ADAM RAMBOUSEK[A], **ALEŠ HORÁK**[B], **KAREL PALA**[C]

Natural Language Processing Centre, Faculty of Informatics, Masaryk University, Brno, Czech Republic
[A]rambousek@fi.muni.cz ; [B]hales@fi.muni.cz ; [C]pala@fi.muni.cz

SUSTAINABLE LONG-TERM WORDNET DEVELOPMENT AND MAINTENANCE: CASE STUDY OF THE CZECH WORDNET

Abstract

Czech WordNet represents one of the first national wordnets created during the EuroWordNet and BalkaNet projects. However, the data contains various issues that affect the use of Czech WordNet in NLP applications. Since the publication of the first CzWN version, the semantic network was augmented in several phases, however, complex final editing and publishing process has not been finished. In 2017, we have started a project to evaluate and update the Czech WordNet, followed by a connection to the Collaborative Interlingual Index.

In this paper, we provide an overview of Czech WordNet data updates and extensions, and present the roadmap to publish a revised version of the Czech WordNet under open license. Moreover, we introduce a developed concept for long-term updates and maintenance of the data based on crowdsourcing activities.

Keywords: EuroWordNet; BalkaNet; wordnet; Czech WordNet; DEBVisDic

1   Introduction and history of the Czech WordNet

After its publication, the Princeton WordNet (PWN; Fellbaum, 1998) proved its usability as a lexical resource, both for common users and various NLP tasks. PWN also inspired many projects aiming either to create semantic networks in other languages, or to extend the wordnet with new features. The first major attempt to build localized wordnets was the EuroWordNet (Vossen, 1998) project started in 1996 and coordinated by Piek Vossen from the University of Amsterdam. In its first phase, EuroWordNet I included Dutch, Italian, Spanish, and English WordNets. In the next phase, EuroWordNet II, German, French, Estonian, and Czech WordNets were added.

EuroWordNet as a whole introduced two new features that were necessary for language compatibility. With the aim to build semantic networks in several languages that share the same language core, a list of Base Concepts was developed and described. The list included 1310 synsets shared

amongst all EuroWordNet languages and represented the part of wordnet that should have been encoded first. Another purpose of Base Concepts was to investigate and capture the individual linguistic differences among the languages.

Since the national wordnets reflect word stocks of various languages displaying specific lexical hierarchies, the Interlingual Index (ILI) was established within the EuroWordNet project. The index was based on a language-independent top ontology. Each wordnet connected its synsets to ILI, thus enabling to create multi-lingual links. The features and processes developed during the EuroWordNet project were later re-used in building other national wordnets.

One of such projects was the BalkaNet (Christodoulakis, 2004) project in 2001–2004, aiming to expand the number of national wordnets for six European languages. BalkaNet project covered Bulgarian, Greek, Romanian, Serbian, and Turkish wordnets. Together with newly developed wordnets, verb synsets in Czech WordNet were extended with valency frames.

As mentioned above, the Czech WordNet (CzWN) was created in EuroWordNet and BalkaNet projects by the Natural Language Processing Centre at the Faculty of Informatics, Masaryk University (NLP Centre). At the beginning, CzWN was published through the ELDA/ELRA agency under closed and paid license. Presently the situation has changed — it is possible to access the Czech WordNet data within the LINDAT/Clarin repository (see below) — thus it is now available in an open form.

Since 2004, there were several subprojects devoted to extending, fixing or updating the Czech WordNet data, which produced several extended datasets. In 2017, we have started a project to evaluate, update and consolidate the Czech WordNet. In the following sections, we present the details of the current state and the consolidation process.

2 Versions of the Czech WordNet

2.1 The original Czech WordNet

The original version of the Czech WordNet (Pala & Smrž, 2004; Horák & Smrž, 2004) is available for licensing from ELDA/ELRA. This is the version created during EuroWordNet and BalkaNet projects, and contains 28,201 synsets with 43,958 literals. All the synsets are linked to their counterparts in Princeton WordNet 2.0. Part of the verbal synsets (824) were also enriched with verb frames. It has to be added that this version was slightly modified (corrected) and is presently accessible in the LINDAT/Clarin repository which is placed at the UFAL MFF UK in Prague.[1]

The primary method for this wordnet creation was the top-down approach (proposed in the EuroWordNet project). Lexicographers consulted several resources, available at the time in electronic form — Czech explanatory dictionary (Filipec et al., 1995), English-Czech dictionary, Czech synonymy dictionary, and the DESAM corpus. Although the explanatory dictionary contained information about hypernyms for some headwords, this information was not entered systematically. This led to the solution that most of the hypernymic relations were directly transferred from the Princeton WordNet. Information on Czech synonyms was more extensive, however not covering all concepts needed. As a result, many synsets were exact translations of synsets from Princeton WordNet.

