



HAL
open science

Text Alignment in Ancient Greek and Georgian: A Case-Study on the First Homily of Gregory of Nazianzus

Tamara Pataridze, Bastien Kindt

► To cite this version:

Tamara Pataridze, Bastien Kindt. 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*, 2018, Special Issue on Computer-Aided Processing of Intertextuality in Ancient Languages, 10.46298/jdmdh.4182 . hal-01294591v2

HAL Id: hal-01294591

<https://hal.science/hal-01294591v2>

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.

Text alignment in Ancient Greek and Georgian. A case-study on the first homily of Gregory of Nazianzus

Tamara Pataridze^{1*}, Bastien Kindt^{1**}

¹ Université catholique de Louvain (Louvain-la-Neuve, Belgium)

*tamar.pataridze@uclouvain.be

** bastien.kindt@uclouvain.be

Abstract

This paper discusses the word level alignment of lemmatised bitext consisting of the Oratio I of Gregory of Nazianzus in its Greek model and Georgian translation. This study shows how the direct and empirical observations offered by an aligned text enable an accurate analysis of techniques of translation and many philological parameters of the text.

Keywords

Ancient Greek; Ancient Georgian; bilingual dictionaries; bitext; corpus; lexical tagging, morphological tagging; text alignment; lemmatisation.

INTRODUCTION

The original Greek texts of the Homilies of Gregory of Nazianzus (329-390 BC; about this author, see [Coulie, 1995]) were from early times translated into the different languages of the Christian East [Coulie, 1994]. This paper offers some conclusions resulting from the analysis of word level alignment of a bitext composed by the Greek model (called ST “source text”) of the first homily of Gregory of Nazianzus, and its Georgian translation (called TT “target text”). This homily, entitled Εἰς τὸ ἅγιον Πάσχα καὶ εἰς τὴν βραδυτῆτα “On Easter and the delay” [CPG 3010], was written in 362 AD. The Georgian translation was made by Ephrem Mtsire (ეფრემ მცირე, *Ephrem Mcire*), also known as Ephrem the Lesser (11th century) [Doborjginidze, 2009:65-93]. This work paves the way for a broader analysis of GreekGeorgian translations, especially, but not exclusively, regarding bilingual lexical correspondences.

The study of the oriental versions of this homily was already initiated in the framework of the Nazianzos Project (see <http://nazianzos.fltr.ucl.ac.be>), which ensured publication of critical editions of Arabic [Tuerlinckx, 2001], Syriac [Haelewyck, 2011], and Georgian [Metreveli et al., 1998] versions of this text, followed by articles, analysing some aspects of their textual correspondence and translation techniques (for example in [Coulie, 2000]).

In this context, our goals are the following:

- Offering multilingual digital dictionaries (for simple words) and translation memory files (for multi-word expressions);
- Offering materials based on the empirical evidence rooted in corpus observations, in order to contribute to the study of the translation methods used by the authors of the Christian East.

To reach these goals, lemmatised corpora and text-alignment tools are used.

I Available linguistic data

1.1 Corpora

The lemmatised concordance of Gregory of Nazianzus' Greek texts is available through the *Thesaurus Sancti Gregorii Nazianzeni* published by [Mossay et al., 1990]. The computerised data of this thesaurus, recovered and updated for the needs of the GREgORI Project, has now been gathered in a corpus based on the Unicode encoding standard and on the TEI guidelines. On the Georgian side, the corpus consists of the Georgian translations of the thirteen homilies published in the *Corpus Nazianzenum*. Note that Gregory of Nazianzus' homilies have been translated from Greek into Georgian several times by different authors. In the case of this first homily, the most important translators are the above mentioned Ephrem Mtsire, as well as Euthymius the Hagiorite (ეფთვიძე მთაწმიდელი *Eptvime Mtaçmideli* “from the Holy Mountain”) (†1028) [about this author, see Kazhdan, 1991]. Here, for our first approach, we have deliberately chosen to use Ephrem's translation because of its literalness in comparison to Euthymius' free style of translation [Metreveli, 1998:XV].

The Georgian version of the first homily by Gregory of Nazianzus was published in [Metreveli, 1998:2-17] and has been lemmatised by the authors of this paper with the collaboration of Professor Bernard Coulie. Table 1 lists the frequencies of the words, the lemmata and the different word-forms attested in these texts.

	Words	Lemmata	Different word-forms
Homilies corpus (Greek: 46 texts)	220,579	9,684	36,543
Homilies corpus (Georgian: 13 texts)	138,751	—	29,313
Homily I (Greek)	777	340	454
Homily I (Georgian; Ephrem)	640	360	454
Homily I (Georgian; Euthymius)	763	—	471

Table 1. Number of words, lemmata and word-forms in Gregory's Homilies (Greek texts and Georgian versions).

1.2 Lexical and morphological tagging

Both ST and TT are lemmatised. Each word is tagged with lexical (i.e. lemma) and morphosyntactic (i.e. part-of-speech) information. The lemmatisation in Greek follows the rules described in [Kindt, 2004]; see the website of the GREgORI Project for the part-of-speech tagset. On the other hand, one can find the lemmatisation principles regarding Georgian texts in [Coulie et al., 2013]. Texts are processed by lexical look-up (with the electronic dictionaries of the GREgORI Project) followed by a step of automatic or manual disambiguation for words corresponding to more than one lemma in the dictionaries; in other words, each word of the corpus receives a single lemma corresponding to its use in the context in which it appears. Lexical look-up and disambiguation are made by using the NLP-software Unitex, described in [Paumier, 2016] (about the role of this software in the project, see the contribution of [Kindt, 2017] in the present issue).

