

TNA ACTIVITY REPORT

PSALM 77

Author: Constanța Burlacu

Current position: PhD student

Affiliation: University of Oxford

Host institution: Institute of the Czech National Corpus

Mentor(s): Michal Křen

Period of stay: 18 September – 11 December



Psalm 77 Project

During my stay at Charles University in Prague I wanted to acquire new research skills which could support me investigate further the linguistic features of Old Romanian¹ as represented by its oldest written and printed textual sources. In particular, I wanted to examine the aspects of (automatic) text annotation and corpus compilation, two underdeveloped areas of research when it comes to Old Romanian. Additionally, I also wanted to explore the possibility of aligning the text chosen for this pilot study (Psalm 77 of the Old Testament) alongside other textual traditions, especially the Slavonic one. For centuries, in fact, Romanian and Church Slavonic have been in close proximity, with the latter having been the Church and State language of the principalities of Wallachia and Moldavia. Inevitably, Church Slavonic had a significant linguistic and cultural influence on Romanian, which based its first translations of religious texts on Church Slavonic sources.

Psalm 77, with its 72 verses, is the second longest text of the Psalter after psalm 118 (as numbered in the Septuagint). For sixteenth century Old Romanian, it has the advantage of being preserved in 7 digitized sources (3 printed editions and 4 manuscripts), of which 4 are monolingual and 3 are bilingual Romanian-Church Slavonic. Additionally, the Slavonic textual tradition of psalm 77 has been studied in depth by MacRobert,² giving us a good philological understanding of the existing varieties amongst the various redactions³ of the Slavonic text and helping us understand which one of them could have been used for the translation of the text into Romanian. With these tools at hand, Psalm 77 project aims at promoting a rather qualitative type of study, where a small amount of data is taken into account with the aim of developing flexible analysis aids which look into the direction of corpus linguistics and morphosyntactic analysis for Old Romanian. Moreover, by offering a collated versions of psalm 77, on the one hand, the project wants to explore the visual possibilities offered by the digital world and present in one webpage the Romanian versions of the psalm alongside the Slavonic ones. On the other hand, by investigating and devising a morphosyntactic tag-set to be used for linguistic annotations, it will give us some fundamental insight into the linguistic dynamics of Old Romanian, especially as represented by translated texts.

Research Visit in Prague

In autumn 2022 I spend twelve weeks in Prague at the Institute of the Czech National Corpus, working on the Psalm 77 project. During this time of research, I worked on the following aspects:

- Data processing
The project required the collection of linguistic data from sixteenth century handwritten and printed books, that is, the transcription of psalm 77 from its various sources. I have done this with the help of Transkribus, working meanwhile on training transcription models that can automatically process old Romanian materials.
- Transcription and Transliteration
A key feature of old Romanian texts is their script – till the nineteenth century, Romanian was written with Cyrillic letters. A second stage of data processing, therefore, has been the compilation of a correspondence table between Cyrillic and Latin letters, which allowed me to transform further the linguistic data.
- Programming

¹ Conventionally, Old Romanian has been placed within the timeframe of the beginning of the 16th century till 1780.

² C.M. MacRobert, ‘Problems in the study of the “Athonite” redaction of the Psalter in South Slavonic manuscripts’, *Studies of Medieval South Slavic Manuscripts. Proceedings of the 3rd International Hilandar Conference* held from March 28 to 30, 1989, ed. P. Ivić (Belgrade, 1995), 195-213.

³ C.M. MacRobert, ‘The Textual Tradition of the Church Slavonic Psalter up to the Fifteenth Century’, *Interpretation of the Bible*, ed. J. Krašovec (Ljubljana - Sheffield, 1998), 921-42.

Parallel to the tasks described above, I have also learned about programming in Python and its employment for linguistic analysis. With the help of a short Python script, the raw linguistic data was transliterated and transformed into machine-readable material. The script evolved in time, so that the language material could be enriched with linguistic information about its components (POS processing)

- Corpus Compilation

Eventually, with the help of my colleagues from the Institute, a small linguistic corpus was compiled on the basis of Psalm 77 as found in seven different Romanian sources. The idea of the corpus is to be parallel and multilingual; the user, therefore, can access the Romanian materials (for which linguistic information is provided) as well as a Slavonic counterpart for each verse of the psalm.

For those who are curious to see the result of the pilot project, please visit the following link: <https://wiki.korpus.cz/doku.php/en:cnk:psalm77>. Here you will find some additional information about the project.

Future Steps

Based on my positive experience in Prague, it is my intention to continue in the field of Corpus Linguistics and promote the diachronic study of Romanian through the compilation of a corpus of its oldest literary works. Ideally, the study of Old Romanian would be performed in context, that is, in comparison with Church Slavonic when the texts under consideration are translations of religious materials (and this is the case for the earliest Old Romanian materials).

Concretely, my future steps are the compilation of an Old Romanian-Church Slavonic linguistic corpus on the basis of a small example of texts (probably the Gospels), the development of automatic POS taggers for both Church Slavonic and Old Romanian, and eventually the modification of the parallel corpus into a historical corpus of Old Romanian.