

The Development of the Open Dictionary of Contemporary Serbian Language Using Crowdsourcing Techniques

Lazić Konjik I., Milenković A.

Institute for the Serbian Language of SASA, Serbia

Abstract

The paper introduces the idea of developing an Open Dictionary of Contemporary Serbian Language using the crowdsourcing method with a clear vision of the necessity of modernizing the work in the field of Serbian lexicography. The crowdsourcing procedure is expected to obtain the latest, present-day vocabulary, i.e. new words which have not yet been presented and described in existing dictionaries of the Serbian language. The crowdsourcing task requires participants to suggest new words, along with the description of their meaning and examples which they find most appropriate. There are two primary goals: (1) solving the problem of timely and comprehensively recording and presenting new vocabulary in real time and (2) creating a user-oriented dictionary of new words and expressions. The main idea is to create a dictionary which would offer simple explanations of the semantic, grammatical and pragmatic features of new words and expressions in the Serbian language. The practical, lexicographic objective is to show the need for the development of such a dictionary which would fill the gap that exists in Serbian descriptive lexicography and neography.

Keywords: e-lexicography, dictionary, neography, crowdsourcing, contemporary Serbian language

1 Introduction

In this paper we will briefly outline the idea of a new online dictionary – Open Dictionary of Contemporary Serbian Language (ODCS). The ODCS would represent a digital lexical database where new words and expressions, which have appeared in the Serbian language in the last ten years, but were not lexicographically processed, would be systematically collected, using a modern lexicographic methodology and explained in a user-accessible way.

We will present the goals which are set before the realization of the planned dictionary and the theoretical and methodological aspects of its development. Accordingly, the paper is structured as follows: in section 2 we present the objectives; section 3 gives a brief overview of the type and purpose of the ODCS; section 4 describes language material collecting procedures; sections 5 and 6 present the macro- and microstructure of the ODCS, and in section 7 we conclude and discuss the future work and perspectives of the ODCS.

2 Objectives

The appearance of new words in languages is an unceasing process. The role of the lexicographer is to record and present new words and expressions in a timely manner, both to the general (laic) and professional public. With the development of digital technologies, this long-standing need of both lexicographers and dictionary users has become attainable, moreover, it stands as a challenge before modern dictionaries (Apresjan 2000; Atkins & Rundell 2008; Hanks 2012). In addition to this demand, a priority is given to the requirement that the dictionary meets the needs and expectations of its future users (Atkins & Rundell 2008). This practically means that it is necessary to anticipate the reasons why the user is searching for a particular word and then try to provide a set of appropriate answers (Varantola 2002: 33). This, furthermore, indicates the necessity of defining and starting from a potential target group of users.

Since we are discussing a dictionary of new words, we emphasize that the target group of the ODCS is a wide range of potential users, non-professionals and professionals, who are usually Serbian native speakers (L1). This is why the main motive for planning and compiling dictionaries of new words is about meeting their communicative needs. By professional users we mean all the people who use dictionaries as reference sources in their work (Varantola 2002: 32) which, in the case of the ODCS, are primarily linguists and lexicographers.

Accordingly, the primary goal of the new ODCS would be to offer simple explanations of the semantic, grammatical and pragmatic features of new words and expressions in the Serbian language. Firstly, to help the broadest range of non-professional users understand the meaning of new words and to encourage them to use these words in naming objects, concepts and phenomena which are part of their everyday life. Secondly, the demands of professional users set a goal before the ODCS to become a modern and reliable lexicographic database of new words and expressions, as well as a reference source for various types of linguistic activities and research. And finally, to apply the user-produced data in the domain of the semantic interpretation of terms into further research of the interpretive user perspective of different types of terms and, in this regard, the ways of defining them (cognitivism) (cf. Bartminjski 2011: 93-168). This is part of our future work and we will give some remarks on it in the concluding section.