This approach caused various issues with the data. Most notable example are the synsets containing words that are not exactly synonyms, or only rare in the Czech language, but present in the Czech WordNet because of the translation from English. For example, English synset *cabriolet:1, cab:2* has the equivalent Czech synset *kabriolet:2, dvoukolový jednospřežní povoz:1, koňská drožka:1* (cabriolet, two-wheeled one horse cart, horse-drawn carriage). Although the translation is correct, this sense of *kabriolet* in Czech is very archaic, in current language the only sense used in spoken language is the *convertible car*. Another problem is the inclusion of multiword expressions

---

[1]See Czech WordNet 1.9 PDT `https://lindat.mff.cuni.cz/repository/xmlui/handle/11858/00-097C-0000-0001-4880-3`

in the synset which are not fixed lexical units in the Czech language (this may be justified in some cases).

2.2 2009 edited version

To deal with some of the issues mentioned above, core synsets of the Czech WordNet were edited by lexicographers in 2009. In total, 2,400 synsets from the Base Concept set were edited. The updates included synonyms revision and definition editing. Total number of synsets is the same (28,201). This version of Czech WordNet was not published, but is available for research.

2.3 Extension with bilingual dictionary

To increase the lexical coverage of the Czech WordNet, a semi-automatic method was proposed in 2011 (Blahuš & Pala, 2012). We acquired machine-readable data from the largest one-volume English-Czech dictionary ever published. It contains more than 100,000 headwords and sub-headwords, more than 200,000 words and phrases and roughly 400,000 equivalents. We used the following algorithm to add new words and synsets:

- Extract translation pairs from the dictionary.

- Keep only pairs in which English literals are monosemous.

- If desired, keep only pairs with unique source literals (one-to-one translations).

- Match English literals with monosemous PWN literals.

The extended version of the Czech WordNet contains 83,769 literals (growth of 76 %) organized into 40,621 synsets (growth of 43 %). Out of the synsets, 27,658 are noun synsets (increase of 6,640, or 31.6 %), 5,852 are verb synsets (increase of 690, or 13.3 %), 5,651 are adjective synsets (increase of 3,522, or 165 %) and 1,457 are adverb synsets (increase of 1,291, or 878 %).

Because of the unsupervised nature of the extension, the newly produced Czech WordNet data need to be inspected manually. We have checked a sample of 600 synsets, with the results that 30 % of the synsets contain wrong or unwanted synonyms, and 20 % of the newly created synsets are connected to an incorrect hypernym. For this reason, the extended Czech WordNet will not be published before a thorough editing, but it is available for research.

2.4 Derivational relations in Czech WordNet

Another enrichment of the Czech WordNet is addition of derivational relations. See Figure 1 for the example of synset with a set of derivational relations (D-relations). As an example we show verbal synset *učit:1, vyučovat: probírat:1, brát:2* (the similar English one is *teach:1, instruct: 1*). It can be seen that there is a derivational subnet with five D-relations associated to *učit:1, ...* (in fact 14 but they are repeating with other literals in the synset as well). Each D-relation is labeled semantically so we have here the following D-relations: agentive, location, deverbative, gerund, passive – the last two may be characterized as more morphological (or surface) than the first three.

2.4.1 What is the nature of the D-relations?

The question may be asked what is the real nature of D-relations, whether it is semantic or rather morphological (formal). The D-relations exist between morphemes, typically between stems and corresponding suffixes (prefixes as well). This formal feature makes them different from the relations between sentence constituents, as e.g. between verbs and their arguments. However, the main criterion for us is whether the particular relation affects meaning irrespective of its formal

Synonyms: probírat:1, brát:2, učit:1, vyučovat:2

Definition:
--->> [deriv-ag] (učitel) učitel:1, instruktor:1
--->> [deriv-dvrb] (učení) učení:2
--->> [deriv-dvrb] (vyučování) vyučování:2
--->> [deriv-pas] (učený) učený:2
--->> [deriv-loc] učebna:1, třída:6
<<-- [deriv-pas] (učit) učen:1
<<-- [deriv-ger] (vyučovat) vyučující:1
<<-- [causes] dávat:1, naučit:1, poučit:1, vyučovat:1
<<-- [deriv-pas] (vyučovat) vyučovaný:1
<<-- [deriv-ger] (probírat) probírající:1
<<-- [deriv-pas] (probírat) probíraný:1
<<-- [deriv-dvrb] (probírat) probírání:1
<<-- [deriv-ger] (učit) učící:1
<<-- [deriv-pas] (vyučovat) vyučován:1
<<-- [hyponym] nastudovat:2
<<-- [hyponym] vstřebat:1, asimilovat:1
<<-- [deriv-pas] (probírat) probírán:1

