Abstrakt:

Responsive dictionaries: bringing dictionaries (and lexicographers) and users closer together

Iztok Kosem

A recent study on the use of (monolingual) dictionaries across Europe (Kosem et al. 2019) has shown that up-to-date content is among the top three dictionary features valued by users. In fact, in countries where existing dictionaries are rather outdated, the users rank up-to-date content as high as reliability, sometimes even higher. The message is clear – users want dictionaries that reflect contemporary language and that react to the changes in language as quickly as possible. The advances in corpus tools and related lexicographic tools (e.g. tools for neologism detection) have enabled lexicographers to address these user demands rather successfully; however, this is mainly the case for languages with good existing dictionary infrastructures, such as English, German etc.

However, in countries such as Slovenia, one of the main problems is the outdatedness of dictionaries, and even the non-existence of, or poor accessibility to certain, more specialised dictionaries. This means that these dictionaries need to be made from scratch, which would traditionally also mean years (even decades) of waiting for the users. In order to address this issue, the team at the Centre for Language Resources and Technologies at the University of Ljubljana has tested and implemented several state-of-the-art approaches into the lexicographic process, such as automatic extraction of lexical data from the corpus (Gantar et al. 2016; Kosem et al. 2013), and crowdsourcing (Čibej et al. 2015). Furthermore, we have used semi-automatic methods for the creation of language resources such as Thesaurus of Modern Slovene (Krek et al. 2017a).

The implementation of the methods mentioned above had one important consequence for the dictionary-making process; namely, the data obtained with (semi-)automatic approaches was already of such quality that we decided to present it to the users. Two dictionaries, *Thesaurus of Modern Slovene* (Krek et al. 2017b) and *Collocations Dictionary of Modern Slovene* (Kosem et al. 2018), were published in this manner. We named them responsive dictionaries as they are able to quickly "respond" to changes, either because of completed lexicographic work and/or availability of new data (e.g. corpora). Another form of responsiveness is related to the users, as we offered them the opportunity to be involved in the dictionary-making process via voting on particular types of information, suggesting new content, or via specific crowdsourcing tasks.

In my talk, I will present both dictionaries, paying particular attention to the lexicographic process behind both dictionaries, and the contributions and feedback received from the users. In addition, as the dictionaries share certain types of data (collocations, examples), I will show how they were integrated into a large database where all the information from the two dictionaries, as well as from

other resources (including crowdsourcing tasks), is combined. I will also discuss some of the solutions that had to be implemented in the dictionary interface to provide the users with the necessary information about the reliability of dictionary data, and to enable their participation in a straight-forward manner. I will argue that the adoption of such a model of dictionary publication brings several benefits to both users and lexicographers. In conclusion, I will present future plans, including the implementation of various methods that are being developed within the ELEXIS (European Lexicographic Infrastructure) project.

References

Čibej, Jaka, Darja Fišer & Iztok Kosem (2015): The role of crowdsourcing in lexicography. In: Kosem, Iztok et al. (eds.): *Electronic lexicography in the 21st century: linking lexical data in the digital age*. Ljubljana: Trojina, Institute for Applied Slovene Studies; Brighton: Lexical Computing, pp. 70-83. <www.elex.link/elex2015/proceedings/eLex_2015_05_ Cibej+Fiser+Kosem.pdf>

Gantar, Polona, Iztok Kosem & Simon Krek (2016): Discovering automated lexicography = the case of Slovene lexical database. In: *International journal of lexicography* 29/2, pp. 200-225. <www.academic.oup.com/ijl/article/29/2/200/2413284/Discovering-AutomatedLexicography-The-Case-of-the?guestAccessKey=95f18766-f10f-4994-a6fa-448cf75ac55e>

Kosem et al. (eds.) (2017): *Kolokacije 1.0: Collocations Dictionary of Modern Slovene*. <viri.cjvt.si/kolokacije>

Kosem et al. (2019): The image of the monolingual dictionary across Europe. Results of the European survey of dictionary use and culture. In: *International Journal of Lexicography* 32/1, pp. 92-114. <academic.oup.com/ijl/article-abstract/32/1/92/5237031?redirectedFrom=fulltext>

Kosem, Iztok, Polona Gantar & Simon Krek (2013): Avtomatizacija leksikografskih postopkov. In: Erjavec, Tomaž & Jerneja Žganec Gros (eds.): *Jezikovne tehnologije, Slovenščina 2.0, 1/2*. Ljubljana: Trojina, zavod za uporabno slovenistiko, pp. 139-164. <www.trojina.org/slovenscina2.0/arhiv/2013/2/Slo2.0_2013_2_07.pdf>

Krek et al. (eds.) (2017b): *Sopomenke 1.0: Thesaurus of Modern Slovene*. <viri-cjvt.si/sopomenke>

Krek, Simon, Cyprian Laskowski & Marko Robnik-Šikonja (2017a): From translation equivalents to synonyms: creation of a Slovene thesaurus using word co-occurrence network analysis. In: Kosem, Iztok et al. (eds.): P*roceedings of eLex 2017: Lexicography from Scratch, 19-21 September 2017, Leiden, Netherlands*. <www.elex.link/elex2017/wp-contentuploads/2017/09/paper05.pdf>

Iztok Kosem
Jožef Stefan Institute (ELEXIS project)
Centre for Language Resources and Technologies,
University of Ljubljana
iztok.kosem@ijs.si