This way, the ODCS would fill the gap which has been noticeable in Serbian descriptive lexicography in the last ten years (the last dictionary of the Serbian language was published in 2011) as a basic lexical supplement to the existing descriptive dictionaries of the Serbian language (the one-volume DSL and the six-volume MSD). At the same time, it

would become a relevant source and reference tool for the development of future dictionaries. The final result would be a user-friendly, easy-to-use dictionary, where new words, which have recently appeared in the language and have, more or less incorporated into it, are explained in a simple, transparent and linguistically reliable way – so that potential users can easily understand and use them.

3 The Type and Purpose of the Dictionary

The proposed dictionary will introduce the present-day vocabulary of contemporary Serbian language covering the last ten years (approximately from the year 2010 until today) which has not been recorded in its descriptive dictionaries: the DSL, MSD and SASAD, nor in the existing dictionaries of foreign or new words: Ćirilov 1982, 1991; Klajn 1992; Otašević 1999, 2008a; Vasić, Prčić & Nejgebauer 2001; Bugarski 2019. These printed dictionaries cover the period up to 2008 (except for the dictionary Bugarski 2019, which, however, brings only lexical blends), so the comparison with them will determine the relevance for their inclusion into the Open Dictionary. Other criteria for the selection of the lemmas are described in section 5. The ODSL will have a descriptive character. There is no intention of it becoming a formal normative source, but certain information of this type will be included by presenting appropriate orthographic and semantic uses of words, in relation to Serbian descriptive lexicography (cf. Ristić 2006: 41-64, 79-92). This is very important because it is assumed that one of the reasons why users search for a certain word is its normative status. The focus will be on introducing words (vocabulary) which are, in conceptual or word-formation terms, a novelty in the Serbian language.

4 Collecting Language Material

The current practice in Serbian descriptive lexicography, especially neography which has, in recent years, stood out in many ways as an independent field (cf. Благоева 2016: 200), is grounded upon traditional methods of collecting words – manual excerption from newspapers and magazines, literary work, radio and television and colloquial language (see the above-mentioned Serbian dictionaries of new words). At the same time examples of recent Slavic and European efforts, as well as the papers, which present the state-of-the-art research into neology and ideas about the modern lexicographic treatment of neologisms, show that the procedures of collecting and excerpting new words for lexicographic presentation have been greatly modernized and improved by introducing the corpus method and digital performance (Sinclair 1991; Meyer 2004; Biber & Reppen 2015). Nevertheless, corpus-based lexicography is still, to a great extent, combined with the traditional manual method (e.g. Благоева 2011: 18; Karlsson 2000), but more frequently uses language tools which enable a semi-automatic or automatic selection of new words (Krek, Kosem & Gantar 2013; Klosa-Kückelhaus & Ilan Kernerman 2020). Concerning the data collection process, we can notice the emergence of a new form, the so-called user-generated dictionaries where word collecting is partially or fully performed by the crowdsourcing method (e.g. the Slovene *Sprotni slovar*,¹ the English Macmillan *Open Dictionary*,² the Swedish *Folkmun*³).

Recent approaches in lexicography show that the crowdsourcing method has been recognized as a suitable method for less complex lexicographic tasks (Čibej, Fišer & Kosem 2015: 71), so collecting new words is, based on previous experience, one of the tasks which can be performed successfully with this method (Rundell 2012; Sköldberg & Wenner 2020). Research has shown that successful crowdsourcing projects can attract a great number of volunteers if the public is well-informed and asked to help, especially when such projects are supported by non-profit organizations with the goal of “general welfare” (Holley 2010). European practice and some successful domestic projects show that there are many people outside formal institutions who have excellent skills and are willing to invest considerable time and effort into linguistic resources which are of general interest, such as free online dictionaries and encyclopedias, e.g. Dict.cc,⁴ Wiktionary,⁵ Lingobee,⁶ Wikipedia,⁷ or the language learning platform, e.g. Duolingo,⁸ Memrise,⁹ or domestic projects, e.g. Jezička laboratorija,¹⁰ Prepis.org,¹¹ etc. (cf. Čibej, Fišer & Kosem 2015: 71; Fišer & Čibej 2017: 214-215).

