

Enriching plWordNet with morphology

Agnieszka Dziob and Wiktor Walentynowicz

G4.19 Research Group, Department of Computational Intelligence

Wrocław University of Technology, Wrocław, Poland

{agnieszka.dziob, wiktor.walentynowicz}@pwr.edu.pl

Abstract

In the paper, we present the process of adding morphological information to the Polish WordNet (plWordNet). We describe the reasons for this connection and the intuitions behind it. We also draw attention to the specificity of the Polish morphology. We show in which tasks the morphological information is important and how the methods can be developed by extending them to include combined morphological information based on WordNet.

1 Introduction

plWordNet is a very large wordnet of Polish. The 4.1 version presented in Dziob et al. (2019) contains 192,495 lemmas, 290,366 lexical units (henceforth, LUs), and 224,179 synsets belonging to four parts of speech: verbs, adjectives, adverbs and nouns. Since 2012, there have been carried out ongoing works on its connection to Princeton WordNet (Rudnicka et al., 2012).

For the description of synsets and LUs, lexical-semantic relations are used, the concept of which was taken from Princeton WordNet (Fellbaum, 1998) and EuroWordNet (Vossen, 2002). In the course of those works, the need emerged to create new relations specific to the Polish language (Piasecki et al., 2012). It is related to the necessity of describing derivational dependencies for a language with a rich derivational morphology. Inflectional morphology was not dealt with in plWordNet. Following Miller (1995), we assumed that describing inflectional morphology is the task of a separate resource that is morphological dictionaries for the Polish language, which support the use of plWordNet.

In plWordNet meaning is defined according to the assumptions of relational semantics (Lyons, 1977) that is as a bundle of lexico-semantic rela-

tions it enters. Thus, LUs having different relations in the language system cannot be treated as synonyms and belong to the same synset (Derwotedowa et al., 2008). Maziarz et al. (2013) formulated the concept of constitutive relations and constitutive features, i.e. those which differentiate the meaning. They include all synset relations, except fuzzynymy, and LU features, such as stylistic register, aspect for verbs, and semantic classes for adjectives and verbs (Maziarz et al., 2016). Therefore, there are no theoretical and methodological assumptions, which would allow to define inflectional features as distinguishing meanings of an LU in plWordNet. At the same time, it was also not explicitly stated that morphological features cannot influence meaning.

Lexicographic works are ongoing, and currently focus on completing the structure of plWordNet with new LUs, increasing the density of lexico-semantic relations, and correcting errors resulting from manual work. The most recent work has consisted of connecting morphological descriptions from the Grammatical Dictionary of Polish (pol. *Słownik Gramatyczny Języka Polskiego*, henceforth SGJP) (Saloni et al., 2007). This process has four practical objectives: 1) investigating the influence of morphological characteristics of LUs on their semantic description; 2) verifying the membership to parts of speech of morphologically ambiguous LUs and lemmas; 3) completing the semantic description of the existing lemmas with new senses, based on their morphological description in the SGJP; 4) building a practical resource, combining semantic and morphological description, for tasks related to the processing of the Polish language. The result of this work is plWordNet combined with the morphological description from SGJP, created automatically with a manual disambiguation of morphologically ambiguous lemmas, i.e. those which have more than one pattern of inflection.

The purpose of this paper is to present the results of combining resources and to indicate the applications of them.

2 The Problem of Morphology Description

2.1 The Morphological Description in Wordnets

The assumption of Princeton WordNet was a semantic description, i.e. including derivational, not inflectional morphology (Miller et al., 1990). Still, morphological data resources are developed as independent linguistic databases. One of them is CELEX – a lexical database for Dutch and English (Van der Wouden, 1990), extended in 2.0 version with German (Baayen et al., 1995). It contains orthographic, phonological, morphological (also inflectional), syntactic, and statistical information (frequency in text corpora), but it does not contain semantics.

The morphological description of the derivation and inflection in CELEX offers great opportunities to enrich it with semantic information. Hathout (2002) describes combining the morphological resources of CELEX for the English language with Princeton WordNet to create a language-independent mechanism for obtaining semantic relational data (synonyms and derivatives). Similar studies with the use of CELEX and semantic-relational thesauri were conducted for Dutch (Kraaij and Pohlmann, 1996). This research is particularly applicable to the extraction of information and knowledge from text.

