Dominika Kováříková/Michal Škrabal

THE DICTIONARY OF CZECH CORE ACADEMIC VOCABULARY

Keywords Academic word list, core academic vocabulary, modularity, Akalex

Over the past two decades, several lists of academic words and academic phrases emerged, often focused on L2 teaching or undergraduate students of academic writing classes (Coxhead 2000; Paquot 2010; Gardner/Davies 2014; Morley 2014). Most of the lists are focused on English as a lingua franca of science and research but with time, lists for other languages, such as Portuguese (Baptista et al. 2010), Swedish (Carlund et al. 2012) or Czech (Kováříková/Kovářík 2021), have been created too.

Academic word list by itself is a powerful tool for both teachers and students, but it can be further enhanced by additional information about the headword and its context as offered by a dictionary. Currently, an online dictionary of core Czech academic vocabulary is being developed, which will contain modules with information relevant to various target groups: undergraduate students (and possibly high school students), philology students, academic writers (professional and in training), and university students of Czech as a second language. Depending on the erudition level, on the field of study, or the specific task, the user will be able to adjust the content of the dictionary by choosing which of the modules will be displayed.

The following information will be included in the modules: 1. frequency information, 2. link to the corpus concordance substituting exemplification, 3. meaning(s) in academic texts, 4. etymology (if relevant, especially for loanwords), 5. common academic collocations including combinations with typical function words, linked to the relevant corpus concordance, 6. synonyms typical for academic texts, and opposites (if relevant), 7. derived words typical for academic texts, 8. translation equivalents in English, 9. translation equivalents in other languages.

Since this is quite an ambitious task, we decided to focus on one specific type of user in this study and elaborate only the modules that we expect would be chosen by an undergraduate student in an academic writing class. Among the modules relevant to this lower-level user are: frequency information, link to corpus concordance, meaning definition(s), etymology, common collocations, and synonyms.

A detailed analysis of several headwords of various word classes focuses on the meaning description module and on the procedure of finding typical collocations and appropriate synonyms. We discuss the benefits and drawbacks of two types of definitions for this specific dictionary (Aristotelian genus-differentia definition vs. full sentence as provided in Collins COBUILD dictionaries). Further, we discuss the possibility of using existing resources for creating the definitions of meaning, such as monolingual and bilingual dictionaries (e.g. Kraus et al. 1995; Sinclair et al. 1998), and the Frequency dictionary of Czech (Čermák/Křen 2010) which can serve as a defining vocabulary.

The dictionary is based on the Czech list of academic words and phrases that has been published recently as a part of an online application Akalex (Kováříková/Kovářík 2021). The list

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.

contains approximately 1,000 single-word and multi-word expressions, and it is based on frequency and distribution criteria similar in some respects to other academic word lists. The material for the academic word list is data from two representative corpora of contemporary written Czech SYN2015 and SYN2020. The words and multi-word units that have been included in Akalex are significantly more common in academic than non-academic texts, they are relatively frequent in academic texts, and are attested and evenly distributed in at least 21 of 24 academic disciplines available in the corpus material. These relatively simple criteria produced outstanding and convincing results comparable to other lists of academic words in size as well as in content (namely the Academic Keyword List by Paquot 2010).

For compiling the dictionary, we utilize some of the online corpus tools available at the Czech National Corpus web page (www.korpus.cz). Apart from the corpus manager KonText (Machálek 2014), which is a primary tool for examining the context of the lemma and for finding collocations, we use the database of translation equivalents Treq (Vavřín/Rosen 2015). Treq can be used not only to search for relevant equivalents in English and other languages but also for finding synonyms through the translation of various equivalents (as suggested by Čibej/Holdt, 2019). Another application, Word at a Glance (Machálek 2020), provides a basic overview of the searched word including collocations and similarly used words.

References

Baptista, J./Costa, N./Guerra, J./Zampieri, M. (2010): P-AWL: Academic Word List for Portuguese. In: Proceedings of Computational Processing of the Portuguese Language, PROPOR 2010, Porto Alegre, Brazil, pp. 120–123.

Carlund, C./Jansson, H./Johansson Kokkinakis, S./Prentic, J./Ribeck, J. (2012): An academic word list for Swedish – a support for language learners in higher education. In: Proceedings of the SLTC 2012 workshop on NLP for CALL, pp. 20–27.

Čermák, F./Křen, M. (2010): Frequency dictionary of Czech: core vocabulary for learners. London.

Čibej, J./Holdt, Š. A. (2019): Repel the syntruders! A crowdsourcing cleanup of the Thesaurus of Modern Slovene. In: Kosem, I. et al. (eds.): Electronic Lexicography in the 21st Century. Proceedings of the eLex 2019 Conference, 1–3 October 2019, Sintra, Portugal. Brno: Lexical Computing CZ, pp. 338–356.

Coxhead, A. (2000): A new academic word list. In: TESOL Quarterly 34 (2), pp. 213–238.

Gardner, D./Davies, M. (2014): A new academic vocabulary list. In: Applied Linguistics 35 (3), pp. 305–327.

Kováříková, D./Kovářík, O. (2021): Akalex: Czech Academic Word List. Prague: Institute of the Czech National Corpus. www.korpus.cz/akalex (last access: 10-01-2022).

Petráčková, V./Kraus, J. (1995): Akademický slovník cizích slov. Prague.

Křen, M. et al. (2015): SYN2015 – representative corpus of contemporary written Czech. Prague: Institute of the Czech National Corpus. www.korpus.cz (last access: 10-01-2022).

Křen, M. et al. (2020): SYN2020 – representative corpus of contemporary written Czech. Prague: Institute of the Czech National Corpus. www.korpus.cz (last access: 10-01-2022).

Machálek, T. (2014): KonText – corpus query interface. Prague: Institute of the Czech National Corpus. kontext.korpus.cz (last access: 10-01-2022).

Machálek, T. (2020): Word at a glance: modular word profile aggregator. In: Calzolari, N. et al. (eds.): Proceedings of the 12th Language Resources and Evaluation Conference. Marseille, pp. 7011–7016.

Morley, J. (2014): Academic phrasebank: a compendium of commonly used phrasal elements in academic English in PDF format. Manchester.

Paquot, M. (2010): Academic vocabulary in learner writing: from extraction to analysis. London/New York.

Sinclair, J. et al. (1998): Anglicko-český výkladový slovník. Prague: Nakladatelství Lidové noviny.

Vavřín, M./Rosen, A. (2015): Treq – database of translation equivalents. FF UK. Prague. https://treq.korpus.cz/. (last access: 10-01-2022).

Contact information

Dominika Kovaříková
Institute of the Czech National Corpus, Charles University, Prague
dominika.kovarikova@ff.cuni.cz

Michal Škrabal
Institute of the Czech National Corpus, Charles University, Prague
michal.skrabal@ff.cuni.cz

Acknowledgements

The paper was supported by the European Regional Development Fund-Project “Creativity and Adaptability as Conditions of the Success of Europe in an Interrelated World” (No. CZ.02.1.01/0.0/0.0/16_019/0000734).

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.