Poetology: Problems of Constructing a Thesaurus and Verse Text Specification

V. N. Boykov*, and M. S. Karyaeva**, **
*Demidov Yaroslavl State University, Yaroslavl, 150003 Russia
**e-mail: boykov_bh@bk.ru
**e-mail: mari.karyaeva@gmail.com
Received October 15, 2017

Abstract—It is a brief version of the report made by the authors at the seminar “Modeling and Analysis of Information Systems,” Yaroslavl, May 17, 2017. The connection between the tasks of automatic thesaurus construction and verse specification is considered.

Keywords: verse, poetology, thesaurus, information analytical system
DOI: 10.3103/S0146411619070253

INTRODUCTION

Poetology is a group of disciplines, focused on a comprehensive theoretical and historical study of poetry (as a way of organizing speech), its texts and verse (as a segment of a text written in verse form). The task of studying a text in verse as an object of poetology is its specification. In other words, on the one hand, it involves defining parameters of specification represented by a certain matrix (or a form) of distinctive attributes (features, properties, characteristics) of a particular class of study objects. On the other hand, it is a procedure of an object identification, which includes detection of unique values of these attributes for a specific object.

Research of this kind is supposed to have a meta description of verse text in terms of a thesaurus in the domain of poetology, which, therefore, requires the existence the thesaurus or its automatic construction as in our case. A thesaurus is a collection of terms (words or word combinations), where each term is a domain specific concept with its definition, as well as relations between them which constitute their specification [3]. Thus, both verses and the terms of the thesaurus are the objects of research as well as the establishment of their specifications (meta-descriptions). The complex solution of these tasks characterizes poetology as an information analytical system [1, 2].

The poetology thesaurus was set up as an instrument for linguo-poetic and historical literary meta-description of poetry and text in verse in previous studies [5—9, 11—14]. The original terminological list of the poetology thesaurus was composed of 1544 words and phrases. A hierarchical index for the upper levels of terms has been created and the lower levels were clustered, so rubrication of thesaurus terms on poetology in the sub-domain is as follows:

1. Poetry.
2. Stylistics.
3. Poetics.
4. Rhetoric.
5. History of literature.
6. Translation studies and literary comparative studies.
7. Textual study.
8. Hermeneutics.
9. Theoretical schools and branches.
10. Logic and methodology of science.
Automatic construction may not apply to all sections of the first level of the thesaurus. Most of the terms of the thesaurus represent the subject subdomain Poetry, the sub-heading of which is the first heading of the second level “Verse”—is divided into 6 subtopics of the third level:

(1.1.1) Metrics.
(1.1.2) The phenomena of the beginning and end of a poetic line.
(1.1.3) Rhythm.
(1.1.4) Stanzas theory.
(1.1.5) Rhyme.
(1.1.6) Verse linguistics.

The main attributes of these subtopics can be automatically analyzed and set the corresponding specification of the verse text. Generally, the purpose of the information analytical system is to establish three levels of text specification:

(C1) Metric-stanza specification of a particular work of poetry (verse text).
(C2) Poetic and stylistic specification of the work of poetry of a particular author based on the metric-stanza specification of their works.
(C3) Historical and literary specification of branches, schools and periods based on the statistical similarity of the poetic and stylistic specifications of various authors.

It is necessary for automatic analysis to distinguish between a work of poetry as an aesthetic entity and a text in verse as an object of study of poetry. Consequently, we should distinguish between linguistic poetic tagging of the verse and the specification of the text in verse. Verse tagging mainly reflects the numerical (quantitative and ordinal) characteristics of the text in verse while specification is its meta-description based on the terms of the thesaurus.

Verse tagging (line tagging) allows to create a formal linguistic syllable representation of the verse [4, 10]:

$$b^{r(1)}_k A_d^{s(1)} b^{r(2)}_k A_d^{s(2)} ... b^{r(k)}_k A_d^{s(k)},$$

where $A$ means ictus (metrically stressed syllables), $k$ is the number of ictus, $b$ and $d$ mean unstressed prepositional and postpositional syllables between ictuses, respectively, $r$ and $s$ stand for the length of the corresponding subchains of unstressed syllables. Subgroups $b^{r(l)}_j$ (anacrusis), $d^{s(k)}_j$ (clausula), $d^{s(j)}_j b^{r(j+1)}_j$ (interictus interval), $j = 1, 2, ..., (k - 1)$, with the number of ictus $k$ and the number of syllables in the line $(k + \sum (r(j) + s(j + 1))_{j=1}^{k-1})$ represent the regulations that determine the repertoire of metro-rhythmic forms of verse in various combinations of their values. The full repertoire of metric schemes should include all forms of omissions of circuit stresses and supercircuit (extrametric) stresses. The definition of the metro-rhythmic form of a verse is an essential characteristic of its specification.

New authors of poetry, contributing something novel to the process of versification, appear all the time. At the same time new researchers, discovering new phenomena in poetology, emerge in study of poetry. These two facts bring both additional terms to thesaurus and changes of specification of the terms. This, in its turn, updates the matrix of the specification of text in verse. Consequently, poetology as an information analytical system cannot and should not be finite.