Crowdsourcing is planned on being used in the development of the new ODCS as the basic way of obtaining material,¹² since there is no available current and up-to-date corpus of Serbian language (the existing corpora SrpKor¹³ and SrWaC¹⁴ cover the period up to 2013 and 2014) or appropriate language tools which would enable (semi-) automatic excerption. Such tools are being developed in the Serbian language community, but for now mostly in the field of terminology (e.g. Krstev et al. 2015; Stanković et al. 2016; Šandrih et al. 2020). The general purpose of crowdsourcing is to accelerate the

¹ Accessed at: <https://fran.si/> [07/07/2020]

² Accessed at: <https://www.macmillandictionary.com/open-dictionary/latestEntries.html> [07/07/2020]

³ Accessed at: <https://folkmun.se/> [07/07/2020]

⁴ Accessed at: <http://www.dict.cc/> [07/07/2020]

⁵ Accessed at: <https://www.wiktionary.org/> [07/07/2020]

⁶ Accessed at: <http://itrg.brighton.ac.uk/simola.org/lingobee/> [07/07/2020]

⁷ Accessed at: <https://www.wikipedia.org/> [07/07/2020]

⁸ Accessed at: <https://en.duolingo.com/> [07/07/2020]

⁹ Accessed at: <https://www.memrise.com/> [07/07/2020]

¹⁰ Accessed at: <http://lab.unilib.rs/> [07/07/2020]

¹¹ Accessed at: <http://www.prepis.org/> [07/07/2020]

¹² On the importance and advantages of electronic corpora as a source of information in lexicography in relation to traditional lexical files see Fillmore 1992; Atkins & Rundell 2008: 53.

¹³ Accessed at: <http://www.korpus.matf.bg.ac.rs/> [07/07/2020]

¹⁴ Accessed at: https://www.clarin.si/noske/all.cgi/first_form?corpname=srwac:align [07/07/2020]

initial (and continuous) process of developing a database (corpus) which will serve as the starting point in forming a lexical vocabulary database, after the process of lexicographic evaluation. The crowdsourcing procedure is expected to obtain the latest, present-day vocabulary such as, for example, the lexicon which has emerged as a result of the outbreak of the coronavirus pandemic.

Crowdsourcing techniques will be combined with the traditional method of manual excerption of targeted sources such as scientific papers, MA theses and PhD dissertations, in which neology topics have been discussed in the last ten years, because of the great amount of new words and lexical material obtained in them.¹⁵ We assume that, this way, a large amount of vocabulary which has remained in the gap between theory and practice, due to the discrepancies between lexicology and lexicography, will successfully be obtained.

Every Serbian native speaker will be able to participate in the development of the planned dictionary. The workflow basically follows the scheme presented in the paper Čibej, Fiser & Kosem (2015: 75-77). An appropriate user interface will be made with an integrated specific proposal form, thanks to which the language community will be involved in resource development. The crowdsourcing task requires participants to suggest new words, as a potential lemma candidate, along with the description of meaning. Also, if they wish, they can provide other information which they consider relevant to the description of the word they are proposing, such as e.g. the circumstances or the situation in which the given word is used, its synonyms, etc. Additionally, votes for other users' proposals whether a word is or isn't new will be enabled. Further verification (validation) of the entries made by the participants in the crowdsourcing task, the selection and lexicographic processing, in accordance with the theoretical concept of the dictionary, will be performed exclusively by lexicographers (see section 5). All the user provided data will be stored in a special database in the form of a list of tokens, available for online viewing and downloading. The dictionary base (i.e. the processed dictionary entries) will be kept separately.