Figure 1: Synset with derivational relations.

realization. If we apply this criterion to the D-relations discussed above, such as deriv-ag, deriv-loc, deriv-instr, deriv-g, deriv-dem, deriv-pos, deriv-pro, we definitely come to the conclusion that their nature is semantic.

Then there are relations like deriv-an, deriv-na, deriv-dvrb, deriv-ger, deriv-aad, deriv-pas that are sometimes characterized as morphological only and their semantics is left aside. The first two relations hold between nouns and adjectives and both denote properties (e.g. deriv-an: nový → novost (new → newness)), but we have to take into account that there is something that may be called semantics of the parts of speech, i.e. in one case property is expressed by the adjective and then by the noun which is derived from the adjective. Deriv-na denotes property as well but here the adjective is derived from noun as in boj → bojovný (fight → combative). The relation deriv-dvrb exists between a verb and noun, e.g. učit → učení (teach → teaching), and it denotes action which is first expressed by the verb and then by the deverbative noun. We can say that in these cases the only difference lies in the optics of the individual parts of speech but this difference should be understood as semantic as well. However, it should be remarked once more that quite often the differences in the semantics of the parts of speech are not treated as truly semantic.

If we have look what standard Czech grammars (Karlík, 1995, pp. 369–546.) say about the semantics of the parts of speech we find the formulations such as: nouns denote independent entities, i.e. persons, animals and things and also properties and actions. Verbs then denote states and their changes and processes (actions) and their mutations. These descriptions certainly refer to the semantics of the nouns and verbs. They are usually followed by the explanations about morphological processes, i. e. usually derivations by which some parts of speech are formed from the others, as we have described them above. What is relevant and what is missing in the standard grammars are more detailed and extensive semantic classifications of nouns, verbs, as well as adjectives and numerals. They are beginning to appear only recently and have the form of ontologies – the standard grammars do not use this term at all.

Until we have such semantic classifications describing semantic relations between the individual parts of speech we can hardly have a full picture that is necessary for automatic processing of the derivational relations. This issue certainly calls for a more detailed examination, which would be

Table 1: Literals count for each type of derivational relation.

deriv-na	641	(property, noun → adj)
deriv-ger	1951	(property, verb → adj)
deriv-dvrb	5041	(action, verb → noun)
deriv-pos	4073	(possessive, noun → adj)
deriv-pas	9801	(passive, verb → adj)
deriv-aad	1416	(property, adj → adverb)
deriv-an	1930	(property, adj → noun)
deriv-g	2695	(gender, noun → noun)
deriv-ag	186	(agentive, verb → noun)
deriv-dem	3695	(diminutive, noun → noun)
Total	31429 literals	

a topic for another paper.

2.4.2 The implementation of D-relations in Czech WordNet

Most wordnet editing tools standardly work with semantic relations between synsets and they treat them as atomic units. In fact, the synsets are not atomic as such and they consist of the smaller units called literals, i.e. for instance the synset *teach:1, instruct:1* contains two literals.

If we want to deal with the D-relations automatically we immediately face a problem: because of their nature they typically hold not between synsets but between literals that as a rule belong to the different synsets, e.g. *teach:1* and *teacher:1*. Therefore we need a tool that is able to define and create derivational links between the literals. The DEBVisdic editor supports this type of relation linking. We have used it for the implementation of the D-relations in Czech WordNet (see Table 1). The DEBVisdic tool is now used for representing and storing all the semantic relations including the D-relations. In our view, the way in which the D-relations (and other relations as well) are represented relevantly depends on the software tools used. This can be demonstrated if we compare the representation of the Czech D-relations in DEBVisdic with the one in PWN 3.0, which appears to be less explicit and rather verbose. This also means that the representation used in PWN 3.0 will be probably less suitable for possible applications.

2.4.3 The results

After processing all D-relations by the derivational morphological analyser Ajka we have added the derived literals (lemmas) to the Czech WordNet. The final result, the number of the literals generated from the individual D-relations is shown in Table 1 together with their semantic labels.

