Ivana Filipović Petrović

A CORPUS-DRIVEN APPROACH TO LEXICOGRAPHIC DEFINITIONS: THE REPRESENTATION OF MEANING IN THE ELECTRONIC *DICTIONARY OF CROATIAN IDIOMS*

Keywords Corpus-driven approach; idioms; representation of meaning in the dictionary; defining strategies; true electronic features

One of the major concerns in lexicography, both paper and electronic, are defining strategies. Several studies investigated what dictionary users want most from their dictionaries and the results showed that finding meaning is among their primary needs (cf. Wingate 2002; Tarp 2009; Lew 2010). The notion of meaning is especially relevant in the case of idioms[1] – conventionalized multiword expressions with figurative meaning, or traditionally, meaning which is 'not the sum of its parts' (see Fernando 1996; Moon 1998). When it comes to the lexicographic treatment of idioms, there were two major milestones: computer corpora, given that corpus data provide many examples of real usage and context in which idioms tend to occur, and digital medium, which increased search options and eliminated space constraints.

In this text we are presenting functionalities that contribute to the representation of meaning in the electronic corpus-driven *Dictionary of Croatian Idioms*, which is currently available in a beta version on the Lexonomy platform. The macrostructure is organized as follows: the headword is the most frequent variant form of the idiom, i. e., the whole construction. The dictionary entry also contains other variant forms of the idiom, explanation in the form of a reduced sentence (which is best suited to Croatian) and examples of use. Boxes with additional information regarding usage or wordplay are placed at the bottom of entry. A search box gives results if the user inserts any component of an idiom, as well as any variant component which is noted in the entry. It is a true electronic feature (cf. Prinsloo/van Graan 2021), which significantly upgraded the issue of searching in regard to printed dictionaries.

In addition, given that Lexonomy enables cross referencing, idioms with a similar and/or opposite meaning are connected via hyperlinks, thus creating a conceptual network of idioms around a common concept, such as anger, mental condition, or happiness. Idioms are included in a conceptual network depending on their meaning, structure, and use, according to corpus data from the Croatian web corpus hrWaC. For example, idioms that are connected through the concept of anger are very frequent, so they are further grouped according to structural features which contribute to the meaning of the whole construction: one group contains idioms with adjectives and the meaning 'angry': *crven od bijesa* (lit. red with rage) and *ljut kao pas/ris* (lit. angry as a lynx/dog), and the other group contains idioms with perfective verbs and the meaning 'get angry': *pozelenjeti od bijesa* (lit. turn green with rage) and *pao je mrak na oči komu* (lit. darkness fell on someone's eyes).

Furthermore, according to the corpus data, significant number of highly flexible idioms occur with a common lexis, but different grammatical forms (*biti u škripcu* 'be in a corner',

[1] In this paper we use the term 'idiom' in its narrower uses, as a translation equivalent for the Croatian word 'frazem'.

This paper is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.

doći u škripac 'get into a corner', *dovesti u škripa koga* 'back someone into the corner', *izvući se iz škripca* 'get out of a corner'). In the *Dictionary of Croatian Idioms*, they are treated as variations and are listed in a single entry, in line with the cognitive linguistic view that variations present the same event in different ways (Langlotz 2006; Parizoska/Omazić 2020). The defining technique implemented here includes a definition of the first, most frequent variant listed in the entry: 'to be in a difficult situation', and other variants are shown underneath with examples of use. Leaving the definition out, we emphasize the role of a variant form which shows lexico-grammatical pattern of use and the role of examples as well, in order to provide not only decoding information in the *Dictionary*, but also encoding. Although studies on the role of examples in language production (Frankenberg-Garcia 2014, 2015) point out that examples that help with encoding and decoding simultaneously are difficult to find, this is exactly what we do. By careful hand-picking, we are looking for examples that bring sufficient context for comprehension and characteristic collocation at the same time.

Overall, this text makes two contributions. Firstly, it shows the connection between corpus data and meaning representation in the electronic *Dictionary of Croatian Idioms*, which is directly reflected in defining strategies. Secondly, it shows that the representation of meaning is organized using available digital functionalities in order to create an up-to-date user-friendly dictionary of Croatian figurative language.

References

Fernando, C. (1996): Idioms and idiomaticity. Oxford.

Frankenberg-Garcia, A. (2014): The use of corpus examples for language comprehension and production. In: ReCALL 26, pp. 128–146.

Frankenberg-Garcia, A. (2015): Dictionaries and encoding examples to support language production. In: International Journal of Lexicography 28 (4), pp. 490–512.

Langlotz, A. (2006): Idiomatic creativity: a cognitive-linguistic model of idiom-representation and idiom-variation in English. Amsterdam.

Lew, R. (2010): Multimodal lexicography: the representation of meaning in electronic dictionaries. In: Lexikos 20, pp. 290–306.

Moon, R. (1998): Fixed expressions and idioms in English. A corpus-based approach. Oxford.

Parizoska, J./Omazić, M. (2020): Sheme dinamike sile i promjenjivost glagolskih frazema. In: Jezikoslovlje 20 (2), pp. 179–205.

Prinsloo, D./van Graan, N. D. (2021): Principles and practice of cross-referencing in paper and electronic dictionaries with specific reference to African languages. In: Lexicography: Journal of ASIALEX 8 (1), pp. 32–58.

Tarp, S. (2009): Reflections on lexicographical user research. In: Lexikos 19, pp. 275–296.

Wingate, U. (2002): The effectiveness of different learners dictionaries. Tübingen.

Contact information

Ivana Filipović Petrović
Linguistic Research Institute
Croatian Academy of Sciences and Arts
ifilipovic@hazu.hr

This paper is part of the publication: Klosa-Kückelhaus, Annette/Engelberg, Stefan/Möhrs, Christine/Storjohann, Petra (eds.) (2022): Dictionaries and Society. Book of Abstracts of the XX EURALEX International Congress. Mannheim: IDS-Verlag.