1.3 Alignment

ST and TT are then aligned as shown on Figures 1 and 2. Each token of the texts is followed, enclosed between braces, by its lemma, by a part-of-speech tag and by a sequential identification number (linked in the data-base to the exact references of this token in the original text). Alignment is processed with the mkAlign software [Fleury, 2012]. A first alignment is done automatically; texts are segmented in “translation units” (TU) on the basis of the punctuation marks used as sentence boundaries (Figure 1).

TU	ST	TT
1		წმიდათა_ {წმიდაჲ.A.1-0}
2		შორის_ {შორის.I+Adv.2-0} მამისა_ {მამაჲ.N+Com.3-0} ჩუენისა_ {ჩუენი.PRO+Pos.4-0} გრიგოლი_ {გრიგოლ.N+Prop.5-0} ღმრთისმეტყუელისა_ {ღმრთისმეტყუელი.V+Part.6-0} სიტყუაჲ_ {სიტყუაჲ.N+Com.7-0} პასქაჲსათჳს_ {პასქაჲ@თჳს.N+Prop@I+Prep.8-0} და_ {და.I+Conj.9-0} დაყოვნებისათჳს_ {დაყოვნებაჲ@თჳს.N+Com@I+Prep.10-0} მამაო_ {მამაჲ.N+Com.11-0} გუაკურთხენ!_ {კურთხევა.V+Mas.12-0}
3	Ἀναστάσεως_ {ἀνάστασις.N+Com.1-0} ἡμέρα_ {ἡμέρα.N+Com.2-0} καὶ_ {καί.I+Part.3-0} ἡ_ {ὁ.DET.4-0} ἀρχὴ_ {ἀρχή.N+Com.5-0} δεξιᾶ_ {δεξιός.A.6-0} καὶ_ {καί.I+Part.7-0} λαμπρυνθῶμεν_ {λαμπρύνω.V.8-0} τῆ_ {ὁ.DET.9-0} πανηγύρει_ {πανηγυρίς.N+Com.10-0} καὶ_ {καί.I+Part.11-0} ἀλλήλους_ {ἀλλήλων.PRO+Rec.12-0} περιπτυξώμεθα_ {περιπτύσσω.V.13-0}	აღდგომისა_ {აღდგომაჲ.N+Com.13-0} დღე_ {დღე.N+Com.14-0} და_ {და.I+Conj.15-0} დასაბამი_ {დასაბამი.N+Com.16-0} მარჯუე_ {მარჯულე.A.17-0} და_ {და.I+Conj.18-0} განზრწყინდეთ_ {განზრწყინება.V+Mas.19-0} კრებასა_ {კრებაჲ.N+Com.20-0} და_ {და.I+Conj.21-0} ურთიერთას_ {ურთიერთას.PRO+Rec.22-0} "შევიტკზნეთ;"_ {შეტკზნობა.V+Mas.23-0}
4	εἶπωμεν_ {λέγω.V.14-0} ἀδελφοί_ {ἀδελφός.N+Com.15-0} καὶ_ {καί.I+Part.16-0} τοῖς_ {ὁ.DET.17-0} μισοῦσιν_ {μισέω.V.18-0} ἡμᾶς_ {ἡμεῖς.PRO+Per1p.19-0} μὴ_ {μὴ.I+Neg.20-0} ὄτι_ {ὄτι.I+Conj.21-0} τοῖς_ {ὁ.DET.22-0} δὲ_ {δέ.I+Prep.23-0} ἀγάπην_ {ἀγάπη.N+Com.24-0} τι_ {τις.PRO+Ind.25-0} πεποιηκόσιν_ {ποιέω.V.26-0} ἢ_ {ἢ (καί).I+Part.27-0} πεπονθόσιν_ {πάσχω.V.28-0}	ვჰრქუათ_ {რქუმა.V+Mas.24-0} მამომუღეთაცა_ {მამომუღე@ცა.V+Part@I+Part.25-0} ჩუენთა_ {ჩუენ.PRO+Pers.26-0} ნუ_ {ნუ.I+Part.27-0} ოდენ_ {ოდენ.I+Adv.28-0} სიყუარულისათჳს_ {სიყუარული@თჳს.N+Com@I+Prep.29-0} რასმე_ {რაჲ.PRO+Int.30-0} მოქმედთა_ {მოქმედი.V+Part.31-0} ანუ_ {ანუ.I+Conj.32-0} შემთხუეულთა_ {შემთხუეული.V+Part.33-0}
5	συγχωρήσωμεν_ {συγχωρέω.V.29-0} πάντα_ {πάντα.A.30-0} τῆ_ {ὁ.DET.31-0} ἀναστάσει_ {ἀνάστασις.N+Com.32-0}	შეუნდოთ_ {შენდობა.V+Mas.34-0} ყოველი_ {ყოველი.PRO+Ind.35-0} აღდგომასა_ {აღდგომაჲ.N+Com.36-0}
6	δῶμεν_ {δίδωμι.V.33-0} συγγνώμην_ {συγγνώμη.N+Com.34-0} ἀλλήλους_ {ἀλλήλων.PRO+Rec.35-0} ἐγώ_ {ἐγώ.PRO+Per1s.36-0} τε_ {τε.I+Part.37-0} ὁ_ {ὁ.DET.38-0} τυραννηθεὶς_ {τυραννέω.V.39-0} τὴν_ {ὁ.DET.40-0} καλὴν_ {καλός.A.41-0} τυραννίδα_ {τυραννίς.N+Com.42-0} τοῦτο_ {οὗτος.PRO+Dem.43-0} γὰρ_ {γάρ.I+Part.44-0} νῦν_ {νῦν.I+Adv.45-0} προσίθημι_ {προσίστημι.V.46-0} καὶ_ {καί.I+Part.47-0} ὑμεῖς_ {ὑμεῖς.PRO+Per2p.48-0} οἱ_ {ὁ.DET.49-0} καλῶς_ {καλός.I+Adv.50-0} τυραννήσαντες_ {τυραννέω.V.51-0} ...	ვსცეთ_ {ცემა.V+Mas.37-0} მიტევება_ {მიტევებაჲ.N+Com.39-0} ურთიერთას_ {ურთიერთას.PRO+Rec.38-0} მე_ {მე.PRO+Pers.40-0} მიმძლავრებულმან_ {მიმძლავრებული.V+Part.41-0} კეთილმძლავრებითა_ {კეთილმძლავრებაჲ.N+Com.42-0} ამას_ {οὗτος.PRO+Dem.44-0} რამეთუ_ {რამეთუ.I+Conj.43-0} აწ_ {აწ.I+Adv.45-0} შევსძინებ_ {შეძინება.V+Mas.46-0} და_ {და.I+Conj.47-0} თქუენ_ {თქუენ.PRO+Pers.48-0} კეთილმძლავრებულთა_ {კეთილმძლავრებული.V+Part.49-0} ...