5 The Macrostructure

New words and expressions in Serbian and other languages, include lexical neologisms (novel and recently borrowed words naming new realities and concepts from different fields), semantic neologisms, word-formation neologisms, neearchaisms, recently introduced compounds of terminological character and new phraseology (Otašević 2008: 39-40; Ristić 2012: 9; Dragičević 2018: 237-247; Благоева 2011: 18). New words represent mainly the peripheral lexicon, neologisms, occasionalisms and potential lexis (individualisms), which have a variable status in the dynamics of lexical development. The characteristics of neologisms is that they undertake different phases of usualization, which begins with the phase of occasionalisms. The relation of neologisms to time is direct, while the chronology criterion for occasionalisms and individualisms is relative (Otašević & Sikimić 1991: 77-78; Dragičević 2018: 228-237). From the aspect of neography, the criterion of chronological status is superior, and the question of their status in the lexicon is mainly a question of the test of time. Accordingly, in digital neography, the question of selection, in terms of distinguishing a neologism in the narrower sense vs. a neologism in the broader sense of the term becomes secondary, when it includes individual vocabulary (individualisms, hapaxes, occasionalisms).

As opposed to the paper (printed) editions, the electronic dictionaries impose no restrictions on the extent and amount of language material that can enter the dictionary, which allows for the gathering of different layers of new lexical phenomena, including: jargon, speech lexis, newly borrowed words, acronyms, loan translations (calcs) and other types of new vocabulary (Котелова 1978: 25-26; Ristić 2012: 1, f.10; Bugarski 2006: 42 & 2019; Dragičević 2018: 222-247). This is a major advantage knowing that languages are constantly changing, evolving, new words and expressions appearing, old ones disappearing, the vocabulary is constantly regrouping, words moving from the peripheral registers into the general vocabulary funds and vice versa, which is a common and integral part of its development dynamics (Ristić 2012: 11). The recording of all new words, even one-day words which quickly disappear from the language, is significant not so much from the point of view of their acquisition and knowledge, but from the point of studying the principles, tendencies and perspectives in the development of the vocabulary of a language, from the synchronic and diachronic aspect. New technologies create a great opportunity for all lexicographers because they enable an exhaustive presentation of the most recent lexical changes and appearances which take place in one language's development and growth.

The main criteria for the selection of entries will be both chronological and lexicographic (cf. Otašević 2008: 19) – that is, all the words which have not yet been announced and recorded in corresponding Serbian dictionaries from the year 2010 (see section 3). Other, special ways of detecting neologisms, essentially, do not exist, but there are various ways of checking whether an unrecorded word is a lemma candidate in general (cf. Trap-Jensen 2020). The basic way of checking is the frequency factor. Namely, the high frequency of a word in the corpus indicates the institutionalization of the lexical unit, but the speakers' perception about the novelty of these words will also be taken into account (cf. Freixa & Torner 2020), thanks to the user participation in dictionary development. Therefore, for each user suggestion, it will be checked whether there is evidence that the word is used (based on reference sources, in the existing available corpora of the Serbian language, and given that it is a current vocabulary that dictionaries and corpora, as a rule, do not yet record, the

¹⁵ Selected PhD and MA theses defended at the University of Novi Sad: The Faculty of Philosophy, Serbia: Rakić, K. (2016). Lexicological and lexicographical position of the current loanwords in the Serbian language; Savić, M. (2013). Semantico-Pragmatic Pseudoanglicisms in Serbian; Aćimović, S. (2013). Anglicisms in the Novels of Modern Serbian Women Writers of the Younger Generation; Rodić, S. (2013). Facebook communication on reality shows: the lexical influence of English on Serbian language; Jerković, V. (2012). Understanding Anglicisms in specialized teenage magazines in Serbian; Janjatović, V. (2010). The use of Anglicisms in Women's Fashion Magazines in Serbian, etc.

check will also be done online). If there is evidence showing that the word is in use, it will be selected and added to the dictionary after lexicographic processing. This way the corpus will exclude misspellings, tendentiously invented words, vulgar and obscene words, and other similar words which will still be stored in the primary database containing all user-produced suggestions. Schicchi & Pilato (2018) describe a semi-automatic method for creating neologisms and show that the number of tendentiously created words for commercial purposes in order to produce a creative name which will sell a product or service, can make a significant share in the vocabulary as a whole. The so-called vulgar and obscene lexicon, which is traditionally omitted from Serbian dictionaries, poses a special problem. While contemporary understandings indicate the need for an objective representation of all lexical units of a language, the real situation shows that the parameters for identifying such vocabulary have not yet been developed, neither in terms of content nor in terms of use (Ristić 2006: 91).

