
zbirka
trojinski
konj

6 Summary

The development of Slovenian terminologies of different professions has been intense especially since the 18th century. Glosses in *The Stična manuscript* can be regarded as the beginning of the Slovenian lexicography and the Scientific *Terminology with a special regard to secondary schools* published in 1880 by Matej Cigale as the first Slovenian terminological dictionary.

The creation of terminological dictionaries differs from making of general language dictionaries in several segments; therefore, it is reasonable to distinguish between terminography and lexicography. Cognitions of Slovenian pre-corpus and partially corpus terminography were presented in four dictionaries: *Slovenian electro-technical dictionary* (1957–), *Dictionary of military terms* (1977/2002), *Mountain terminological dictionary* (2002), and *Islovar* (2001–). These dictionaries were examined in three points: material for the dictionary, criteria for inclusion of lexis into headword list, and the data in a dictionary entry. Comparatively, we joined the *Slovenian-German dictionary* (1894/95) by Maks Pleteršnik and *Dictionary of the standard Slovenian language* (1970–1991).

The selected terminological dictionaries differ in their concepts. The Slovenian electro-technical dictionary is to a great extent an implementation of the international standard; it contains terms in four languages and their definitions. The concept of the Dictionary of military terms was made in 1970s, and was partially prompted by the need and decision to show that the Slovenian military language differs for the Serbo-Croatian; it contains several types of grammar information and translations into Serbo-Croatian. The Mountain terminological dictionary is a multilingual explanatory dictionary, containing five languages, that differs from the other three in larger inclusion of style and registry marks. The Islovar is created continuously online using a four-level editorial procedure and contains definitions and English translations. The authors are using a corpus of informatics as one of the sources of terms or as the source of data in the editorial procedure.

Pleteršnik's dictionary and the Dictionary of the standard Slovenian language are based on extensive and documented material, written out from texts that were chosen according to pre-determined criteria, while the reviewed terminological dictionaries have a weak material basis; the documented material was available only for a part of the Dictionary of military terms, and the corpus has been used for only a small part of new terms in the Islovar; the other two dictionaries are not supported by any printed or other source. Consequently, the frequency of occurrence could be one of the criteria for the inclusion of words into a headword list only in the Dictionary of military terms (partially) and the Islovar (also partially). With other dictionaries, the decision for (non)inclusion of terms into headword list was made by members of the editorial board or was based on foreign dictionaries (the Slovenian electro-technical dictionary). The review also showed that Slovenian terminological dictionaries contain a lot of entry information. The Dictionary of military terms stands out with phrases, which we interpreted as tendency to show typical context of terms, although in the dictionary those phrases are not clearly separated from multi-word military terms.

In the applicative part of the research we built a specialized corpus: a corpus of texts in the field of public relations, to be used as a source of terms for a terminological database of the profession. We initially defined the domain and criteria for the selection of texts, and described the procedure of collection. The criteria were: text type, size, completeness, authorship, readership, mode, published vs. non published, time, text origin, authenticity, domain boundaries, and copyrights. We made the scheme of the document's header and included all the information we considered relevant for documenting dictionary material or as potentially relevant parameters for other lexical and textual research. Lemmatised and morpho-syntactically tagged corpus of public relations KoRP is freely available online since July 2007. It contains 1.8 million words and is monolingual, synchronic, written and static corpus of professional texts. Searching the corpus can be conducted via the Amebis Concordance programme ASP32. From 2013 the KoRP corpus has been available in the NoSketch Engine and CUWI concordancer at *http://nl.ijs.si* as well.

Corpus-based dictionaries use frequency as the main criterion for inclusion of words into the headword list. A corpus can be used to obtain various lists of words, the frequency list being among the most important ones. The comparison of the frequency list of the KoRP corpus and the frequency list of the reference corpus of Slovene Gigafida has shown that among thirty most frequent lemmas from the KoRP corpus there are six nouns that are lower on the frequency list of the Gigafida corpus and that all of these nouns are terms in the field of public relations.