The inflectional morphology is less of a problem for languages with residual inflection, such as English, but a serious challenge for highly inflected languages. In the paper Henrich et al. (2012), semantic data from Princeton WordNet and GermaNet (Hamp and Feldweg, 1997) and morphological data for these languages from Wiktionary were used to create a method for sense-annotation and automatically annotated text corpora.

Slavic languages have an even more complicated inflection than Germanic ones. The paper of Pala and Hlaváčková (2007) presents the results of the work consisting of adapting a mechanism of the Czech morphological analyzer Ajka (Sedláček and Smrž, 2001) to extend the Czech WordNet with derivational relations. For Slavic languages, the research on derivational morphol-

ogy has also been carried out on a larger scale for Bulgarian (Koeva, 2008).

The works that are the closest to those presented in this paper were described in Obradović and Stanković (2007). The authors have developed a tool for the creation of complex lexicographic data obtained from wordnets, morphological dictionaries, and text corpora.

It is possible to highlight several important aspects of morphological description in wordnets: 1) on a wider scale it describes derivational, not inflectional morphology; 2) there is a regular relationship between inflection and derivation, but these two levels of description are not treated as equivalent; 3) the combination of semantic and morphological description (both, at the inflectional and derivational level) is useful, e.g. for the tasks related to Word Sense Disambiguation and, in connection with it, the extraction of information from texts and building knowledge bases.

2.2 The Specificity of the Polish Morphology

As already mentioned, plWordNet describes four parts of speech: verbs, adjectives, adverbs, and nouns. In Polish, nouns and adjectives are inflected by seven cases in two numbers: singular and plural. Furthermore, adjectives are inflected for gender, while nouns are always lexically specified for grammatical gender.

There are five genders: three masculine (personal, animate, and inanimate), feminine and neuter (Lewandowska-Tomaszczyk et al., 2012). For example, *pies* (zw) ‘dog’ and *pies* (os) ‘cop’ (depreciative and colloquial) differ in their genders and, consequently, in the pattern of the inflection. The gender of adjectives depends on the gender of nouns they have syntactic relations in the text, e.g. *szafa* ‘wardrobe’, feminine, *czerwon-a* ‘red’ an adjective, feminine. Moreover, adjectives and adverbs have three degrees: positive, comparative, and superlative.

Verbs in Polish are inflected for two numbers and three persons for each of them, four tenses (including two future ones that differ with each other only grammatically), and three modes of expressing modality (indicative, conditional, imperative). They have an assigned aspect having grammatical and semantic functions (Dziob and Piasecki, 2018). They form gerunds and four types of participles, which are treated as forms of the verb. A semantic-syntactic feature of the verb (to a lesser

extent also of other parts of speech) is valence, understood as the ability of predicates to attach arguments in specific forms and syntactic positions (Przepiórkowski et al., 2014). For example, the verb *jeść* ‘to eat’ opens three syntactic positions, for a subject, an object, and a circumstance. In this case, the object is expressed as a noun in the accusative case. Walenty is connected at the semantic layer to the plWordNet by using synsets to determine a semantic preference, e.g. ‘*jeść*’ + *jedzenie 2* ‘food’.

Traditional Polish grammars cf. (Grzegorzyczyk et al., 1998) distinguish a quite large group of numbers and pronouns. In syntactically oriented grammars, e.g. (Saloni, 2012), they are combined with other parts of speech (nouns, adjectives, adverbs, and verbs) depending on the pattern of the inflection and morphosyntactic functions. This solution was also applied in plWordNet.

2.3 The Morphological Description in the NLP Practice

Morphological analysis is widely used in the syntactic processing of the Polish language. The task of morphological tagging is to process a sequence of tokens – sentence – and assign each of them a unified morphological interpretation. In the task of morphosyntactic tagging the morphological analysis is most often used in two ways: 1) as a set of input features to the tagger (Waszczuk, 2012, Wróbel, 2017), 2) as a tagger output filter (Georgiev et al., 2019, Walentynowicz et al., 2019).