The advantages of digital technology provide the possibility of making rapid and permanent changes in the composition of dictionary entries (supplementation, rearrangement, etc.). Also, the lemmas will be presented in a concise or an extended form. In that way, the entry will be complemented with new contents and information which have not been customary for printed dictionaries of new words, and are, also, in line with the requirements of monolingual descriptive lexicography: the pronunciation of words, full forms of declension and conjugation, pragmatic information, thesaurus functions, established collocations, multimedia contents.

The dictionary will be incrementally updated and the entries will appear in the dictionary immediately after lexicographic processing, so we can say that the dictionary will be constantly complemented and created before the eyes of the user, reflecting the living nature of language and its continuous development. The database will be permanently available for download in at least one of the standard text formats (XML, CSV, JSON, YAML). A cross-section is planned yearly (once a year) in the form of individual dictionary versions.

The accompanying, not less important parts of the dictionary will introduce additional search criteria which is intended to meet the specific needs of users and different levels of their competence (Varantola 2002: 39). Different choices of displaying and searching will be offered according to: the standard alphabetical order, different functional-stylistic and semantic registers, etymology (foreign word / domestic word), the criterion new word / new meaning of an existing word. Labels for semantic fields in each entry can be used as a navigational tool to display a list of all entries from the given field, enabling thematic browsing through the collection (see Tasovac & Petrović 2015: 392). The thematic (semantic) classification of vocabulary in Serbian studies is mainly considered in connection with the dialectal and traditional vocabulary in dialectology, ethnolinguistics and dialectal lexicography (Miloradović 2012: 146; Tasovac & Petrović 2015; Lazić Konjik 2017), from which we will start, continue and develop in relation to the material. The labels for the semantic fields will also show which semantic domains are current in the contemporary linguistic picture of the world. In the case of a new meaning of an existing word the lemma will have in its exponent a symbol (mark) which will indicate that it's about a new meaning of an existing word, and not a new word.

In this way, the multi-layered content of the dictionary will be closer to potential users and will provide easy and quick access to the information they need.

6 The Microstructure

The basic characteristics of the microstructure provide that the design is entirely at the service of the function to be fulfilled by the dictionary in the presented context (Kiefer & Sterkenburg 2003: 351). The lexicographical processing is completely under the jurisdiction of the lexicographer.

- basic lemma (the most common, as well as all the acceptable spelling forms);
- the stressed form and alternatively, all acceptable pronunciation forms in use;
- the grammatical category of: word type, gender for nouns, verb type, gender forms for adjectives;
- etymology (original form of a foreign word);
- word-formation structure;
- inflectional grammatical forms (declension and conjugation);
- notes on pragmatic and stylistic values;
- definition (presenting the explanation of the new word's meaning, given in full sentences, with high frequency common words which are easy to understand);
- information on lexical and semantic compatibility, collocations in use;
- examples (whole sentences, as many examples as possible);¹⁶
- paradigmatic lexical relations (synonyms, antonyms, opposites);
- phraseology (all phraseologisms will be united in a separate dictionary part, connected to their primary lemmas);
- multimedia content.

7 Conclusion and Perspectives

In this paper we have introduced the idea of developing an Open Dictionary of Contemporary Serbian using the innovative crowdsourcing method, with a clear vision of the necessity of modernizing the work in the field of Serbian lexicography. In addition to the questions of the dictionary concept, the methodology of making and organizing the content on the micro and macro plan, which are mostly considered, the development of a new dictionary always opens numerous theoretical questions, especially in the domain of the presentation of semantic content and interpretation of the

¹⁶ The usage degree of the entry-word will be indirectly indicated by the number of examples and collocations cited.

ways in which words function in language and texts. This leads to current cognitive theoretical approaches in which the problem of the boundary between the so-called linguistic and non-linguistic knowledge and ways of defining different types of concepts is discussed, which will also be the subject of our future work. Cognitivism, as pointed out by J. Bartminjski, requires respect for the interpretive perspective adequate to the competence of the user of the language that is the subject of analysis (Bartminjski 2011: 97). We expect that a significant contribution to research in this area will be given by the data obtained from our users.