These numbers also tell us how productive the particular relations are. Note that the most frequent is passive relation which is followed by the deverbative (action) relation. The third most frequent relation is a possessive one. It would be interesting to examine what these facts can tell us about semantic structure of texts.

Though the presented analysis is far from complete at the moment the number of the generated items has led us to the decision to include them in Czech WordNet and enrich it considerably with the derivational nests (subnets). In our view, this kind of enrichment makes Czech WordNet more suitable for some applications, namely for searching.

The second and even more important reason for doing all this is a belief that the derivational relations and derivational subnets created by them reflect basic cognitive structures existing in

natural language. More effort is needed for exploring them from the point of view of now so popular ontologies – they certainly offer a formal ground (they are expressed by the individual morphemes) for natural language based ontologies.

2.5   Connection to VerbaLex

VerbaLex (Hlaváčková, Horák, & Kadlec, 2006) is a large lexical database of Czech verb valency frames which has been under development at NLP Centre since 2005. The organization of lexical data in VerbaLex is derived from the WordNet structure and entries follow the form of synsets. The current version of VerbaLex contains 6,360 synsets, 21,193 verb senses, 10,482 verb lemmata and 19,556 valency frames. When possible, a synset from VerbaLex is linked to its equivalent in Princeton WordNet. Out of the total number, 3,725 synsets have English equivalent, remaining 2,635 are verbs specific for the Czech language and not having lexicalized counterparts in English (i.e. in ILI).

2.6   Added definitions

Because many synsets in the Czech WordNet miss a definition, students of a linguistics course at the Faculty of Arts were asked to update the missing parts. The Czech definitions were written for 5,676 synsets from the Base Concepts set, consulting both Princeton WordNet definitions and Czech explanatory dictionaries. These revisions are currently only saved in text files and were not inserted into the Czech WordNet.

3   DEBVisDic integration with Open Multilingual WordNet

Since the BalkaNet project, NLP Centre is developing a browser and editor for wordnet-like lexical databases — VisDic (Horák & Smrž, 2003), later reimplemented as DEBVisDic (Horák, Pala, Rambousek, & Povolný, 2006; Rambousek & Horák, 2016). The editor stores the wordnet data in the XML format, thus making the wordnet databases standardized and exchangeable. The current DEBVisDic version is based on the DEB platform, a general lexicographic platform based on client-server architecture and adaptable to wide range of dictionary projects.

DEBVisDic is available as a web application and offers various features for wordnet browsing and editing. Users may work with several wordnets at once, utilizing linking and referencing between dictionaries. The application allows any user to create a new wordnet, without any complicated set-up, and start editing in a few minutes (Rambousek & Horák, 2016). To promote wordnet sharing, DEBVisDic supports export to the WordNet-LMF (Soria, Monachini, & Vossen, 2009) format.

As the part of preparation of new version of Czech WordNet, DEBVisDic editor will be updated to offer better integration with the Open Multilingual WordNet (OMW; Bond & Foster, 2013) repository. Users will be able to easily connect synsets to the Collaborative Interlingual Index (Bond, Vossen, McCrae, & Fellbaum, 2016) and upload data to OMW repository directly from the DEBVisDic.

4   Open Czech WordNet

The main impulse to promote the creation of a new version of the Czech WordNet was the proposal of integrating all available wordnets in the Global WordNet Association repository with Collaborative Interlingual Index. However, current Czech WordNet is not published under an open licence. Another important motivation is the need to fix various linguistic issues that may pose problems in using the Czech WordNet data in NLP applications.

Figure 2: User feedback form to provide synset data suggestions.

We have decided to evaluate and combine all the available updates and extensions to the Czech WordNet. The NLP Centre team has compiled the following roadmap that will lead to the publication of Open Czech WordNet:

- Start with 2009 Edited version and combine it with definitions created for Base Concepts.

- Check synonyms present in synsets, remove unnecessary synonyms and add missing words.

- Revise or create definitions where missing. Join or split synsets to follow word senses used in Czech language, where necessary.

- Verify all types of relations between synsets semi-automatically and fix broken relations.

- Link Czech synsets to their equivalents in Princeton WordNet 3.1 and to Collaborative Interlingual Index.

We plan to include the extensions from the semi-automatically translated Czech WordNet, but the data have to be evaluated by lexicographers first. Evaluation is planned during 2018.

