SlaviCorp 2018

24–26 September 2018
Charles University, Prague

Book of Abstracts

Valeria Kolosova
Institute for Linguistic Studies, Russian Academy of Sciences
chakra@eu.spb.ru

Ksenia Zaytseva
Austrian Centre for Digital Humanities, Austrian Academy of Sciences
Ksenia.Zaytseva@oeaw.ac.at

Kira Kovalenko
Institute for Linguistic Studies, Russian Academy of Sciences
Kira.Kovalenko@gmail.com

PhytoLex – the Database of Russian Phytonyms: from Idea to Implementation

Plants have always played an extremely important role in any traditional culture. They served as food, forage, medicine, material for building, clothes, dying, etc. Later some of them were domesticated to make their usage easier. Ancient plant names reflect mythological ideas and language worldview. Among sources, which can tell us about the role of plants in ancient cultures there are archeological findings, anthropological facts, and, of course, texts and inscriptions that allow us knowing plant names and plant knowledge.

In spite of the importance of the plant names investigation, collecting phytonyms is difficult and time consuming, especially for early periods. For some of them, we even do not know the time of their appearing in this or that language and/or their referents. The situation is especially difficult for the Russian language, as the first texts come from the 11th century, which is rather late; at that, most of them being translated from Greek and describing the culture of other people. Though there exist some databases of the early Church Slavonic and Old Russian literature, such as Historical sub-corpus inside the National corpus of the Russian language [1], Corpus Cyrillo-Methodianum Helsingiense – An Electronic Corpus of Old Church Slavonic Texts [2], Old Slavonic Corpus of the University of South California [3], and others, they do not provide semantic search. In fact, all modern sub-corpora of the National corpus of the Russian language have semantic search, while in the Historical part including the Old East Slavic, Birch Bark manuscript, the Old Russian, and Church Slavonic corpora the semantic search is supposed to be provided in the future only for the first one. That means that it is impossible now to have a list of all plant names occurred in the old texts, and that is the reason why most research projects are often based on limited amount of texts or just on the lexicographical materials.

The current project PhytoLex will create favorable conditions for the introduction of new materials into scientific use, for future comparative and typological studies on phytonymy, ethnobotany, folk taxonomy, folk medicine and magic. It will also help overcome fragmentation in Russian studies on folk botany, and provide their compliance with the level and requirements of modern ethnobotanical researches.

Collecting plant names starts from the earliest manuscripts of the Russian literature. The texts chosen for analysis are supposed to cover all the main Old Russian genres from 11th up to 17th centuries, such as religious literature, chronicles, travelogues, lexicographical works (lexicons and phrase-books), herbal books, medicine manuscripts, medical prescriptions and other papers of Apothecary Chancery (Rus. *Aptekarskij Prikaz*).

The sources are being well attributed, including author's name, title, text creation time, as well as time and place of the copy used by a researcher, and information about the book in case if the text was published. For identifying plants, we actively use historical dictionaries and academic books, articles, and theses concerning Old Russian plant names.

To give the full information about plant and its name(s), we create and fill the following description including standard, scientific, and Latin plant names, functions (food, medicine, etc.), metaphorical meaning (if any), word in simplified spelling (close to modern), citation (simplified and as in a source), life form, part of the plant which was mentioned in a text, ways of rendering the foreign phytonym (translation, transliteration, calque, generalisation, etc.) and its foreign etymon.

To make the data more unified, comparable and suitable for analysis, and also to avoid discrepancies, we created a number of controlled vocabularies which, for example, describe functions of plants (decoration, medicine), plant parts (branch, fruit, leaves, root, etc.), literature genres (chronicle, travelogue, herbal book), languages (used or mentioned in texts as plant names sources) and other attributes. In the process of controlled vocabularies creation we are following SKOS [4] standard recommendations and planning to link PhytoLex concepts to external existing thesauri, in particular to the General Multilingual Environmental Thesaurus (GEMET) [5].

The technical implementation of PhytoLex includes data modelling, creation and normalization of controlled vocabularies, development of database and web application for project's data curators and anonymous users on the web, visualization of available geographical data. The project also aims to integrate PhytoLex resources with open access resources like Geonames [6] for georeferencing places mentioned in manuscripts, and Catalogue of Life [7] for scientific name reference.

Overall, the main goal of PhytoLex project is to collect and harmonize data from analogue resources in order to make it available for exploration and analysis, access for further research and reuse.

Acknowledgements

The research is supported by the RFBR (the Russian Foundation for Basic Research), project 17-06-00376 "Russian Phytonyms in the Diachronic Aspect (11-17 cc.)".

References

1. National corpus of the Russian language, http://ruscorpora.ru/. Accessed February 16, 2018.
2. Corpus Cyrillo-Methodianum Helsingiense – An Electronic Corpus of Old Church Slavonic Texts, http://www.helsinki.fi/slaavilaiset/ccmh/. Accessed February 16, 2018.
3. The Historical Syntax of South Slavic, http://www-bcf.usc.edu/~pancheva/HistoricalSyntaxSouthSlavic.html. Accessed February 16, 2018.
4. SKOS Simple Knowledge Organization System Reference, W3C Recommendation, August 18, 2009, https://www.w3.org/TR/skos-reference/. Accessed February 16, 2018.
5. General Multilingual Environmental Thesaurus (GEMET), https://www.eionet.europa.eu/gemet/en/themes/. Accessed February 16, 2018.
6. GeoNames geographical database, http://www.geonames.org/. Accessed February 16, 2018.
7. Catalogue of Life, http://www.catalogueoflife.org/. Accessed February 16, 2018.

Lucie Kopáčková
Czech Language Institute of the Czech Academy of Sciences
kopackova@ujc.cas.cz

Oprahin or Opražin? How to Correctly Form Possessive Adjective from Female First Name or Surname of Foreign Origin in Contemporary Written Czech Language?

How to form possessive adjectives with the suffix *-in* describes every grammar book of Czech language, currently Štícha et al. (2013: 198). Usually, the suffix *-in* is added to the word base without the nominative ending with the consistent consonantic alternation. But how to correctly form these adjectives from some types of female first names of foreign origin shortly describes only Pravdová - Svobodová (2014: 231-233). There are some useful ideas how to form these adjectives, but some first names are missing on the list.

This paper shows how are these adjectives formed by authors of written texts (writers, translators, journalists etc.) in the contemporary Czech language. All presented linguistics data were found in the Czech National Corpus - SYN version 5. The data were obtained by following method. Possessive adjective with the suffix *-in* has its own tag („AU…F.* „); but most of the here presented variants are under the tag „X.* „. It is necessary to search them by a wordform and then all results sort manually, which is time-consuming and laborious process. (see note 1).

The comprehensive analysis showed that from one first name occasionally also from one surname (see note 2) there are often two or three variants. Some adjectives derived from first names with ending *-y (Daisy, Hillary)* or *-ey (Britney)* have as much as four different variants. Also, four variants occurred by names *Sarah* and *Rebecca*. (see note 3). The quantity of variants rises from the ignorance of some irregularities associated with the forming of possessive adjectives from names of foreign origin and probably also from the insecurity about the correct pronunciation of these sometimes exotic sounding names. It is possible to distinguish three basic problems: