Konferenca
Jezikovne tehnologije in digitalna humanistika
Ljubljana, 2022

Conference on
Language Technologies & Digital Humanities
Ljubljana, 2022

A Parallel Corpus of the New Testament:

Digital Philology and Teaching the Classical Languages in Croatia

Petra Matović,[*] Katarina Radić[†]

* Faculty of Humanities and Social Sciences, University of Zagreb
Ivana Lučića 3, 10000 Zagreb, Croatia
pmatovic@ffzg.hr

† Faculty of Humanities and Social Sciences, University of Zagreb
Ivana Lučića 3, 10000 Zagreb, Croatia
katarina.radic1@gmail.com

1. Introduction

Corpus linguistics has been one of the liveliest disciplines in Croatian linguistics, and parallel corpora have been established by Croatian scholars since the 1960s (Tadić 1997, 2001; Simeon 2002). These corpora normally include Croatian and another living language, while corpora consisting of texts in Croatian and at least one of the so-called "dead" languages are still underrepresented, although there are corpora including languages like Ancient Greek, Latin, Sanskrit, Arabic, Persian and Akkadian, to be found on the World Wide Web (The Alpheios Project 2019; Palladino et al., 2021). The Department of Classical Philology at the University of Zagreb can already boast one of the earliest online (monolingual) corpora of Latin texts, the CroALa database, built and curated by Neven Jovanović (CroALa, 2014). In the last few years the said department has been steadily building small parallel corpora, and this paper aims to describe one of them, the Greek-Croatian parallel corpus of the New Testament, currently in the making, and furthermore discuss its educational uses in teaching Ancient Greek.

2. Goal of the paper

Building parallel copora has been garnering more and more attention in the field of classical philology. The Department of Classical Philology at the University of Zagreb has been building smaller corpora, both as a part of several small-scale projects lead by Neven Jovanović and courses on Greek and Latin language (e.g. Soldo and Šoštarić 2019). Since 2021, several professors and students at the department have been working on project titled "A Linguistic Analysis of Selected Early Christian Writings", lead by Petra Matović. Within the scope of the project we have started building a parallel corpus of the New Testament, so far comprising the Gospel of Mark and a part of the Apocalypse. The texts are aligned using the Alpheios tool for text alignment at the Perseids environment (The Alpheios Project, 2019; The Perseids Project, 2017). Alpheios enables the user to align words or word combinations in the source text with corresponding parts of its translation (The Alpheios Project, 2019). In this poster we firstly aim to explain the principles of alignment we followed while building the corpus, and, secondly, discuss some peculiarities in aligning Ancient Greek with Croatian. Finally we aim to look at the corpus from an educational point of view and discuss its possible uses in teaching Ancient Greek today.

Text aligment was done by 4 students (Mateo Cader, Ružarijo Lukas, Katarina Radić, Luka Šop) and supervised by Petra Matović. The editions of the texts were Nestle-Aland 28 (Greek New Testament) and the so-called Zagreb Bible (https://biblija.ks.hr/). Initially, the main principle of alignment was to align units (words or word combinations) in the Greek text with their Croatian counterparts; these units had to be as small as possible. Full stops and commas were aligned, too. After the initial period it became clear that additional rules were necessary. While the students did not struggle with the meaning of the Greek text, they were sometimes unsure how to align the Greek with the Croatian. These uncertainties typically arose in the following situations due to specific linguistic features of the two languages:

- the use of the article (exist in Greek, but not in Croatian: ὁ Ναζαρηνός = Nazarećanin, Mark 10,47)
- commas can be aligned with conjunctions
- participles (extensively used in Greek, not common in Croatian: ἀκούσας = kad je čuo, Mark 10,47)
- particles (Greek is rich in particles, Croatian often lacks equivalents: the particle δέ is translated as "ali" in Mark 13,5, but left untranslated in Mark 13,13)
- features of Hellenistic Greek (The New Testament was written in this later variety of the Greek language, which is often different from the Classical, 5th century BC Attic dialect of Greek which is mainly taught in schools and universities; one of theseis the preterite form ἤμην διδάσκων = „naučavah" Mark 14,49).

Konferenca
Jezikovne tehnologije in digitalna humanistika
Ljubljana, 2022

Conference on
Language Technologies & Digital Humanities
Ljubljana, 2022

There was also one unexpected problem: students often struggled with aligning prepositions, for example in Mark 1,6: ἐνδεδυμένος τρίχας καμήλου καὶ ζώνην δερματίνην περὶ τὴν ὀσφὺν, the preposition "s" was left unaligned in the Croatian translation ("odjeven u devinu dlaku, s kožnatim pojasom oko bokova").

Consequently, the following set of rules was formed:

1. The article is aligned together with the corresponding nouns, unless translated separately.
2. Conjunctions should be aligned either with conjunctions, particles or punctuation.
3. Punctuation should be aligned whenever possible.
4. Participles should be aligned with the corresponding word combination, even if it is an entire sentence.
5. If something is left out in the translation, the Greek original is left unaligned and *vice versa*, for example the verb "to be".
6. Prepositions should never be left unaligned. Whenever possible, they should be aligned with a corresponding Greek preposition. In the case where a preposition is added in Croatian, together with its noun it should be aligned with the corresponding noun in Greek.

The work done on this corpus highlights several problems in teaching not only Ancient Greek, but also Croatian. Students are unsure of the uses of certain parts of speech, usually those parts of speech that do not have an equivalent in their mother tongue. They are also unaware of the nature of the comma, which can connect (or divide) two words just like a conjunction. Prepositions are often an obstacle because their meaning can be incorporated into a nominal form in Greek and does not have to be expressed separately. These problems probably arise because the school curriculum for Croatian is different from the curricula for Greek and Latin: the curricula for the classical languages pay more attention to grammar, while Croatian has to include both language and literature. Hopefully, projects like this one can highlight specific problems that can then be resolved either by adapting the school curricula or the teaching of classical languages on university level.

3. References

The Alpheios Project. 2019. https://alpheios.net/.

CroALa (Croatiae Auctores Latini). 2014. http://croala.ffzg.unizg.hr.

Chiara Palladino, Maryam Foradi, and Tariq Yousef. 2021. Translation Alignment for Historical Language Learning: a Case Study. *Digital Humanities Quarterly*, 15(3). https://www.proquest.com/openview/e048d32e8e991c67282c3fbda5c1f0d4/1?pq-origsite=gscholar&cbl=5124193.

The Perseids Project. 2017. https://www.perseids.org/.

Ivana Simeon. 2002. Paralelni korpusi i višejezični rječnici. *Filologija*, 38-39: 209–15.

Petar Soldo and Petra Šoštarić. 2018. Treebanking Lucian in Arethusa: Experiences, Problems and Considerations. *Studia UBB Digitalia* 63(2):7–18.

Marko Tadić. 1998. Raspon, opseg i sastav korpusa suvremenoga hrvatskoga jezika. *Filologija* (30-31):337–47.

Marko Tadić. 2001. Procedures in Building the Croatian-English Parallel Corpus. *International Journal of Corpus Linguistics*, *Special issue, pages* 1–17.

Novum Testamentum Graece, Neste-Aland 28. https://www.academic-bible.com/en/online-bibles/novum-testamentum-graece-na-28/read-the-bible-text/bibel/text/lesen/stelle/51/10001/19999/ch/418f354347a79b322324823db62504dc/.

The Zagreb Bible https://biblija.ks.hr/.