It was not yet decided, in which way to include the VerbaLex data. However, the best option for the wordnet composition is to create new synsets based on the VerbaLex entries, including only the synonyms and definition to the wordnet data and linking to the VerbaLex for full verb valency information. VerbaLex does not contain relations between synsets, thus hyperonymy and troponymy relations have to be set in the wordnet.

Figure 3: Administrator view of suggested changes.

5   New crowdsourcing tool for update suggestions

During the integration we will not have enough resources and lexicographers to check all synsets and relations in the Czech WordNet. We have thus developed a new software tool that allows any authorized wordnet user to report issues and suggest specific corrections they spot in the data. Although we are testing the tool on the Czech WordNet it is language-independent and available for all wordnets developed using the DEBVisDic editor.

The tool is not directly integrated into DEBVisDic editor, but rather uses the DEBVisDic server API to access the wordnet data. On the other hand, all available synset representations (editor or simplified visualization) will enable users to easily move to the error reporting application. Users are presented with the data from the synset they were browsing and may propose an update of any data value — change existing value, add a new one if some part of synset is missing, or remove unwanted value. See Figure 2 for an example of the user feedback form. Updates are stored in a separate database as suggestions. Each value (e.g. gloss or relation) is stored as a single suggestion.

Any member of the editing team with an authorization to the given wordnet may browse all user suggestions (or filter them by reporting user, information type, or review status). The editors may approve or reject any single suggestion, or approve/reject all suggestions for any synset at once. Of course, it is also possible to approve/reject all suggestions based on the selected filter. Before deciding, the editor may compare user feedback with previously approved or rejected updates for the selected synset. See Figure 3 for a preview of the editor's interface.

All approved suggestions are immediately transferred to the development version of the wordnet database and presented to users. When a user's feedback is rejected by the editor, the information is kept in the database and future users trying to suggest the same update are notified about the previous refusal.

The application is automatically available for all wordnets stored at the selected DEBVisDic server. Its back-end is implemented in the Python programming language, stores all the suggestions, together with their history, and communicates with the DEBVisDiC server using the API. The user interface is a web application usable in any internet browser, implemented in HTML and JavaScript. The source codes for both server and client parts are published under an open license at `https://nlp.fi.muni.cz/projects/wordnet-report/`.

6 Conclusions and future work

The wordnet editing work is already under way with the plan to release the version of Czech WordNet linked to Collaborative Interlingual Index in 2018 under an open license and then continue with the evaluation of the previously prepared translated data. Depending on the funding and resources available, we plan to expand Czech WordNet and reach the coverage of Princeton WordNet. We would like to point out that the resulting data will be made accessible in LINDAT/Clarin Repository at UFAL MFF UK in Prague.

The DEBVisDic-Report application was designed for the purpose of long-term development and updates of the wordnet database by freely expandable editing teams. The core teams are still responsible for the final authorization, but the actual editing is kept as quick as possible. After successful evaluation with the Czech WordNet, the reporting tool will be available to any other wordnet databases.

References

Blahuš, M., & Pala, K. (2012). Extending Czech WordNet using a bilingual dictionary. In C. Fellbaum & P. T. J. M. Vossen (Eds.), *Proceedings of the 6th Global WordNet Conference, Matsue, Japan* (pp. 50–55). Brno: Tribun EU.

Bond, F., & Foster, R. (2013). Linking and extending an Open Multilingual WordNet. In *51st Annual Meeting of the Association for Computational Linguistics: Proceedings of the conference* (Vol. 1, pp. 1352–1362). Stroudsburg, PA: Association for Computer Linguistics.

Bond, F., Vossen, P., McCrae, J. P., & Fellbaum, C. (2016). CILI: The Collaborative Interlingual Index. In V. Barbu Mititelu, C. Forascu, C. Fellbaum, & P. Vossen (Eds.), *Proceedings of the Eighth Global WordNet Conference* (pp. 50–57). Bucharest: Global WordNet Association. Retrieved from `http://jiangbian.me/papers/2016/gwc2016.pdf`

Christodoulakis, D. (2004). *Balkanet Final Report.* University of Patras, DBLAB, No. IST-200029388.

Fellbaum, C. (Ed.). (1998). *WordNet: An electronic lexical database.* Cambridge, MA: MIT Press.

Filipec, J., et al. (1995). *Slovník spisovné češtiny* [SSČ]. Praha: Academia. (Electronic version LEDA, Praha).