Indirectly, morphology can be used as a method of regularization tagger learning process (Straka et al., 2019). Also in the task of inflection language chunking, full morphological information is used (Goldberg et al., 2006).

The task which depends on the morphological information and, at the same time, has a semantic aspect, is lemmatization. Words have different patterns of inflection by case, depending on their meaning. Choosing the right base form affects the result of tasks occurring after the lemmatization process such as Word Sense Disambiguation or Named Entity Recognition.

The combination of morphological information with semantic information will allow for better results in syntactic tasks due to better differentiation of contexts of the occurrence of given expression forms. The semantic dimension, which Word-

Net contains, allows searching for new patterns in data.

3 Resources

3.1 Morfeusz and SGJP

SGJP (Saloni, 2012) aims to give grammatical characteristics of Polish words. The main element of this characteristic is an open description of the inflection of units taken into account by giving all their forms of inflection and determining their grammatical functions. The dictionary does not contain lexeme sense. Morfeusz2 (Woliński, 2014) is a morphological analyzer that can use data from SGJP as a basis for a dictionary. It has the ability to analyze word forms, without recognizing out-of-vocabulary words. It is also able to perform morphological synthesis – the creation of a modified word form by indicating the lemma and the desired inflection characteristics.

The above-mentioned combination of Morfeusz2 with the SGJP dictionary was used to prepare the projection of forms found in plWordNet on the morphology available in SGJP. The list of lemmas available in plWordNet was processed by Morfeusz2, and then all word form varieties were prepared with the help of a morphological synthesis. This process had to be supervised by a linguist, because not all lemmas change in the same way – depending on the sense of a word, differences in the inflection can occur. This is when a linguist decided which inflection scheme to use.

3.2 Combining of Resources

As already mentioned, plWordNet describes four parts of speech. For all of them, the morphological information has been drawn from the SGJP. However, as a result, not all units have been given a pattern of inflection, see 4.1. In the task described in this article, new patterns of inflection were not added. Morphological disambiguation consisted of manual removal of the excess patterns for ambiguous lemmas. The task of linguists was to leave a set of forms confirmed for a given meaning in text corpora, even when they were rare or used in a specific context. In the cases where the morphological description did not match any of the available meanings, it was removed, and a new LU has been added to plWordNet without any inflection. Five linguists and a coordinator, who controlled the quality of works, worked on this task. Each of the linguists worked on one set of morpholog-

ical features. The inter-annotator agreement was not measured. The morphological information for these units shall also be added in the further iteration. The lexicographic works were realized using the WordNet Loom editing system (Naskret et al., 2018), using a specially constructed field to edit morphological data.

4 The Alliance of Morphology and Semantics

4.1 Morphological Disambiguation

Two lists of lemmas were the result of a comparison of plWordNet and SGJP. The first one indicated those lemmas, which are described in plWordNet, but not in the SGJP. The second list included lemmas that are morphologically ambiguous in plWordNet. Those lemmas were manually disambiguated at the level of LUs. In total, 3,733 lemmas were ambiguous, including 772 adjectives, 200 adverbs, 2,309 nouns, and 452 verbs.

Among the ambiguous lemmas were those that belong to the following groups: 1) lemmas that have an adjective and a noun pattern of inflection, e.g. *white 1* (color, adjective) and *white 1* ‘White’ (person, noun); 2) nouns which may have two grammatical genders depending on their meaning, e.g. *pies 1* ‘dog’ and *dog 4* ‘cop’; 3) proper names (surnames and names of places), which in SGJP have a given masculine gender, in plWordNet are not described at all (the work of linguists consisted in removing excess patterns); 4) elements of multi-word LUs that are not one-words lemmas in plWordNet¹; 5) lemmas, which in one meaning belong to parts of speech described in plWordNet, but not in another, e.g. noun *jeden 1* ‘short, nip’ and the numeral *jeden* ‘one’; 6) lemmas which, depending on their meaning, may belong to: a) nouns or verbs (gerunds), e.g. *uczulenie* as a noun (‘allergy’) and gerund from the verb *uczulić* ‘to sensitize’; b) adjectives or participles, e.g. adjective *zabłąkany* ‘as an adjective (‘confused’) and participles from the verb *zabłąkać się* ‘get lost’; 7) two-aspectual verbs, i.e. those which have the same form in the perfective and imperfective, cf. (Grzegorzczkowska et al., 1998), e.g. *izolować 1* ‘to isolate (perflimperf)’; 8) lemmas marked by pragmatics (e.g. high, formal, book) and unmarked meaning, e.g. *miły 1* ‘dear’ has the general form