We also conducted an automatic term extraction using the LUIZ term extraction tool. We extracted from the KoRP corpus: (a) single-word term candidates (nouns, verbs, adjectives, and adverbs); and (b) multi-word term candidates (noun phrases and verb phrases). Both single- and multi-word term candidates were extracted using morphosyntactic patterns and term weights, calculated by comparing the frequency in the KoRP corpus and the frequency in a reference corpus (in our case, the reference corpus of Slovene FidaPLUS), and considering phraseological stability of an extracted terminological unit.

We have used 39 morphosyntactic patterns in total: 30 with noun as a headword, 9 with verb as a headword. Our results show that the terminological relevance of extracted nouns is indeed higher than that of merely frequent nouns, and that verbal phrases are rarely proper terms, but can be often listed as collocations within other entries. Terminologically most productive patterns are those for multi-word terms with noun as a headword: [adjective + noun], [adjective + *and* + adjective + noun] and [adjective + adjective + noun].

Apart from the headword list, we paid special attention to three parts of a dictionary entry: typical context of words, definition, and norm. For the dictionary database, we specified two rubrics for recording typical context: collocations and examples of use.

The method used for extracting lexical information (syntactic relations, collocations, and examples) for single and multi-word terms uses the Sketch Engine tool and its Word sketch function, so we had to prepare and upload the KoRP corpus in our local installation of the Sketch Engine. Some changes had to be made to the extraction algorithm and its constituent parts. For example, sketch grammar had to be slightly adapted, new GDEX configurations for good example extraction had to be prepared, and minor tweaks to API script had to be made. In addition, a new DTD for the Termania dictionary portal had to be prepared to enable importing of information in the database, as well as its visualisation. Using this approach, we avoided manual corpus analysis for nearly 2000 terms and reduced the time of editing the terminological database.

Extracted lexical information was imported into the dictionary editor of the Termania portal where the rest of the editing was performed.

The terminological database for the field of public relations contains over 2000 terms with information on norm, explanations, English translations, typical collocations and examples. Each entry contains links with related entries, as well as links to the concordances in the KoRP corpus and the Gigafida corpus. The database is freely accessible online at *http://www.termania.net.*

7 Literatura

Adamič, Š., in dr., ur. (2002): *Slovenski medicinski slovar*. Ljubljana: Medicinska fakulteta.

Aijmer, K., Altenberg, B., ur. (1991): *English corpus linguistics*. London, New York: Longman.

Arhar Holdt, Š. (2011): *Luščenje besednih zvez iz besedilnega korpusa z uporabo dvodelnih in tridelnih oblikoskladenjskih vzorcev.* Ljubljana: Trojina, zavod za uporabno slovenistiko.

Arhar, Š. (2004): *Gradnja specializiranega korpusa: diplomsko delo*. Ljubljana: Filozofska fakulteta.

Arhar, Š., Gorjanc, V. (2007): Korpus FidaPLUS: nova generacija slovenskega referenčnega korpusa. *Jezik in slovstvo* 52/2. 95–110.

Atkins, S. B. T., Clear, J., Ostler, N. (1992): Corpus design criteria. *Literary and linguistic computing* 7/1. 1–16.

Atkins, S. B. T., Rundell, M. (2008): *The Oxford guide to practical lexicography*. Oxford: Oxford University Press.

Batis, J., in dr., ur. (1982–). *Veterinarski terminološki slovar*. Ljubljana: Založba ZRC, ZRC SAZU.

Béjoint, H. (2004): *Modern lexicography: an introduction*. Oxford: Oxford University Press.

Beran, J., in dr., ur. (1999): *Pravni terminološki slovar: do 1990, gradivo*. Ljubljana: Založba ZRC, ZRC SAZU.

Berčič, B., in dr., ur. (1996): *Bibliotekarski terminološki slovar: poskusni snopič.* Ljubljana: NUK.