Hlaváčková, D., Horák, A., & Kadlec, V. (2006). Exploitation of the Verbalex Verb Valency Lexicon in the syntactic analysis of Czech. In P. Sojka, I. Kopeček, & K. Pala (Eds), *Lecture Notes in Computer Science: Vol. 4188. Text, Speech and Dialogue: 9th International Conference, TSD 2006, Brno, Czech Republic, September 11–15, 2006: Proceedings* (pp. 79–85). Berlin: Springer–Verlag. `https://doi.org/10.1007/11846406_10`

Horák, A., & Smrž, P. (2003). VisDic — wordnet browsing and editing tool. In P. Sojka, K. Pala, P. Smrž, C. Fellbaum, & P. Vossen (Eds.), *Proceedings of the Second International WordNet Conference — GWC 2004: Brno, Czech Republic, January 20–23, 2004* (pp. 136–141). Brno: Masaryk University.

Horák, A., & Smrž, P. (2004). New features of wordnet editor VisDic. *Romanian Journal of Information Science and Technology*, 7, 1–13.

Horák, A., Pala, K., Rambousek, A., & Povolný, M. (2006). DEBVisDic — First Version of New Client-Server Wordnet Browsing and Editing Tool. In P. Sojka, K.-S. Choi, C. Fellbaum, & P. Vossen (Eds.), GWC 2006: *Third International WordNet Conference, GWC 2006 Jeju Island, Korea, January 22–26, 2006: Proceedings* (pp. 325–328). Retrieved from `http://semanticweb.kaist.ac.kr/conference/gwc/pdf2006/gwc06.pdf`

Horák, A., & Rambousek, A. (2018). Wordnet consistency checking via crowdsourcing. In J. Čibej, V. Gorjanc, I. Kosem, & S. Krek, *Proceedings of the XVIII EURALEX International Congress: Lexicography in Global Contexts* (pp. 1023–1029). Ljubljana: Ljubljana University Press, Faculty of Arts.

Jakubíček, M., Kovář, V., & Šmerk, P. (2011). Czech morphological tagset revisited. In R. Horák (Ed.), *Proceedings of recent Advances in Slavonic Natural Language Processing* (pp. 29–42). Brno: Tribun EU.

Karlík, P. (1995). *Příruční mluvnice češtiny*. Praha: Nakladatelství Lidové noviny.

Pala, K., & Smrž, P. (2004). Building Czech Wordnet. *Romanian Journal of Information Science and Technology, 7*(1–2), 79–88.

Rambousek, A., & Horák, A. (2016). DEBVisDic: Instant Wordnet Building. In V. Barbu Mititelu, C. Forascu, C. Fellbaum, & P. Vossen (Eds.), *Proceedings of the Eighth Global WordNet Conference* (pp. 317–321). Bucharest: Romanian Academy. Retrieved March 1, 2018, from `http://gwc2016.racai.ro/proceedings.html`

Rambousek, A., Horák, A., Klement, D., & Kletečka, J. (2017). New features in DEBVisDic for WordNet visualization and user feedback. In A. Horák, P. Rychlý, & A. Rambousek, *Proceedings of the 11th Workshop on Recent Advances in Slavonic Natural Languages Processing, RASLAN 2017* (pp. 3–10). Karlova Studánka: Tribun EU.

Rambousek, A., Pala, K., & Tukačová, S. (2017). Overview and future of Czech Wordnet. In J. P. McCrae, F. Bond, P. Buitelaar, P. Cimiano, T. Declerck, J. Gracia, I. Kernerman, E. Montiel Ponsoda, N. Ordan, & M. Piasecki, *LDK Workshops: OntoLex, TIAD and Challenges for Wordnets* (pp. 146–151). Galway: CEUR-WS.org.

Soria, C., Monachini, M., & Vossen, P. (2009). Wordnet-LMF: Fleshing out a standardized format for WordNet interoperability. In *Proceedings of IWIC2009*. New York, NY: ACM Press. `https://doi.org/10.1145/1499224.1499246`

Vossen, P. (Ed.). (1998). *EuroWordNet: A multilingual database with lexical semantic networks for European languages*. Dordrecht: Kluwer. `https://doi.org/10.1007/978-94-017-1491-4`

The authors declare that they have no competing interests.

The authors' contribution was as follows: concept of the study and Czech WordNet overview: Karel Pala; development of WordNet-related tools: Aleš Horák, Adam Rambousek; writing: Aleš Horák, Karel Pala, Adam Rambousek.

Publisher: Institute of Slavic Studies, Polish Academy of Sciences, University of Silesia & The Slavic Foundation