К вопросу об электронном представлении комментированного авторского словаря (на материале словаря риторических терминов М. В. Ломоносова)

Филиппов Андрей Константинович[1] **Добров Алексей Владимирович**[2]

[1]Санкт-Петербургский государственный университет (СПбГУ), Санкт-Петербург, Россия

[2]Санкт-Петербургский государственный университет (СПбГУ), Санкт-Петербург, Россия

Towards an Online Representation of a Commented Author's Dictionary (as exemplified in M. Lomonosov's Rhetoric Terms Dictionary)

Filippov Andrei[1] **Dobrov Alexei**[2]

[1]Saint Petersburg State University (SPSU), St.Petersburg, Russia
[2]Saint Petersburg State University (SPSU), St.Petersburg, Russia)

Аннотация

Создание электронного представления для авторских словарей, снабженных научным комментарием, сопряжено со специфическими сложностями. К ресурсам подобного рода относится создаваемый в ИЛИ РАН словарь «Риторика М. В. Ломоносова». Наиболее распространенным среди открытых стандартов электронного представления и структурно-семантической разметки словарных данных является формат XDXF. Для отражения структуры статьи риторического словаря потребовалась доработка этого формата, включая введение новых элементов и расширение области применения старых.

Abstract

Creating an online representation for authors' dictionaries with a scientific commentary implies facing difficulties of a specific kind. One of such resources is «M. Lomonosov's Rhetoric Dictionary» (work in progress at the Institute for Linguistic Studies, Russian Academy of Sciences). One of the most popular open standards for dictionary online

representation and structural-semantic tagging is the XDXF format. To represent the rhetoric dictionary's structure, an adaptation of the format was needed, including introduction of new elements and widening the usage sphere for some of the old ones.

Ключевые слова: XVIII век, авторская лексикография, компьютерная лексикография, структурно-семантическая разметка, словарные платформы.

Keywords: 18th century, author's lexicography, computer lexicography, structural-semantic tagging, dictionary platforms.

Work at «M. Lomonosov's Dictionary» was started at the Institute for Linguistic Studies, Russian Academy of Sciences, St. Petersburg in 2003 (project headed by S. Volkov). Following modern tendencies in author's lexicography [Shestakova, p. 31–33], the project team see the scientific-descriptive function of the dictionary as the principal one, but do overlook its referential and educational functions as well, which implies addressing to a wider circle of readers. This could be facilitated by creating an online version of the dictionary.

Parallel to the Lomonosov's complete dictionary, the project team are working at a number of partial dictionaries. Among them is «Lomonosov's Rhetoric dictionary» (project headed by P. Bukharkin since 2012), a description of rhetorical terminology in Lomonosov's chief work on rhetoric, «Brief Guide to Eloquence». This dictionary is intended to provide a lexicographic description of rhetorical terms used by Lomonosov and to supplement it with a vast historic-cultural commentary [Bukharkin, Volkov, Matveev, 2013a, p. 139].

The rhetorical dictionary was chosen as a starting point for creating electronic version of Lomonosov's complete dictionary. Its entry structure consists of (1) a rhetoric term's lexicographic description and (2) a commentary, or so-called diachronic rhetorical context [Bukharkin, Volkov, Matveev, 2013b, p. 19]. The lexicographic description includes a number of sections common for most kinds of dictionaries (definition, examples, grammar, statistics, etc.) as well as some non-standard ones (e.g. unmodernized form, i.e. the word written according to the 18th century spelling rules). For polysemic words, the definition section is further subdivided into senses belonging to the domains of rhetoric, linguistics in general and non-specific senses. As for the diachronic rhetorical context, this part presents examples of usage for a rhetorical term and its semantic analogs in works on rhetoric from antique authors up to modern reference books. The dictionary consists of 45 entries, each of them being 10–20 times greater size than a conventional dictionary entry.