Apart from the fact that such a dictionary would be an important source of data for the understanding of novel lexical meanings and word usage, it would also serve as an important language infrastructure and source for linguistic studies of contemporary Serbian, for monitoring its development, language standardization and linguistic planning. All of this together stands as a prerequisite for the modernization of scientific practice in the field of language science and other humanities in the modern IT society.

With the help and financial support of the competent institutions, the ODCS would be available, free of charge, to all interested users, researchers and all who wish to develop applications, programs and tools for the Serbian language based on this dictionary as a source of lexical information. A well-structured data set, as will be the outcome of this project, can encourage further scientific analysis in the domain of language technologies, both within the dictionary itself or linking up with other structured data collections, such as e.g. (morphological) e-dictionaries, electronic corpora or other future digitized dictionaries of the Serbian language.

8 References

- Apresjan, J. (2000). *Systematic Lexicography*. Oxford: Oxford University Press.
- Atkins, S.B.T., Rundell, M. (2008). *The Oxford Guide to Practical Lexicography*. Oxford: Oxford University Press.
- Bartminjski, J. (2011). *Jezik – slika – svet. Etnolingvističke studije*. D. Ajdačić (ed.). Beograd: SlovoSlavia.
- Biber, D., Conrad, S. & Reppen, R. (1998). *Corpus Linguistics: Investigating Language Structure and Use*. Cambridge: Cambridge University Press.
- Bugarški, R. (2006). *Žargon – lingvistička studija*. Drugo, prošireno izdanje. Beograd: Biblioteka XX vek.
- Bugarški, R. (2019). *Srpske slivenice, monografija sa rečnikom*. Novi Sad: Akademska knjiga.
- Čibej, J., Fišer, D. & Kosem, I. (2015). *The role of crowdsourcing in lexicography*. Accessed at: https://elex.link/elex2015/proceedings/eLex_2015_05_Cibej+Fiser+Kosem.pdf [05/15/2020].
- Ćirilov, J. (1982). *Rečnik novih reči*. Beograd: Narodna knjiga.
- Ćirilov, J. (1991). *Novi rečnik novih reči*. Beograd: Bata.
- Dragičević, R. (2018). *Srpska leksika u prošlosti i danas*. Novi Sad: Matica srpska.
- DSL (2007). *Rečnik srpskoga jezika*, Novi Sad: Matica srpska.
- Fillmore, Ch. (1992). Corpus Linguistics vs. Computer Aided Armchair Linguistics. In J. Svartvik (ed.). *Directions in Corpus Linguistics*. Berlin and New York, pp. 35-60.