¹The description of multi-words is planned for the further work by linking it to the SEJF Dictionary (Czerepowicka and Savary, 2015).

of a comparative *milszy* and superlative *najmilszy* and a marked (old, book) comp. *milejszy* and sup. *najmilejszy*; 9) lemmas non-inflectable according to any Polish pattern of inflection, belonging to a part of speech disambiguated contextually, e.g. *extra* ‘additionally’ (adverb) and *extra* ‘additional’ (adjective).

These are the most common problems defined in the course of manual work. They result from the ambiguity of the Polish language and its rich grammar, on the one hand, and, on the other hand, from the way of defining the LU in plWordNet. And they also have their consequences for this definition.

5 Towards the Definition of Meaning

The third list, resulting from the manual connection of the resources, contains senses missing in plWordNet. It includes about 800 lemmas which appear in plWordNet in a different sense and whose missing sense was made possible to be completed by a morphological disambiguation.

Among them there are the following groups of senses which have been qualified as plWordNet LUs: 1) lemmas, which plWordNet contains only in the adjective meaning, but not in their noun meaning, e.g. *hotelowy* ‘hotel’ (adjective) and ‘the person who serves guests in the hotel, boy’ (noun)²; 2) the sub-group, which contains missing adjective senses, but not distinct also in traditional dictionaries; those lemmas have in texts the function of a subject or an object (like a noun), not an attributive (like an adjective), e.g. *otyły* ‘fat’ (adjective, described in plWordNet) and *otyły* ‘a fat person’ (noun, none); 3) missing sense, which are pragmatically marked; they are included in plWordNet only if their appearance can be confirmed in corpora (Maziarz et al., 2014), e.g. *napaść* ‘to fatten’, which has a different inflection than *napaść 1* ‘to attack’; 4) uninflected lemmas belonging to a part of speech disambiguated by the context, e.g. *bordo* ‘maroon’ as an adjective (described) or adverb (none); 5) ambiguous nouns whose grammatical gender is contextually disambiguated, e.g. *kapo*-female ‘female guard in a concentration camp’ (in plWordNet is only *kapo*-male); 6) inflected nouns, which can have different grammatical genders, depending on their meaning, e.g. *przewodnik* ‘guide’ (person,

²This group includes many representatives of less popular or former occupations and positions (functions).

personal gender), *przewodnik* ‘guidebook’ (thing, inanimate gender), *przewodnik* ‘guide dog’ (none; animate); 7) bearers of features, in the case of which meanings can be distinguished by particular lexico-semantic relations, e.g. *garbusek* ‘little hunchback’ (person who has a hump) and *garbusek* ‘little beetle’ (car).

In addition to the above, the method of manually enriching plWordNet allows to reveal other LUs, the meaning of which is connected with less regular morphological processes. These are such LUs as e.g. *podskarbiostwo* ‘the married couple, a former Polish court’s clerk and his wife’ *flop* ‘the computer power unit’, *flop* ‘men’s hairstyle’ etc.

It is interesting that for many nouns with lemmas identical to adjectives, whose morphological disambiguation is contextual, plWordNet describes only masculine nouns, e.g. there is *gruby* as a ‘fat person’ but not *gruba* as a ‘fat women’. Let us recall that nouns do not inflect by gender, but have their gender assigned. On this basis, it can be concluded that the existence of these meanings as separate is non-intuitive and not obvious in Polish.