Bergenholtz, H., Nielsen, S. (2006): Subject-field components as integrated parts of LSP dictionaries. *Terminology* 12/2. 281–303.

Bergenholtz, H., Tap, S., ur. (1995): *Manual of specialised lexicography*. Amsterdam, Philadelphia: John Benjamins.

Biber, D. (1993): Representativness in corpus design. *Literary and linguistic computing* 8/4. 243–257.

Biber, D. (2009): A corpus-driven approach to formulaic language in English: multi-word patterns in speech and writing. *International journal of corpus linguistics* 14/3. 275–311.

Biber, D., Conrad, S., Reppen, R. (1998): *Corpus linguistics: investigating language structure and use*. Cambridge: Cambridge University Press.

Bosanac, M., Mandić, O., Petkovič, S., ur. (1977): *Rječnik sociologije i socijalne psihologije*. Zagreb: Informator, Izdavačka kuća.

Bowker, L. (1996): Towards a corpus-based approach to terminology. *Terminology* 3/1. 27–52.

Bowker, L., Pearson, J. (2002): *Working with specialized language*. London, New York: Routledge.

Burnik, V. (2000): *Terminološki slovarji – kritike zasnov: diplomska naloga*. Ljubljana: Filozofska fakulteta.

Caruso, V. (2011): Online specialised dictionaries: a critical survey. *Proceedings of eLex 2011*. Ljubljana: Trojina, zavod za uporabno slovenistiko. 66–75.

Chromá, M. (2006): Synonymy and polysemy in a bilingual law dictionary. *Proceedings of the XII EURALEX international congress*. Torino: Edizioni dell'Orso. 735–741.

Church, K., Hanks, P. (1990): Word associations norms, mutual information and lexicography. *Proceedings of the 27th annual conference of the Association for Computational Linguistics*. Vancouver: Association for Computational Linguistics. 76–82.

Church, K., in dr. (1991): Using statistics in lexical analysis. U. Zernik, ur.: *Lexical acquisition*. Englewood Cliff, NJ: Erlbaum. 115–64.

Clear, J. (1993): From Firth principles: computational tools for the study of collocation. M. Baker, G. Francis, E. Tognini - Bonelli, ur.: *Text and technology: in honour of John Sinclair*. Amsterdam: John Benjamins. 271–292.

Corréard, M. (2002): Are space-saving strategies relevant in electronic dictionaries? *Proceedings of the 10th EURALEX international congress*. Copenhagen: Center for Sproktehnologi. 463–470.

Čermák, F. (2005/1995): Jezikovni korpus: sredstvo in vir spoznanj. V. Gorjanc, S. Krek, ur.: *Študije o korpusnem jezikoslovju*. Ljubljana: Krtina. 137–171./*Slovo a slovesnost* 56. 119–140.

Dobnikar, M., in dr., ur. (2005): *Gemološki terminološki slovar*. Ljubljana: Založba ZRC, ZRC SAZU.

Drapal, A. (2000): Prispevek h konstruktivnemu sovražnemu dialogu med odnosi z javnostmi in trženjem. *Akademija MM* 4/6. 77–79.

Drstvenšek, N. (2003): Vloga besedilnega korpusa pri postavitvi geselskega članka v enojezičnem slovarju. *Jezik in slovstvo* 48/5. 65–81.

Erjavec, T. (1996/97): Računalniške zbirke besedil. *Jezik in slovstvo* 42/2–3. 81–96.

Erjavec, T. (1998): Oznake korpusa FIDA. *Uporabno jezikoslovje* 6. 85–95.

Erjavec, T. (2003): Označevanje korpusov. *Jezik in slovstvo* 48/3–4. 61–76.

Erjavec, T. (2009): Odprtost jezikovnih virov za slovenščino. M. Stabej, ur.: *Infrastruktura slovenščine in slovenistike (Obdobja 28)*. Ljubljana: Znanstvena založba Filozofske fakultete. 115–121.