Figure 1. First ('sentence') alignment.

A second alignment process is done manually in order to identify more specific “translation units”, as close as possible to the “lexical units” (Figure 2). This word-by-word alignment process will become increasingly automatized when all the resources, such as translational memories will be exploited.

TU	ST	TT
1		წმიდათა_ {წმიდაჲ.A.1-0}
2		შორის_ {შორის.I+Adv.2-0} მამისა_ {მამაჲ.N+Com.3-0} ზუენისა_ {ზუენი.PRO+Pos.4-0} გრიგოლი_ {გრიგოლ.N+Prop.5-0} ღმრთისმეტყუელისა_ {ღმრთისმეტყუელი.V+Part.6-0} სიტყუა_ {სიტყუა.N+Com.7-0} პასქასათვს_ {პასქა@თვს.N+Prop@I+Prep.8-0} და_ {და.I+Conj.9-0} დაყოვნებისათვს_ {დაყოვნება@თვს.N+Com@I+Prep.10-0} მამაჲ_ {მამაჲ.N+Com.11-0} გუაკურთხენ!_ {კურთხევა.V+Mas.12-0}
3	Ἀναστάσεως_ {ἀνάστασις.N+Com.1-0}	აღდგომისა_ {აღდგომაჲ.N+Com.13-0}
4	ἡμέρα_ {ἡμέρα.N+Com.2-0}	დღე_ {დღე.N+Com.14-0}
5	καί_ {καί.I+Part.3-0}	და_ {და.I+Conj.15-0}
6	ἡ_ {ὁ.DET.4-0}	
7	ἀρχή_ {ἀρχή.N+Com.5-0}	დასაზამი_ {დასაზამი.N+Com.16-0}
8	δεξιὰ_ {δεξιός.A.6-0}	მარჯუე_ {მარჯუე.A.17-0}
9	καί_ {καί.I+Part.7-0}	და_ {და.I+Conj.18-0}
10	λαμπρυνθῶμεν_ {λαμπρύνω.V.8-0}	განვზრწყინდეთ_ {განვზრწყინება.V+Mas.19-0}
11	τῆ_ {ὁ.DET.9-0}	
12	πανηγύρει_ {πανήγυρις.N+Com.10-0}	კრებასა_ {კრებაჲ.N+Com.20-0}
13	καί_ {καί.I+Part.11-0}	და_ {და.I+Conj.21-0}
14	ἀλλήλους_ {ἀλλήλων.PRO+Rec.12-0}	ურთიერთას_ {ურთიერთას.PRO+Rec.22-0}
15	περιπτωξώμεθα_ {περιπτύσσω.V.13-0}	"შევიტკბნეთ;"_ {შეტკბობა.V+Mas.23-0}
16	εἶπωμεν_ {λέγω.V.14-0}	ვჰრქუათ_ {რქუმა.V+Mas.24-0}
17	ἀδελφοί_ {ἀδελφός.N+Com.15-0} καί_ {καί.I+Part.16-0} τοῖς_ {ὁ.DET.17-0} μισοῦσιν_ {μισέω.V.18-0}	ძმამოძულეთაცა_ {ძმამოძულე@ცა.V+Part@I+Part.25-0}
	ἡμᾶς_ {ἡμεῖς.PRO+Pers.19-0}	ზუენთა_ {ზუენ.PRO+Pers.26-0}
18	μὴ_ {μή.I+Neg.20-0}	ნუ_ {ნუ.I+Part.27-0}
19	ὅτι_ {ὅτι.I+Conj.21-0}	ოღენ_ {ოღენ.I+Adv.28-0}
20	τοῖς_ {ὁ.DET.22-0} δι' _ {διά.I+Prep.23-0} ἀγάπην_ {ἀγάπη.N+Com.24-0}	სიყუარულისათვს_ {სიყუარული@თვს.N+Com@I+Prep.29-0}
21	τι_ {τις.PRO+Ind.25-0}	რასმე_ {რაჲ.PRO+Int.30-0}
22	πεποικόσιν_ {ποιέω.V.26-0}	მოქმედთა_ {მოქმედი.V+Part.31-0}
23	ἢ_ {ἢ (καί).I+Part.27-0}	ანუ_ {ანუ.I+Conj.32-0}
24	πεπονθόσι_ {πάσχω.V.28-0}	შემთხუელთა_ {შემთხუელი.V+Part.33-0}
25	συγχωρήσωμεν_ {συγχωρέω.V.29-0}	შეუნდოთ_ {შენდობა.V+Mas.34-0}