- Fišer, D., Čibej, J. (2017). The potential of crowdsourcing in modern lexicography. In V. Gorjanc, P. Gantar, I. Kosem & S. Krek (eds.) *Dictionary of Modern Slovene: Problems and Solutions*. Ljubljana: Ljubljana University Press, Faculty of Arts, pp. 212-229.
- Freixa, J. & Torner, S. (2020). Beyond Frequency: On the Dictionarization of New Words in Spanish. In *Dictionaries: Journal of the Dictionary Society of North America*, 41(1), pp. 131-153.
- Hanks, P. (2012). The corpus revolution in Lexicography. In *International Journal of Lexicography*, 25/4, pp. 398-436.
- Holley, R. (2010). Crowdsourcing: How and Why Should Libraries Do It? In *D-Lib Magazine*, 16 (3/4). Accessed at: <http://www.dlib.org/dlib/march10/holley/03holley.html> [05/15/2020].
- Jovanović, J. (2019). Nova upotreba prideva human u savremenom srpskom jeziku. In *Novorečje*, 1, Beograd: Alma, pp. 5-8. Accessed at: <http://alma.rs/Novorecje/Novorecje1.pdf> [05/15/2020].
- Karlsson, F. (2000). Lexikografie a korpusova lingvistika. In F. Čermák, J. Klimov, V. Petkevič (eds.) *Studie z korpusové lingvistiky*. Praha: Karolinum, pp. 427-454.
- Kiefer, F., Sterkenburg, P. (2003). Design and production of monolingual dictionaries. In P. Sterkenburg (ed.) *A Practical guide to lexicography*. Amsterdam/Philadelphia: John Benjamins Publishing, pp. 350-365.
- Klajn, I. (1992). *Rečnik novih reči*. Novi sad: Matica srpska.
- Klosa-Kückelhaus, Ilan Kernerman (2020) (ed.). Global Viewpoints on Lexicography and Neologisms. In *Dictionaries: Journal of the Dictionary Society of North America*, 41(1). Accessed at: https://muse.jhu.edu/issue/42292#info_wrap [07/07/2020].
- Krek, S., Kosem, I. & Gantar, P. (2013). *Predlog za izdelavo Slovarja sodobnega slovenskega jezika (A Proposal for a new Dictionary of Contemporary Slovene)*. Version 1.1. Accessed at: http://trojina.org/slovar-predlog/datoteke/Predlog_SSSJ_v1.1.pdf [05/15/2020].
- Krstev, C., Stanković, R., Obradović, I. & Lazić, B. (2015). Terminology Acquisition and Description Using Lexical Resources and Local Grammars. In T. Poibeau & P. Faber (eds.) *Proceedings of the 11th Conference on Terminology and Artificial Intelligence, Granada, Spain, 2015*. LexiCon (Universidad de Granada), pp. 81-89.
- Lazić Konjik I. (2017). Leksika tradicionalne kulture prema tematskim poljima. In P. Piper & V. Jovanović (eds.) *Slovenska terminologija danas*, Beograd: SANU, pp. 613-623.
- Meyer, Ch. F. (2004). *English Corpus Linguistics. An Introduction*. Cambridge: Cambridge University Press.
- Miloradović, S. (2012). Lingvistički atlas – „centralni instrument“ savremene dijalektologije. In *Zbornik radova*

