

Iztok Kosem

## THE COMPREHENSIVE SLOVENIAN-HUNGARIAN DICTIONARY: BILINGUAL LEXICOGRAPHY MEETS MONOLINGUAL LEXICOGRAPHY

**Keywords** Bilingual dictionary; Digital Dictionary Database; Slovenian; semantic concepts; database model

The last decade has been relatively quiet for bilingual lexicography in Slovenia, with only few bilingual dictionaries being published. The methodology has not changed much either, with the usual approach including the use of an existing bilingual dictionary database, and replacing the target language with new language translations. However, even this has become a problem, particularly when compiling a dictionary of Slovene to another target language, mainly due to the fact that there is currently no dictionary, bilingual or monolingual, that would satisfactorily describe contemporary Slovene.

In recent years, various developments in Slovenian and international lexicography have made it possible to start addressing this gap in Slovenian bilingual lexicography. Among the most important were the initiatives by linguists and politicians to start the work on a new Slovenian-Hungarian dictionary, an important education and cultural resource for the two neighbouring countries. The main result of the initiatives was a project funded by the Slovenian Research Agency and the Ministry of Education, Science and Sport of the Republic of Slovenia focussed on producing a concept for a Slovenian-Hungarian dictionary (Kosem et al. 2018b). The project provided a unique opportunity to thoroughly investigate the state-of-the-art in bilingual lexicography, and design and test the best methodology for the compilation of not only this particular dictionary, but any others that may benefit from its data. After the project ended, the work on the Comprehensive Slovenian-Hungarian Dictionary began in 2018 at the Centre for Language Resources and Technologies of the University of Ljubljana (CJVT UL). The dictionary compilation is co-funded by the Slovenian Research Agency (ARRS) through CJVT-dedicated funds as part of the Network of Infrastructure Centres at the University of Ljubljana.

In parallel, CJVT UL has been compiling various other dictionaries, e. g. the Collocations Dictionary of Modern Slovene (Kosem et al. 2018a) and the Thesaurus of Modern Slovene (Arhar et al. 2018). Moreover, CJVT UL has started developing a Digital Dictionary Database, which aims to become a one for-all database for the Slovenian language, to be used both in the compilation of language resources and for natural language processing tasks. The plans for the database have been described in detail in Klemenc et al. (2017). The main aspect of the Digital Dictionary Database relevant for this paper is that it intends to use the same concepts, which translate into dictionary senses, for all the resources coming out of the database. So any resource being compiled would use the senses of existing lexical units found in the database, or contribute new senses for lexical units not yet analysed. In terms of the Comprehensive Slovenian-Hungarian Dictionary, this meant that the workflow was split into two major stages: in the first stage the (monolingual) entries are compiled com-

pletely from scratch, with a great amount of contextual information (collocations, examples etc.) being provided. In the second stage the translation into Hungarian is conducted, with the lexicographers-translators translating the senses and selected parts of contextual information. By using this approach, we are excluding target language bias (i. e. providing only contextual data of contrastive interest for a particular language pair) and creating a database that can be used in the compilation of language resources of other Slovenian-foreign language pairs.

In October 2021, version 1.0 of the Comprehensive Slovenian-Hungarian Dictionary (Kosem et al. 2021a) was published, containing 10,946 headwords, 33,298 translations, 15,265 collocations and other word combinations, and 2,416 examples. The Comprehensive Slovenian-Hungarian dictionary is a growing dictionary, which means that new headwords will be added in regular intervals. Importantly, the dictionary database is available under CC BY-SA 4.0 license in the CLARIN.SI repository (Kosem et al. 2021b). In our presentation, we will discuss the lexicographic workflow, the steps and recording of data relevant for other bilingual and monolingual dictionaries, and some of the challenges of adopting the same sense structure for all the dictionaries in the Digital Dictionary Database. We will also give a short demonstration of the dictionary interface, and present future plans, especially those related to improving the efficiency of lexicographic workflow.

## References

- Arhar Holdt, Š./Čibej, J./Dobrovoljc, K./Gantar, P./Gorjanc, V./Klemenc, B./Kosem, I./Krek, S./Laskowski, C./Robnik Šikonja, M. (2018): Thesaurus of modern Slovene: by the community for the community. In: Čibej, J./Gorjanc, V./Kosem, I./Krek, S. (eds.): Proceedings of the XVIII EURALEX International Congress: Lexicography in Global Contexts. Ljubljana, pp. 401–410.  
<https://e-knjige.ff.uni-lj.si/znanstvena-zalozba/catalog/view/118/211/3000-1> (last access: 25-04-2022).
- Klemenc, B./Robnik-Šikonja, M./Fürst, L./Bohak, C./Krek, S. (2017): Technological design of a state-of-the-art digital dictionary. In: Gorjanc, V./Gantar, P./Kosem, I./Krek, S. (eds): Dictionary of modern Slovene: problems and solutions. Ljubljana, pp. 10–22.
- Kosem, I./Krek, S./Gantar, P./Arhar Holdt, Š./Čibej, J./Laskowski, C. (2018a): Collocations dictionary of modern Slovene. In: Čibej, J./Gorjanc, V./Kosem, I./Krek, S. (eds.): Proceedings of the XVIII EURALEX International Congress: Lexicography in Global Contexts. Ljubljana, pp. 989–997.  
<https://e-knjige.ff.uni-lj.si/znanstvena-zalozba/catalog/view/118/211/3000-1> (last access: 25-04-2022).
- Kosem, I./Bálint Čeh, J./Gorjanc, V./Kolláth, A./Kovács, A./Krek, S./Novak-Lukanovič, S./Rudaš, J. (2018b): Osnutek koncepta novega velikega slovensko-madžarskega slovarja. Ljubljana.  
<https://www.cjvt.si/komass/wp-content/uploads/sites/17/2020/08/Osnutek-koncepta-VSMS-v1-1.pdf> (last access: 25-04-2022).
- Kosem, I. et al. (2021a): Comprehensive Slovenian-Hungarian Dictionary. Ljubljana.
- Kosem, I. et al. (2021b): Comprehensive Slovenian-Hungarian Dictionary 1.0, Slovenian language resource repository CLARIN.SI. <http://hdl.handle.net/11356/1453> (last access: 25-04-2022).

## Contact information

### Iztok Kosem

Faculty of Arts, University of Ljubljana  
 iztok.kosem@cjvt.si