26	πάντα_ {πᾶς.A.30-0}	ყოველი_ {ყოველი.PRO+Ind.35-0}
27	τῆ_ {ὁ.DET.31-0}	
28	ἀναστάσει_ {ἀνάστασις.N+Com.32-0}	ადღვომასა_ {ადღვომაჲ.N+Com.36-0}
29	δῶμεν_ {δίδωμι.V.33-0}	ვსცეთ_ {ცემა.V+Mas.37-0}
30	συγγνώμην_ {συγγνώμη.N+Com.34-0}	მიტევება_ {მიტევებაჲ.N+Com.39-0}
31	ἀλλήλους_ {ἀλλήλων.PRO+Rec.35-0}	ურთიერთას_ {ურთიერთას.PRO+Rec.38-0}
32	ἐγώ_ {ἐγώ.PRO+Per1s.36-0}	მე_ {მე.PRO+Pers.40-0}
33	τε_ {τε.I+Part.37-0}	
34	ὁ_ {ὁ.DET.38-0}	
35	τυραννῆεις_ {τυραννέω.V.39-0}	მიმძლავრებულმან_ {მიმძლავრებული.V+Part.41-0}
36	τῆν_ {ὁ.DET.40-0}	
37	καλῆν_ {καλός.A.41-0} τυραννίδα_ {τυραννίς.N+Com.42-0}	კეთილმძლავრებითა_ {კეთილმძლავრებაჲ.N+Com.42-0}
38	τοὔτο_ {οὗτος.PRO+Dem.43-0}	ამას_ {იგი.PRO+Dem.44-0}
39	γάρ_ {γάρ.I+Part.44-0}	რამეთუ_ {რამეთუ.I+Conj.43-0}
40	νῦν_ {νῦν.I+Adv.45-0}	აწ_ {აწ.I+Adv.45-0}
41	προστίθῃμι_ {προστίθῃμι.V.46-0}	შევსძინებ_ {შეძინება.V+Mas.46-0}
42	καί_ {καί.I+Part.47-0}	და_ {და.I+Conj.47-0}
43	ὅμοιως_ {ὅμοιος.PRO+Per2p.48-0}	თქუენ_ {თქუენ.PRO+Pers.48-0}
44	οἱ_ {ὁ.DET.49-0}	
45	καλῶς_ {καλός.I+Adv.50-0} τυραννήσαντες_ {τυραννέω.V.51-0}	კეთილმძლავრებულთა_ {კეთილმძლავრებული.V+Part.49-0}
...

Figure 2. Second ('lexical') alignment.

The Georgian translator Ephrem the Minor belongs to the so-called “hellenophile” school. This literary trend adopts the principle of formally equivalent translation [Doborjginidze 2009:65-90], almost slavishly reproducing all the particularities of the source language, leading to the translation being positioned as close as possible to its model. Accordingly, the source and target sentences of this bitext enjoy a very similar structure, to the point that their respective segments may be delimited along the same boundaries. Therefore, the translation units frequently link one word of the ST with one word of the TT:

TU	ST	TT
4	ἡμέρα_ {ἡμέρα.N+Com.2-0}	დღე_ {დღე.N+Com.14-0}

Despite this, several other cases are also possible:

- two words from in ST are translated by one token in TT:

TU	ST	TT
37	καλὴν_{καλός.A.41-0} τυραννίδα_{τυραννίς.N+Com.42-0}	კეთილმძლავრებია_{კეთილმძლავრება.N+Com.42-0}

- a word from ST is omitted in TT:

TU	ST	TT
6	ἡ_{ὁ.DET.4-0}	

The article ἡ is omitted in TT because this part-of-speech does not exist in Georgian.

- the word order may be different between ST and TT:

TU	ST	TT
29	δῶμεν_{δίδωμι.V.33-0}	ვსცეთ_{ცემს.V+Mas.37-0}
30	συγγνώμην_{συγγνώμη.N+Com.34-0}	მიტევება_{მიტევება.N+Com.39-0}
31	ἀλλήλοισ_{ἀλλήλων.PRO+Rec.35-0}	ურთიერთას_{ურთიერთას.PRO+Rec.38-0}

Here, in TT, the translator did not respect the word order of the ST and altered the sequence of the noun and pronoun. This change explains the discrepancy between TU numbers and identification numbers.

The result of the alignment process is saved in a Translation Memory eXchange file (.tmx) called 'bitext'. This TMX format is extended from the XML format, as shown below:


```

<body>
<tu>
<tuv xml:lang="EL">
<seg> Ἀναστάσεως_{ἀνάστασις.N+Com.1-0}
</seg>
</tuv>
<tuv xml:lang="KA">
<seg> აღდგომისა_{აღდგომაჲ. N+Com.13-0}
</seg>
</tuv>
</tu>
<tu>
<tuv xml:lang="EL">
<seg> ἡμέρα_{ἡμέρα.N+Com.2-0}
</seg>
</tuv>
<tuv xml:lang="KA">
<seg> დღე_{დღი.N+Com.14-0}
</seg>
</tuv>
</tu>
<tu>
<tuv xml:lang="EL">
<seg> καὶ_{καί.I+Part.3-0}
</seg>
</tuv>
<tuv xml:lang="KA">
<seg> და_{და.I+Conj.15-0}
</seg>
</tuv>
</tu>
<tu>
<tuv xml:lang="EL">
<seg> ἡ_ {ὁ.DET.4-0}
</seg>
</tuv>
<tuv xml:lang="KA">
<seg> დასაბაბი_{დასაბაბი.N+Com.16-0}
</seg>
</tuv>
</tu>
<tu>
<tuv xml:lang="EL">
<seg> ἀρχή_{ἀρχή.N+Com.5-0}
</seg>
</tuv>
<tuv xml:lang="KA">
<seg> მარჯუე_{მარჯულები.A.17-0}
</seg>
</tuv>
</tu>
...