The list of deficiencies also includes meanings that will not be included in plWordNet due to its theoretical-methodological limitations of defining meanings: 1) senses which are understood in the SGJP as adjectives, whereas in plWordNet interpreted as participles, e.g. *kupujący* ‘someone who buys’ (derived from the verb *kupować* ‘to buy’); 2) occasionalisms, e.g. *bufetowa* ‘the person managing the canteen’ (the journalists called in this way the former Mayor of Warsaw, Hanna Gronkiewicz-Walcz); 3) nouns that may be personal or non-personal, e.g. *pięciolatek* ‘a person who is a five years old’ or ‘animal who is a five years old’; in plWordNet this is the same sense; 4) archaisms whose occurrence is not confirmed by corpora, e.g. *majorat* as a person; 5) lemmas which occur only in multi-word units cf. (Maziarz et al., 2015), e.g. *warcabnik* as ‘butterfly’ occurs only in species names *warcabnik ślazowiec* ‘*Carcharodus alceae*’ and *warcabnik szantowiec* ‘*Carcharodus floccifera*’; 6) proper names not constituting the basis for derivation of relational adjectives used in general language, cf. (Maziarz et al., 2012), e.g. *Koto* ‘a part of Warsaw’s Wola district’; 7) acronyms from proper names which are not described in plWordNet, e.g. *LP* - Legiony Polskie ‘Polish Legions’.

The most important conclusion for the methodology and procedure of distinguishing senses is that the morphological pattern cannot be treated as a distinguishing feature. Instead, it can be a strong argument for manual work, which consists of verifying in corpora previously not described LUs. This is especially true in the case of regularities connected with distinguishing LUs belonging to different parts of speech, as well as grammatical genders (masculine and feminine). The existence of two masculine patterns of inflection next to each other needs to be verified every time, because plWordNet often treats meanings more generally than it is established in the Polish grammatically oriented linguistics.

6 New Possibilities

A wordnet combined with morphological information can be used by NLP tools such as taggers and shallow parsers. The use of wordnet-based context vectors (Patwardhan and Pedersen, 2006) or the combination of word embeddings with wordnet information (Mao et al., 2018) have already been applied in NLP tasks. Current taggers most often rely on a vector-based word representation, so a wordnet-based context vector could be attached to the input representation to better represent the dependencies between tokens in a sentence. More semantic information should improve the lemmatization process, which cannot be based solely on morphological information, since there are cases where a pair of word forms and tags has several possible lemmas. The task of extracting key phrases requires morphological information to obtain good results. The word relations that are in wordnet are an additional element that should improve the results (Kardan et al., 2013). So far, this task has not used the combined morphology information with the data from wordnet.

The combination of information contained in wordnet and morphological dictionaries opens up new paths for the development of a method in various NLP tasks, like the ones mentioned above. Moreover, from the implementation side, such integration will allow reducing the number of dependencies in the already functioning methods, which use both the morphology and wordnet. Examples of such systems can be a system of clustering terms from the field of economics, which uses morphological information and relationships from wordnet (Mykowiecka and Marciniak, 2012), or

wordnet-based morphological analysis (Geum and Park, 2016).

Acknowledgments

The work co-financed as part of the investment in the CLARIN-PL research infrastructure funded by the Polish Ministry of Science and Higher Education.