- etnografskog instituta SANU: Terenska istraživanja – poetika susreta*, 27, pp. 141-151.
- MSD (1967–1976). *Rečnik srpskohrvatskoga književnog jezika*, I–VI, Novi Sad (– Zagreb): Matica srpska (– Matica hrvatska).
- Otašević, Đ. (1999). *Rečnik novih i nezabeleženih reči*. Beograd: Enigmatski savez Srbije.
- Otašević, Đ. (2008). Nove reči i značenja u savremenom standardnom srpskom jeziku. Beograd: Alma.
- Otašević, Đ. (2008a). *Rečnik novih reči*. Beograd: Alma.
- Otašević, Đ. & Sikimić, B. (1991). Odnos okazionalizama prema vremenu. In *Naš jezik*, XXIX, 1–2, pp. 77-81.
- Ristić, S. (2006). *Raslojenost leksike srpskog jezika i leksička norma*. Beograd: Institut za srpski jezik SANU.
- Ristić, S. (2012). *O rečima u srpskom jeziku*. Beograd: Institut za srpski jezik SANU.
- Rundell, M. (2012). The wisdom of crowds: can it work for dictionaries? Accessed at: <http://www.macmillandictionaryblog.com/the-wisdom-of-crowds-can-it-work-for-dictionaries> [05/15/2020].
- SASAD (1959–2019). Речник српскохрватског књижевног и народног језика, I–XXI, Београд: САНУ.
- Schicchi, D., Pilato, G. (2018). WORDY: A Semi-automatic Methodology Aimed at the Creation of Neologisms Based on a Semantic Network and Blending Devices. In L. Barolli, O. Terzo (eds.) *Complex, Intelligent, and Software Intensive Systems*. CISIS 2017. Advances in Intelligent Systems and Computing, vol. 611. Springer, Cham. Accessed at: https://link.springer.com/chapter/10.1007/978-3-319-61566-0_23 [05/15/2020].
- Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.
- Sköldberg, E. & Wenner, L. (2020). Folkmun.se: A Study of a User-Generated Dictionary of Swedish. In *International Journal of Lexicography*, 33(1), pp. 1-16.
- Stanković, R., Krstev, C., Obradović, I. Lazić, B. & Trtovac, A. (2016). Rule-based Automatic Multi-word Term Extraction and Lemmatization. In N. Calzolari et al. (eds.) *Proceedings of the 10th International Conference on Language Resources and Evaluation, LREC 2016, Portorož, Slovenia, 23–28 May 2016*, pp. 507-514.
- Šandrih, B., Krstev, C. & Stanković, R. (2020). Two approaches to compilation of bilingual multi-word terminology lists from lexical resources. In *Natural Language Engineering*. Cambridge: Cambridge University Press. DOI: 10.1017/S1351324919000615.
- Tasovac, T., Petrović, S. (2015). Multiple Access Paths for Digital Collections of Lexicographic Paper Slips. In *eLex 2015 – Electronic Lexicography in the 21st Century: Linking Lexical Data in the Digital Age*. Accessed at: <https://elex.link/elex2015/conference-proceedings/paper-25/> [05/15/2020].
- Trap-Jensen, L. (2020). Language-Internal Neologisms and Anglicisms: Dealing with New Words and Expressions in The Danish Dictionary. In *Dictionaries: Journal of the Dictionary Society of North America*, 41(1), 11-25. doi:10.1353/dic.2020.0002.
- Varantola, K. (2002). Use and Usability of Dictionaries: Common Sense and Context Sensibility? In M. Corraed (ed.) *Lexicography and Natural Language Processing: A Festschrift in Honour of B.T.S. Atkins, Euralex 2002*, pp. 30-44.
- Vasić, V., Prčić, T. & Nejgebauer, G. (2001). *Rečnik novijih anglicizama*. Novi Sad: Zmaj.
- Благоева, Д. (2011). Принципи за съставяне на словник за неологичен речник. In: *Многообразие в единството*, 1, pp. 17-21.
- Благоева, Д. (2016). Славянската неография – минало и настояще. In: *За словото – нови търсения и подходи: Юбилеен сборник в чест на чл.-кор. проф. д.ф.н. Емилия Пернишка*. Издателство на БАН „Проф. Марин Дринов“, pp. 200-210.
- Котелова З. Н. (1978). Первый опыт лексикографического описания русских неологизмов. In Н. З. Котелова (ed.) *Новые слова и словари новых слов*. Ленинград, pp. 5-26.
- Плотникова А.А. (2009). *Материалы для этнолингвистического изучения балканословянского ареала* (Plotnikova A.A. Materijali za etnolingvističko proučavanje balkansko-slovenskog areala / Prevod M. Pić.) М: Институт славяноведения РАН. (1st ed. 1996).
- Поликарпов А.А. (2013). Модель жизненного цикла знака: к теоретическим основаниям исторической лексикологии и дериватологии. In: М. И. Чернышева (ed.) *Славянская лексикография. Международная коллективная монография*, М.: Издательский центр «Азбуковник», pp. 679-702.
- Поликарпов, А.А., Кукушина, О. В. & Токтонов, А. Г. (2008). Проверка теоретически предсказанных неодериватологических закономерностей данными русской корпусной неодериватографии. In: М. И. Чернышева (ed.) *Теория и история славянской лексикографии, Научные материалы к XIV съезду славистов*, Москва, pp. 392-426.

Acknowledgements

This paper has been developed as a result of the financing Contract 451-03-68/2020-14/200174 with the Ministry of Education, Science and Technological Development of the Republic of Serbia. The research is also related to the activities of the Cost Action CA16105 enetCollect *European Network for Combining Language Learning with Crowdsourcing Techniques*, to whom we, by this occasion, express our sincere gratitude.