```

This bitext is then loaded into the database of the GREgORI Project and processed with specific software, allowing to edit the bilingual concordances and create the bilingual dictionary or ‘translation memory’. Figure 3, below, shows a bilingual Greek-Georgian concordance of the verbs ἀποδίδωμι and δίδωμι.

All this results in a need of the well annotated bitext. Morphosyntactic tagging is, in such case, an indispensable step to provide exhaustive information about each unit of bitext enabling, therefore, precise and well specified requests and accurate extraction of information.

In addition, one needs to take into consideration that for our purpose the identification well discriminated and equivalent units from ST and TT is essential, since the project GREgORI is conceived for philologists editing ancient texts and studying ancient translational technics. The ancient models are systematically translated many times by different translators using different translation technics: this makes the link between the ST and TT subtle and variable as the case might be. These subtleties must be accurately discriminated thanks to the morphosyntactic tagging, since this is the main purpose of the GREgORI project.

Summing up, the morphosyntactic annotation and alignment enable detection of related units within the bitext given that our study is aiming to the highest philological precision. These strategies are indispensable for supporting the accurate extraction of information when general context related to this bitext is marked by scarcity of the comparative studies, by lacking software tools for the ancient languages and, and by the usage of morphosyntactically quite distinct pair of languages.

II Case studies

2.1 Lexical equivalence

As noted before, the bitext offers a formally equivalent translation of a high degree of precision. This often leads to two consequences in TT: a very low level of terminological fluctuation and the creation of neologisms.

2.1.1 A very low level of terminological fluctuation

We generally observe a strict terminological correspondence between ST and TT. Usually, no fluctuation occurs when translating ST's terminology in TT, even for frequently used terms. For example, the occurrences of ἀνάπαυσις (N+Com) (1-2), used twice in ST, are only rendered by the word განსუენებაჲ (N+Com) in TT, despite synonyms being available for this lexical unit.

- (1) καὶ εἰς τὴν ἐκεῖθεν ἀνάπαυσιν [PG 35, col. 401A].
და მუნასა განსუენებასა
da munasa gansuenebasa
“du repos qui (nous attend) là-bas” [SC, p. 82, § 7, l. 22]
“unto the (heavenly) rest” [Shaff, 2007:204]
- (2) ἐπὶ ὕδατος ἀναπαύσεως ἐκτρεφόμενοι [PG 35, col. 400D]
ზრდილნი წყალთა ზედა განსუენებისათა
zrdilni c̣qalta zeda gansuenebisata
“nourris près des eaux du repos” [SC, p. 80, § 7, l. 7]
“being fed by water of refreshment” [Shaff, 2007:204]

The lemma μυστήριον (N+Com) (3-4), used three times in ST, is always translated by the same საიდუმლოჲ (N+Com) in TT.

- (3) **Μυστήριον** ἔχρισέ με, **μυστηρίω** μικρὸν ὑπεχώρησα [PG35, col. 396B]
საიდუმლომან მცხო მე, **საიდუმლოსა** მცირედ განვეშორე
saidumloman mcxo me, saidumlosa mcired ganvešore
 “j’ai reçu l’onction du mystère, j’ai manifesté un certain recul devant le mystère” [SC, p. 74, §2, l. 1]
 “a Mystery anointed me; I withdrew while at a Mystery” [Shaff, 2007:203]
- (4) **μυστηρίω** καὶ συνεισέρχομαι [PG35, Col. 396B]
საიდუმლოსაცა თანა შემოვალ
saidumlosaca tana šemoval
 “je reviens avec le mystère” [SC, p. 74, § 2, l. 2]
 “now I come in with a Mystery” [Shaff, 2007:203]

2.1.2 Creation of neologisms

The other consequence of the formally equivalent translation is the rise in TT of neologisms. Some of them, being constructed in the same way as their Greek models, are contrived, unnatural words in Georgian. They are reproducing their model slavishly, accurately reflecting Greek structures alien to genuine native usage. This is for example the case of *παρέρχομαι* (V), rendered by *თანაწარგვლმა* (V+Mas) in (5):

- (5) καὶ ἡμᾶς **παρήλθεν** ὁ ὀλοθρεῖων [PG35, col. 397A]
 და ჩვენ **თანაწარგვლდა** მომსრველი
da čuen tanačarguqda
 “l’exterminateur est passé à côté de nous” [SC, p. 74, § 3, l. 2]
 “and the (Destroyer) passed us over” [Shaff, 2007:203]

Or *ἀποδίδωμι* (V), translated through *უკუნცემა* (V+Mas) in (6):

- (6) **ἀποδῶμεν** τῇ εἰκόνι τὸ κατ’ εἰκόνα [PG35, col. 397B]
უკუნცეცეთ ხატსა ხატებაჲ
ukunvscet xaṭsa xaṭebaj
 “restituons à l’image ce qui est de l’image” [SC, p. 76-77, § 4, l. 9-10]
 “let us give back to the Image what is made after Image” [Shaff, 2007:203]

In the first example (5), the word *თანა-წარგვლდა* begins with the element *თანა-* “with”, which is used as a postposition in Ancient Georgian and never to build verbs. Its use as though it was a preverb, in combination with a conjugated form of a verb, is not natural. Instead, it is a slavish reproduction of the preverb *παρα* present in the Greek form *παρήλθεν*. Similarly, *უკუნ-ვცეცეთ*, in (6), is faithfully reproducing the structure of *ἀπο-δῶμεν* (see Coulie, 2000:255).

