35,856,C,2 GRNA SYC:Or:13,52,4,15,21	έκζηται τό άσθενές ένίσχωε το ίσχυρόν φύλασσε. Τοσαύτην άχαιτοθμέν σε την έν τοίς ανευματικούς εύδοκμησιν δοην έπι τῶν κοσμικῶν	
	🔮 διέρχομαι { V }		سه تعديد خبر يعه وصيب سد. يعه وسيه به مده منه في سمه وحسب بره حمد وسيه مع الحسب علاقه منه وسيه وسيه وسيه	
			© UCL - GREGORI Project - 2017 Page 1 sur 17	>

# Fig. 18. Bilingual Greek-Syriac (S2) concordance of the 13th homily by Gregory of Nazianzus (light concordance)

### **CONCLUSION**

For many years, all these technical developments and linguistic data are used to provide scholars with lemmatized concordances. These concordances were formerly published by Brepols Publishers on microfiches in the *Thesaurus Patrum Graecorum* series (see [Coulie, 1996; Kindt, 2010]. This use of microfiches is however rightly characterized as a "doubtful aspect" since [Trapp, 2008:97]. Now concordances and others lexical outputs are directly available on the website of the GREgORI project.

Some collaborators made use of these computerized resources to build lemmatized indexes (*index nominum, index verborum*, etc.) included in their editions of ancient texts (see for instance [Schmidt, 2001], [Auwers, 2005], [Sanpeur, 2007], [Amato, 2009], [Auwers, 2011] and more recently [Calzolari, 2017]). The GREgORI project also takes part in international research programs, such as the *Syriac Galen Palimpsest* project led by the University of Manchester or the study undertaken at the University of Lausanne and devoted to the comparative analysis of the *Iliad* and its Byzantine translation contained in the manuscript Genavensis 44.

To date, a web-based interface allowing to consult lemmatized concordances on personal computers via Internet, is being prepared in collaboration with the Cental. A beta-version of this interface is used by Peeters Publishers (Leuven) in order to pave the path for a forthcoming digital version of the well-known *Corpus Scriptorum Christianorum Orientalium*, the series dealing with some of the most important ancient sources written in languages of the Christian Middle East.

### References

- Amato E. Severus Sophista Alexandrinus Progymnasmata quae exstant omnia [...] accedunt Callinici Petraei et Adriani Tyrii sophistarum testimonia et fragmenta necnon incerti auctoris ethopoeia nondum vulgata. Bibliotheca Scriptorum Graecorum Romanorum Teubneriana, Walter de Gruyter (Berlin, New York), 2009.
- Auwers J.-M. Concordance du Siracide (Grec II et Sacra Parallela), avec la collaboration d'Églantine Proksch-Strajtmann. Cahiers de la Revue Biblique, 58. Gabalda (Paris), 2005.
- Auwers J.-M. Procopii Gazaei Epitome in Canticum Canticorum, cum praefatione a Jean-Marie Auwers et Marie-Gabrielle Guérard curata. Corpus Christianorum. Series Graeca, 67. Brepols Publishers (Turnhout), 2011.
- Calzolari, V. Apocrypha Armeniaca. Acta Pauli et Theclae, Prodigia Theclae, Martyrium Pauli. Corpus Christianorum. Series Apocryphorum, 20. Brepols Publishers (Turnhout), 2017.
- Coulie B. La lemmatisation des textes grecs et byzantins: une approche particulière de la langue et des auteurs. *Byzantion*. 1996;66:35-54.
- Coulie B., Kindt B. and Pataridze T. Lemmatisation automatique des sources en géorgien ancien. Le Muséon. 2013;126:161-201.
- Fleury S. mkAlign, Manuel d'utilisation, Université Sorbonne Nouvelle Paris 3, juin 2012. [get PDF]
- Gallay P. and Jourjon M. Lettres théologiques. Sources Chrétiennes, 280. Les Éditions du Cerf (Paris), 1976.
- Gallay P. Saint Grégoire de Nazianze, Lettres. Collection des Universités de France. Les Belles Lettres (Paris), 1964-1967.
- Kindt B. Des concordances et du *Thesaurus Patrum Graecorum*: la ligne éditoriale du Projet de recherche en lexicologie grecque (1990-2010). Schmidt A. *Studia Nazianzenica* II. Corpus Christianorum. Series Graeca. 73. Corpus Nazianzenum, 24. Brepols Publishers (Turnhout), 2010:43-120.
- Kindt B. La lemmatisation des sources patristiques et byzantines au service d'une description lexicale du grec ancien. Les principes de formulation des lemmes du Dictionnaire Automatique Grec (D.A.G.). *Byzantion*. 2004;74:213-272. [get PDF]
- Liddell H.G., Scott R. and Jones H.S. A Greek-English Lexicon, 9th ed, Oxford University Press. Clarendon Press (Oxford), 1940, repr. 1977 with Barber E. Greek-English Lexicon. A Supplement, Oxford University Press. Clarendon Press (Oxford), 1968; see also Glare G. W. Greek-English Lexicon. Revised Supplement, Oxford University Press. Clarendon Press (Oxford), 1996.
- Pataridze T. and Kindt B. Text Alignment in Ancient Greek and Georgian: A Case-Study on the First Homily of Gregory of Nazianzus. *Journal of Data Mining and Digital Humanities*. 2017;Special Issue.
- Paumier S. Unitex 3.1 User Manual, Université Paris-Est Marne-la-Vallée, March 27, 2016. [get PDF]
- Sanspeur Cl. Sancti Gregorii Nazianzeni Opera. Versio armeniaca, IV. Oratio VI. Corpus Christianorum. Series Graeca, 61. Corpus Nazianzenum, 21. Brepols Publishers (Turnhout), 2007.

Schmidt A.B. Sancti Gregorii Nazianzeni Opera. Versio Syriaca. II. Orationes XIII, XLI. Corpus Christianorum. Series Graeca, 47 ; Corpus Naizanzenum, 15. Brepols Publishers (Turnhout, Leuven), 2002.

Schmidt Th. Basilii Minimi in Gregorii Nazianzeni Orationem xxxviii Commentarii. Corpus Christianorum. Series Graeca, 46. Corpus Nazianzenum. 13. Brepols Publishers (Turnhout, Leuven), 2001.

Trapp E. Lexicography and electronic textual resources. Jeffreys E., Haldon J.F., Cormack R. (eds) The Oxford Handbook of Byzantine Studies. Oxford University Press (Oxford), 2008:95-100.

- Tuerlinckx L. La lemmatisation de l'arabe non classique. Dister A., Fairon C. and Purnelle G. (eds) Le poids des mots. 7<sup>es</sup> Journées internationales d'Analyse statistique des Données Textuelles, 10-12 mars 2004, Louvain-la-Neuve. Vol 2. Cahiers du Cental, II. Presses Universitaires de Louvain (Louvain-la-Neuve), 2004:1069-1078. [get PDF]
- Van Elverdinghe E. Recurrent Pattern Modelling in a Corpus of Armenian Manuscript Colophons. *Journal of Data Mining and Digital Humanities*. 2017;Special Issue.