There are different formats of electronic representation of dictionary data, but most of them are visual, which means that the logical structure of dictionary entry is not represented and thus complex queries by attributes cannot be executed. The framework of this study implied logical representation of dictionary data, which narrowed the choice of formats to XDXF and TEI. Of these two, XDXF [*xdxf_makedict/xdxf_description.md* at master soshial/xdxf_makedictGitHub] was chosen as a proven dictionary interlingua that had already aggregated hundreds of dictionaries in nearly all existing open formats, e.g. DSL (ABBYY Lingvo), mueller7, apresyan, stardict, dictd, Mova, PtkDic. XDXF is a subset of XML so it can be queried with XPATH

and converted with XSLT to any needed representation. Moreover, it can be expanded with new elements and attributes, producing a new format with a wider functionality.

Due to the highly specific entry structure of «Lomonosov's Rhetoric dictionary», even its logical representation in XDXF essentially functions as a visual one. To this end, a minimal adaptation of XDXF project was carried out. This included introducing new tags and attributes and widening the usage sphere for some of the old ones. Following are some examples: *<defgroup>* (a new subsection in the definition section), *lang* as an attribute of *<dtrn>* (language label for a term's semantic analog), *<author>* (a semantic analog's author; previously intended to be use only as an indication of people who took part in making a dictionary).

XDXF is native format for modern free open-source lexicographical platforms like StarDict, GoldenDict, Atlantida, WiseDict, mDictionary, Alpus, SimpleDict, etc. These platforms are standalone applications allowing users to retrieve, browse, and navigate through dictionary entries in different ways. In this research, the dictionary was tested with StarDict and GoldenDictplatforms [*GoldenDict:* dictionary lookout program].

In order to provide compatibility of the developed format with these platforms, an XSLT-transformation table was created. This table can be used, if necessary, to convert all the new logical elements and attributes into visual format, preserving the original layout for other purposes — e.g., for its use in the case of developing a new lexicographic platform.

The expanded list of tags and attributes was used for manually tagging all the dictionary articles. This process revealed some elements of the entries that demanded a further adaptation of XDXF

format. However, current version of the dictionary is functional and can be uploaded into StarDict/GoldenDict dictionary base. As a perspective, compatibility with other platforms (e.g. ABBYY Lingvo via DSL format) is planned. Further steps may involve development of a multi-functional web platform for the dictionary (functions may include: search by any tag or attribute, online editing of any dictionary uploaded, automatic export of dictionaries in different formats, construction of a web connecting multiple dictionaries via weblinks, instruments for citing and commenting the dictionary data, etc.).

Литература

*Bukharkin P. E., Volkov S. S., Matveev E. M.*Problemy leksikografičeskogo opisaniâ ritoričeskoi terminologii M. V. Lomonosova (The problems of lexicographical description of M. V. Lomonosov's rhetorical terms) // Indoevropejskoeâzykoznanieiklassičeskaâfilologiâ (Indo-European Linguistics and Classical Philology-XVII (Joseph M. Tronsky memorial Conference). Proceedings of the International Conference, St. Petersburg, 24–26 June, 2013. St. Petersburg: «Nauka», 2013. — P. 137–159.

GoldenDict: dictionary lookout program. URL http://goldendict.org/ (Accessed date: 30.06.2016).

Ritorika M. V. Lomonosova: proektslovaria (M. V. Lomonosov's Rhetoric: A Dictionary Project) / scientific editors *P. E. Bukharkin, S. S. Volkov, E. M. Matveev.* St. Petersburg: «Gelikon Plus», 2013. — 132 p.

*Shestakova L.L.*Russkaâavtorskaâleksiko
   grafiâ: Teoriâ, istoriâ, sovremennost'
   (Russian Author's Lexicography:
   Theory, History, Modern Times).
   Moscow: «Âzykislavianskihkul'tur»,
   2011. — 464 p.—
   (Studiaphilologica).

*xdxf_makedict/xdxf_description.md*at
   master soshial/xdxf_makedictGitHub.
   URL
   https://github.com/soshial/xdxf_make
   dict/blob/master/format_standard/xdx
   f_description.md (Accessed date:
   30.06.2016).

References