Consequently, lemmata corresponding to such lexical units (*თანა-ჯუარსვეცმოდე*, *თანა-ვიდიდები*, *თანა-დავევლვოდე* ect.), including analysed *უკუნცემა* *ukuncema* are not recorded in the dictionaries of Ancient Georgian — see (Abulaze, 1973:175-178). [Coulie, 2000:255] draws attention to the fact that these terms are paraphrased in other ways by the other translator, the above-mentioned Euthymius, who respects more closely the natural functioning of the Georgian language.

2.2 Lexical variations

2.2.1 Terminological fluctuation

In spite of the principle of a formally equivalent translation, it must be taken into account that any translation shows at least a few instances of terminological fluctuation. For example, αιδέσιμος “respectable, venerable”, occurring twice in ST, has been translated in two different ways in TT. Its first rendering, in (7), is the adjective განსაკრთომელ *gansakrtomel* “fearful, frightening”, while its second instance, in (8), is the locution ღირს პატივთაჲ *girs pativtaj* “worthy of respect”.

- (7) ἡ σφραγὶς φοβερὰ καὶ αἰδέσιμος [PG35, col. 397A]
 და ბეჭედი საშინელ და განსაკრთომელ
da bechedi saşinel da gansakrtomel
 “le sceau lui a inspiré crainte et respect” [SC, p. 74, § 3, l. 3]
 “the Seal was dreadful and reverend” [Shaff, 2007:203]
- (8) ἡ τιμία κεφαλὴ καὶ αἰδέσιμος [PG35, col. 400C]
 პატიოსანი თავი და ღირს პატივთაჲ
patiosani tavi da girs pativtaj
 “cette tête précieuse et respectable” [SC, p. 80, § 7, l. 2]
 “this honourable and reverend Head” [Shaff, 2007:204]

2.2.2 Failure of word-by-word correspondence

A correspondence between the units of ST and TT does not necessarily imply that strictly one token on the one side is equivalent to another one on the other side. Translation equivalence is a relation between two units with the same meaning from both sides but, obviously, word-by-word correspondence is sometimes impossible to achieve, as in (9-10).

- (9) ἀδελφοὶ καὶ τοῖς μισοῦσιν [PG35, col. 396A]
 ძმამოდულებთაჲ
zmatmozuletaca
 “(appelons) frères ceux-là même qui nous haïssent” [SC, p. 72, § 1, l. 3]
 “(let us say) Brethren, even to those who hate us” [Shaff, 2007:203]
- (10) τὴν καλὴν τυραννίδα [PG35, col. 396A]
 კეთილმძლავრებითა
ketilmzlavrebita
 “de cette belle tyrannie” [SC, p. 72, § 1, l. 6]
 “for the noble tyranny” [Shaff, 2007:203]

In the case of (9), the best solution for identifying bitextual pairs in this case would be to use a more accurate lemmatisation on the Georgian side, indicating that the root ძმა-მოდულე is a composite root consisting of two different elements, such as noun and verb (participle): ძმა-*zma*- “brother” and -მოდულე *-mozule* “one who hates”. Similarly, in (10), კეთილ-მძლავრებითა brings together the elements კეთილ *ketil* “good” and მძლავრებითა *mzlavrebita* “in a coercive way”.

[ST] PRO+Pers + V vs V [TT]

A conjugated verb accompanied by a personal pronoun in ST often matches with a conjugated verb in TT, without a personal pronoun. In Georgian, pronouns are directly included in the verbal structure through an appropriate morphological mark. In (11), the morpheme მ- *m-* in მაბრალობთ *m-abralobt* is the equivalent of the Greek pronoun μοι. In other words, the

Greek personal pronoun will no longer be present in the Georgian translation, leading to the frequent asymmetry of this type.

- (11) **μοι** μέμφοισθε [PG35, col. 396A]
 მამრალობთ
mabralobt
 “au cas où vous auriez quelque reproche à me faire” [SC, p. 74, § 1, l. 9]
 “if you had cause to blame my tardiness” [Shaff, 2007:203]

2.3 Symmetrical equivalences

[ST] V = V+Mas [TT]
 [ST] N+Com = N+Com [TT]
 [ST] A = A [TT]
 [ST] PRO+Rec = V+Mas [TT]

Unsurprisingly, in most bitextual pairs, each term pertains to a similar morphological category in ST and in TT.

ἀκούω_{ἀκούω.V} vs ესმის_{ესმის.V+Mas}
 ἡμέρα_{ἡμέρα.N+Com} vs დღე_{დღე.N+Com}
 δεξιὰ_{δεξιός.A} vs მარჯუე_{მარჯუე.A}
 ἀλλήλους_{ἀλλήλων.PRO+Rec} vs ურთიერთს_{ურთიერთი.PRO+Rec}

2.4 Asymmetrical equivalences

[ST] A = V+Part [TT]

We emphasise below some asymmetrical equivalences from the Greek-Georgian bitext. The correspondence [ST] A = V+Part [TT] is frequently attested, and is justified on the linguistic level as well; a Georgian lexeme being morphologically a participle is often performing the function of an adjective qualifying the name in the sentence [Coulie et al., 2013:183-184]. Some grammars categorise such words as “verbal adjectives”. However, given that from a morphological point of view these units are clearly participles in Georgian, we opted to label them as “V+Part”, as seen in ἀγαθὸν_{ἀγαθός.A} vs კეთილად_{კეთილი.V+Part} in (12):

- (12) **πλάστην ἀγαθὸν** [PG35, col. 397A]
 მოქმედად კეთილად
mokmedad ketilad
 “un bon modeleur” (SC, p. 74, § 2, l. 6)
 “as a good modeler” [Shaff, 2007:203]

The word კეთ-ილ-ო shows the morphology of a past participle, based on a verbal root კეთებ-ა “doing, performing, realising”, with the morpheme -ილ- proper to past participles, and the nominative case ending -ო (participles can be declined in Georgian). Thus, კეთ-ილ-ო has a meaning of “done, performed”, but the unit is used as an adjective with the extensional meaning of “good, well done”.

[ST] N+Com = V+Part [TT]

[ST] I+Adv = V+Part [TT]
 [ST] I+Adv = A [TT]
 [ST] A = N+Com [TT]
 [ST] PRO+Ref1s = N+Com [TT]
 [ST] I+Part = I+Conj [TT]

Applied to the case of substantivised participles in Georgian, the asymmetric equivalence [ST] N+Com = V+Part [TT] makes sense. Concerned Georgian tokens exhibit a clearly participial morphology and despite their substantivised character, they are still nevertheless labelled as participles, such as κτήμα_{κτήμα.N+Com} vs მონაგები_{მოგება.V+Part} in (13) [Coulie et al., 2013:184]:

- (13) τὸ τιμιώτατον Θεῷ κτήμα καὶ οἰκειότατον [PG35, col. 397B]
 უპატიოსნესი და უთვსესი მონაგები ღმრთისა
upatiosnesi da utwsesi monagebi gmrtsisa
 “le plus précieux aux yeux de Dieu et le plus proche de lui” (SC, p. 76, § 4, l. 8)
 “the possession most precious to God, and most fitting” [Shaff, 2007:203]

The form მონაგები is a past participle from the verb მოგება *mogeba* “to earn, to get, to win something”. The word მონაგები *monagebi* means “something that was obtained, earned” leading to the meaning of “goods, properties”.

Adverbs in ST are widely affected by asymmetrical renderings in TT, and the following formulae are common:

[ST] I+Adv = V+Part [TT]
 [ST] I+Adv = A [TT]

Indeed, a considerable number of adverbs in Georgian is formed through inflecting *adjectives* and *participles* in the adverbial case. This relates in particular to the so-called *adverbs of manner* characterising the manner by which the action expressed by the verb is performed. They are considered as “derivative” adverbs, in contrast with the “primary” ones. We tagged as “adverbs” only “primary” forms, while “derivatives” — being adjectives or participles declined in the adverbial case — are merely considered as declined adjectives and participles, and are labelled as such, e.g. καθαρῶς_{καθαρῶς.I+Adv} vs წმიდად_{წმიდა.A} in (14) [Coulie et al., 2013:192-194]:

- (14) σήμερον καθαρῶς ἐφύγομεν Αἴγυπτον [PG35, col. 397A]
 დღეს წმიდად განვიღებენით ეგვიპტით
dges cmidad ganvilvenit egwptit
 “aujourd’hui nous avons échappé totalement à l’Égypte” (SC, p. 74-76, § 3, l. 4-5)
 “today we have clean escaped from Egypt” [Shaff, 2007:203]

The adjective წმიდა *cmidaj* “pure” has been put in the adverbial case (წმიდად *cmid-ad*), which enables to express the meaning of “purely”.

Similarly, the participle კეთილი *ketili* discussed above, once put in the adverbial case, will express the meaning of “well, nicely, pleasantly”, and its matched pair in the ST will be an adverb, such as καλῶς_{καλῶς.I+Adv} vs კეთილადა_{კეთილი.V+Part} in (15):

- (15) τὸ *καλῶς* ποιμαίνεσθαι [PG 35, col. 400C]
 კეთილად მწყსილებასა
ketilad mcqsilebasa
 “bien conduire au pâturage” [SC, p. 80, § 7, l. 7]
 “to nicely drive to the pasture” [translated by the authors]

It is thus generally true that an adverbial form of an adjective or participle in Georgian will correspond to an adverb in many other languages.

[ST] A=N+Com [TT]

Similarly, the genitive case of the common names in TT will quite often express the same meaning as the adjectives of ST:

λιθίναις_{λίθινος.A} vs ქვისათა_{ქვაო.N+Com}
 σαρκίνας_{σάρκινος.A} vs ჯორცთანი_{ჯორცი.N+Com}

[ST] PRO+Ref1s vs N+Com [TT]

The Georgian noun თავი *tavi* “head” is often used in the function of reflexive pronoun, a fact that will also generate a frequent asymmetrical equivalence, since the usage of a reflexive pronoun is common, such as in (16):

- (16) μυστηρίῳ μικρὸν ὑπεχώρησα ὅσον *ἐμαυτὸν* ἐπισκέψασθαι [PG 35, col. 396B]
 საიდუმლოსა მცირედ განვეშორე რაოდენ *თავის* განცდადმდე
saidumlosa mcired ganvešore raoden tavis gancdadmde
 “j’ai manifesté un recul devant le mystère, le temps de m’examiner” (SC, p. 74, § 2, l. 2)
 “I withdrew a little while at a Mystery, as much as was needful to examine myself” [Shaff, 2007:203]

Yet many other situations of asymmetry are frequent, especially with “functional” words that are differently classified in the grammar of the respective languages of ST and TT. For example, the widespread word *καί* is tagged as a particle in Greek while its Georgian equivalent, *და*, is considered a conjunction: *καί*_{*καί*.I+Part} vs *და*_{*და*.I+Conj}.