**FUNDING**

This work was supported by RFBR, project nos. 16-07-01180 and 16-06-00497.

**CONFLICT OF INTEREST**

The authors declare that they have no conflicts of interest.

**REFERENCES**

1. Zakharov, V.E., Boykov, V.N., Vigurskii, K.V., and Pil’shchikov, I.A., Russkaya poeziya: Problemy konsolidatsii i analiza v elektronnom formate (Russian Poetry: Problems of Consolidation and Analysis in Electronic Format), Moscow, 2004.
2. Boykov, V.N., Zakharov, V.E., Pil’shchikov, I.A., and Sysoev, T.M., Thesaurus as a poetological tool, Model. Anal. Inf. Sist., 2010, vol. 17, no. 1, pp. 5–23.
3. Lukashevich, N.V., Tseaurusy v zadachakh informatsionnogo poiska (Thesauri in Information Retrieval Problems), Moscow: Izd. Mosk. Univ., 2011.
4. Boykov, V.N., A context-free grammar of one rhythmic model of Russian verse, *Model. Anal. Inf. Sist.*, 2012, vol. 19, no. 4, pp. 154–167.

5. Boykov, V.N. and Pil’shchikov, I.A., The semantic model of “Thesaurus on Poetology” as part of the information-analytical system, *Internet i sovremennoe obshchestvo, Trudy XVI Vserossiiskoi konferentsii IMS-2013* (Internet and Modern Society, Proceedings of the XVI All-Russian Conference IMS-2013), St. Petersburg, 2013, pp. 273–279.

6. Boykov, V.N., Zakharov, V.E., Karyaeva, M.S., and Sokolov, V.A., Domain-specific thesaurus as a tool for information retrieval and collection of knowledge, *Model. Anal. Inf. Sist.*, 2013, vol. 20, no. 4, pp. 125–135.

7. Boykov, V.N., Zakharov, V.E., Karyaeva, M.S., and Sokolov, V.A., A subject-oriented thesaurus in an open information-analytical system, *Elektronnye biblioteki: Perspektivy, metody i tekhnologii, elektronnye kolektcii, RCDL’2013* (Digital Libraries: Prospects, Methods, and Technologies, *Digital Collections, RCDL’2013*), 2013, pp. 70–76.

8. Boykov, V.N., Zakharov, V.E., Karyaeva, M.S., and Sokolov, V.A., On the automatic rubrication of thesaurus terms of an open information-analytical system, *RCDL-2014*, Dubna, 2014.

9. Boykov, V.N., Zakharov, V.E., Karyaeva, M.S., and Sokolov, V.A., An information and analytical block of the thesaurus as a learning tool, *Obrazovanie, nauka i ekonomika v vzakh i shkolakh. Integratsiya v mezhdunarodnoe obrazovatel’noe prostranstvo. Trudy mezhdunarodnoi konferentsii* (Education, Science, and Economics at Universities and Schools. Integration into International Educational Space. Proc. Int. Conf.), Tsagkhadzor, 2014, pp. 205–209. http://nuclphys.sinp.msu.ru/math/conf/tsagkhadzor_conference_part1.pdf.

10. Boykov, V.N., Karyaeva, M.S., Sokolov, V.A., and Pil’shchikov, I.A., On the automatic specification of a verse in an information-analytical system, *DAMDID/RCDL-2015*, Obninsk, 2015.

11. Karyaeva, M.S., Linguistic and statistical analysis of the terminology for constructing the thesaurus of a specified field, *Model. Anal. Inf. Sist.*, vol. 22, no. 6, pp. 834–851.

12. Karyaeva, M.S. and Sokolov, V.A., An overview of methods for extracting hypo/hyperonyms from document collections, in *Zametki po informatike i matematike, Sbornik nauchnyh statei* (Notes on Computer Science and Mathematics, Collection of Scientific Articles), Yaroslavl: Yarosl. Gos. Univ. im. P. G. Demidova, 2015, vol. 7, pp. 53–58. http://www.lib.uniyar.ac.ru/edocs/iuni/20150406.pdf.

13. Karyaeva, M.S. and Sokolov, V.A., About information technologies for solving the problem of extracting domain terms, *Matematika, fizika, informatika i ikh prilozheniya v naute i obrazovanii, Sbornik tezisov dokladov Mezhdunarodnoi shkoly-konferentsii molodykh uchenykh* (Mathematics, Physics, Informatics, and Their Applications in Science and Education. Abstracts of the International School—Conference of Young Scientists), Moscow, 2016, pp. 143–145. https://elibrary.ru/item.asp?id=28373830.

14. Karyaeva, M.S. and Sokolov, V.A., On the problem of multi-word term extraction from a domain-specific document collection, *Data Analytics and Management in Data Intensive Domains (DAMDID/RCDL’2017)*, *Proceedings of the XIX International Conference*, Moscow, pp. 218–221.