
Collaborative Approaches to Open up Russian Manuscript Lexicons

Kira Kovalenko

kira.kovalenko@gmail.com

Austrian Academy of Sciences, Austria

Eveline Wandl-Vogt

eveline.wandl-vogt@oeaw.ac.at

Austrian Academy of Sciences, Austria

In this paper, the authors introduce a new research collaboration between the Institute for Linguistic Studies (St. Petersburg, Russia) and the Austrian Academy of Sciences, Austrian Centre for Digital Humanities (Vienna, Austria), and exemplify multicultural, crossdisciplinary collaboration on the Ph.D project of Russian manuscript lexicons.

The project is aimed at the online representation of the Russian manuscript lexicons, which reflect a very important stage in Russian lexicography and served as a basis for the printed dictionaries appeared later. This lexicographical genre (called *azbukovnik* in Russian) appeared in the middle of the 16th century on the base of glossaries, and differ from them by the alphabetical order of the word entries. About 150 manuscripts, mostly from the 17th century, containing the lexicons came into our hands. The tradition of copying manuscript lexicons preserved in the Old Believers communities, and the latest lexicon dates from the beginning of the 20th century. Lexicons were classified by L.S.Kovtun in the 1989, and some additions were made later by K.I.Kovalenko (for details see Kovalenko, 2016).

Obligatory parts of the word entries were head word (or collocation) and explanation. Word entries could also contain inscription of the language source for loan words, lexicographical or literary sources, citations, and information about topic groups (animals, birds, plants, etc.). Sometimes, observations on various linguistic topics were included in the word entries as well. Despite the great importance of these lexicons as Russian language history resources, only small part of them is available for researchers (2

published in Kovtun, 1975 and 1989, and 16 can be found [online](#) as a facsimile.

The research idea is formulated collaboratively by the Russian and Austrian scholars, experienced in (traditional) lexicography as well as in digital transformation and research infrastructure development. An architecture to embed the Russian manuscript lexicons into a larger lexical resources infrastructure is developed based on these visions and needs at the Austrian Academy of Sciences. Cornerstones of the architecture are opening up lexical resources and develop as much as possible (due to different types of licensing) freely available resources to the broader public which meet [the open science definition](#).

The established architecture consists of three layers:

1. **The human interface layer:** supports data access e.g. list of headwords, geographical map, timeline and other visual approaches etc., which is further developed with every new project. Furthermore, via authentication it offers a private workspace.
2. **The persistent layer:** offers a data repository, e.g. triple store, open access repository to sustainable store and interlink data.
3. **The enrichment layer:** offers opportunities and tools to interlink data and collaboratively enrich data, e.g. with Wikidata or Europeana.

Ideally, all Russian manuscript lexicons in text format and as facsimiles should be included in the research data environment. As a first step, main representatives of each lexicon type are represented. The plain text is marked up in TEI and is provided with metadata which include standardised form of the head word, information about its origin, foreign etymon for loan words (such as *ἀπόστολος* for *апостолъ*) and topic group designation. Word entries contain metadata about lexicographical or literary sources. In case that the facsimile is available, it will be presented along with the text form of the lexicon.

The project is aimed, on one hand, at the representation of the most important lexicons in marked up textual form and as facsimile (if possible). On other hand, the search engine embedded in the infrastructure will make it possible to create enquiries to the data and make selections such as: words according to their origin, literary or lexicographical

sources, words belonging to a particular topic groups and so on. The manuscripts will be semantically exploited and concept-based, interlinked with knowledge resources in the linked open data framework (for more details see Kovalenko et al., 2016). For example, interlinking with Europeana will increase the accessibility and reusability of the data. Thus, the project will make Russian lexicons available and open them for further research and for public curiosity.

The poster will introduce the collaborative approach, the roles of the partners and their contributions, discuss the technical framework based on the research approach and present recent results the first time.

The project is connected to the COST action IS 1305 European Network of electronic Lexicography (ENeL).

Acknowledgements

This research was supported by the RGNF (the Russian Humanitarian Scientific Fund), project № 16-34-01008 “Russian manuscript dictionaries as a cultural phenomenon: the history of the genre and literary context”.

Bibliography

- Kovalenko, K., Wandl-Vogt, E., Schopper, D., Declerck, T.** (2016). “Opening up Russian Manuscript Lexicons for Cultural Heritage Studies.” In *El'Manuscript-2016. Rašytinis palikimas ir informacinės technologijos. VI tarptautinė mokslinė konferencija. Pranešimai ir tezės. Vilnius, 2016 m. rugpjūčio 22–28 d.* Vilnius, pp. 71-74.
- Kovalenko, K.** (2016). “On the Classification of the Russian Manuscript Dictionaries.” In *Words across History: Advances in Historical Lexicography and Lexicology.* Las Palmas de Gran Canaria, pp. 276-286.
- Kovtun, L.S.** (1975). *Leksikografiya v Moskovskoy Rusi XVI — nachala VXII v.* [Lexicography in the Moscow Rus' in the 16th - Beginning of the 17th Century]. Leningrad: Nauka.
- Kovtun, L.S.** (1989). *Azbukovniki XVI-XVII vv. (starshaya raznovidnost) [Azbukovniki of the 16th and 17th Centuries (the Earliest Edition)].* Leningrad: Nauka.