III Conclusion

The work on word-level alignment between the Greek and Georgian texts of the first homily of Gregory of Nazianzus was carried out in the framework of the GREgORI Project. This work complements the previous studies focused on the analysis of techniques of translation from Greek into the different languages of the Christian East. This method — being based on the direct and empirical observations offered by an aligned bitext — enables systematising those previous researches. Applying the same method to the twelve other already published Georgian homilies, a corpus of 138,741 words, is the next step. Of course, increasing bilingual data consisting of previously identified TU-s will allow for an ever increasing automatised alignment process. Other alignment strategies, such as statistical methods, will be tested before being applied to new texts. At the same time, the GREgORI Project is beginning to run the same methodology on the versions of the works of the Theologian translated into other languages of the Christian East.

References

- Abulaze I. *ძველი ქართული ენის ლექსიკონი* (Dictionary of Ancient Georgian Language), Mecniereba (Tbilisi), 1973.
- Coulie B. Les traductions géorgiennes du Discours 1 par Euthyme l'Hagiorite et Éphrem de la Montagne Noire, in Coulie B. *Studia Nazianzenica. I*, Brepols Publishers (Turnhout), 2000:245-258.
- Coulie B. Les versions orientales des œuvres de Grégoire de Nazianze in *Mémoire des saints Grégoire le Théologien et Photios le Grand, archevêques de Constantinople. Actes du symposium scientifique (14-17 oct. 1993) (in Greek) (Thessalonique)*, 1994:53-59.
- Coulie B. Gregor v. Nazianz d. J. *Lexikon für Theologie und Kirche*, Herder (Freiburg.Basel.Rom.Wien) 1995;4:1004-1007.
- Coulie B., Kindt B., Pataridze T. Lemmatisation automatique des sources en géorgien ancien. *Le Muséon*, 2013;126(1-2):161-201.
- CPG *Clavis Patrum Graecorum*, cura et studio Geerard M. (*Corpus Christianorum*), 5 vol., Brepols Publishers (Turnhout), 1974-1987; *Clavis Patrum Graecorum. Supplementum*, cura et studio Geerard M. et Noret J., ibid., 1998; *Clavis Patrum Graecorum*, Volumen III A. *Addenda uolumini III*, a Noret J., ibid., 2003.
- Doborjginidze N. *Die georgische Sprache im Mittelalter (Sprachen und Kulturen des christlichen Orients, 17)*, Reichert (Wiesbaden), 2009.
- Fleury S. *mkAlign. Manuel d'utilisation*, Juin 2012. <http://www.tal.univ-paris3.fr/mkAlign/mkAlignDOC.pdf>.
- GREgORI Project <http://www.uclouvain.be/gregori-project>.
- Haelewyck J.-C. *Sancti Gregorii Nazianzeni Opera. Versio syriaca, IV. Orationes, I, II, III (Corpus Christianorum, 77; Corpus Nazianzenum, 25)*, Brepols Publishers (Turnhout, Leuven), 2011.
- 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.
- Kindt B. Processing Tools for Greek and Other Languages of the Christian Middle East. *Journal of Data Mining and Digital Humanities*. 2017; Special Issue.
- Metreveli H. et alii *Sancti Gregorii Nazianzeni Opera. Versio iberica. I Orationes, I, XLV, XLIV, XLI (Corpus Christianorum. Series Graeca, 36; Corpus Nazianzenum, 5)*. Brepols Publishers (Turnhout), 1998.
- Mossay J. et alii, *Thesaurus Sancti Gregorii Nazianzeni*, vol. I. *Enumeratio lemmatum, Orationes, Epistulae, Testamentum (Corpus Christianorum. Thesaurus Patrum Graecorum)*, Brepols Publishers (Turnhout), 1990.
- Paumier S. Unitex 3.1rc. User Manual, Université Paris-Est Marne-la-Vallée, March 05, 2016 <http://igm.univ-mlv.fr/~unitex/UnitexManual3.1.pdf>.
- PG *Patrologia Graeca*.
- SC Bernardi, J. *Grégoire de Nazianze, Discours 1-3 (Collection des Sources Chrétiennes, 247)*, Éditions du cerf (Paris), 1978 (repr. 2006).
- Schaff Ph. and Henry Wace H., *A Select library of Nicene and post-Nicene fathers of the Christian church : Second series. The Christian Literature Publishing Company. Volume 7 (Buffalo)*, 1894 (repr. 2007).
- Schmidt A. *Sancti Gregorii Nazianzeni Opera. Versio syriaca, IV. Orationes I, III, V Corpus Christianorum. Series Graeca, 43. Corpus Nazianzenum, 10*, Brepols Publishers (Turnhout, Leuven), 2011.
- Tuerlinckx L. *Sancti Gregorii Nazianzeni opera. Versio Arabica antiqua, II. Orationes I, XLV, XLIV (Corpus Christianorum. Series Graeca, 43. Corpus Nazianzenum, 10)*, Brepols Publishers (Turnhout, Leuven), 2001.