References

- R Harald Baayen, Richard Piepenbrock, and Hedderick Van Rijn. The celex database. *Nijmegen: Center for Lexical Information, Max Planck Institute for Psycholinguistics, CD-ROM*, 1995.
- Monika Czerepowicka and Agata Savary. Sejf-a grammatical lexicon of polish multiword expressions. In *Language and Technology Conference*, pages 59–73. Springer, 2015.
- Magdalena Derwojedowa, Maciej Piasecki, Stanisław Szpakowicz, Magdalena Zawislawska, and Bartosz Broda. Words, Concepts and Relations in the Construction of Polish WordNet. In A. Tanács, D. Csendes, V. Vincze, Ch. Fellbaum, and P. Vossen, editors, *Proc. Fourth Global WordNet Conf.*, pages 162–177, 2008.
- Agnieszka Dziob and Maciej Piasecki. Implementation of the Verb Model in plWordNet 4.0. In *Proceedings of the 9th Global Wordnet Conference*, 2018.
- Agnieszka Dziob, Maciej Piasecki, and Ewa Rudnicka. plwordnet 4.1—a Linguistically Motivated, Corpus-based Bilingual Resource. In *Global Wordnet Conference 2019*, pages 353–362, Wrocław, 2019.
- Christiane Fellbaum, editor. *WordNet – An Electronic Lexical Database*. The MIT Press, 1998.
- Georgi Georgiev, Valentin Zhikov, Petya Osenova, Kiril Simov, and Preslav Nakov. Feature-rich part-of-speech tagging for morphologically complex languages: Application to bulgarian. *arXiv preprint arXiv:1911.11503*, 2019.
- Youngjung Geum and Yongtae Park. How to generate creative ideas for innovation: a hybrid approach of wordnet and morphological analysis. *Technological Forecasting and Social Change*, 111:176–187, 2016.
- Yoav Goldberg, Meni Adler, and Michael Elhadad. Noun phrase chunking in hebrew: Influence of lexical and morphological features. In *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics*, pages 689–696, 2006.
- Renata Grzegorzczkova, Roman Laskowski, and Henryk Wróbel. *Morfologia: gramatyka współczesnego języka polskiego*. Wydawn. naukowe PWN, 1998.
- Birgit Hamp and Helmut Feldweg. Germanet—a lexical-semantic net for german. In *Automatic information extraction and building of lexical semantic resources for NLP applications*, 1997.
- Nabil Hathout. From wordnet to celex: acquiring morphological links from dictionaries of synonyms. In *LREC*, 2002.
- Verena Henrich, Erhard Hinrichs, and Tatiana Vodolazova. An automatic method for creating a sense-annotated corpus harvested from the web. *International Journal of Computational Linguistics and Applications*, 3(2):47–62, 2012.
- Ahmad A Kardan, Farzad Farahmandnia, and Amin Omidvar. A novel approach for keyword extraction in learning objects using text mining and wordnet. *Global Journal of Information Technology*, 3(1), 2013.
- Svetla Koeva. Derivational and morphosemantic relations in bulgarian wordnet. *Intelligent Information Systems*, 16: 359–369, 2008.
- Wessel Kraaij and Renée Pohlmann. *Using Linguistic Knowledge in Information Retrieval Technical Report*. Citeseer, 1996.
- Barbara Lewandowska-Tomaszczyk, Mirosław Bańko, Rafał L Górski, Piotr Pęzik, and Adam Przepiórkowski. *Narodowy korpus języka polskiego*. Wydawnictwo Naukowe PWN, 2012.
- John Lyons. *Semantics*. 1977.
- Rui Mao, Chenghua Lin, and Frank Guerin. Word embedding and wordnet based metaphor identification and interpretation. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1222–1231, 2018.
- Marek Maziarz, Stanisław Szpakowicz, and Maciej Piasecki. Semantic relations among adjectives in polish wordnet 2.0: a new relation set, discussion and evaluation. *Cognitive Studies*, (12), 2012.
- Marek Maziarz, Maciej Piasecki, and Stanisław Szpakowicz. The chicken-and-egg problem in wordnet design: synonymy, synsets and constitutive relations. *Language Resources and Evaluation*, 47(3):769–796, 2013.
- Marek Maziarz, Maciej Piasecki, Ewa Rudnicka, and Stan Szpakowicz. Registers in the system of semantic relations in plwordnet. In *Proceedings of the Seventh Global Wordnet Conference*, pages 330–337, 2014.
- Marek Maziarz, Stan Szpakowicz, and Maciej Piasecki. A procedural definition of multi-word lexical units. In *Proceedings of the International Conference Recent Advances in Natural Language Processing*, pages 427–435, 2015.
- Marek Maziarz, Maciej Piasecki, Ewa Rudnicka, Stan Szpakowicz, and Paweł Kędzia. plwordnet 3.0 – a Comprehensive Lexical-Semantic Resource. In Nicoletta Calzolari, Yuji Matsumoto, and Rashmi Prasad, editors, *COLING 2016, 26th International Conference on Computational Linguistics, Proceedings of the Conference: Technical Papers, December 11-16, 2016, Osaka, Japan*, pages 2259–2268. ACL, 2016.
- George A Miller. Wordnet: a lexical database for english. *Communications of the ACM*, 38(11):39–41, 1995.
- George A Miller, Richard Beckwith, Christiane Fellbaum, Derek Gross, and Katherine J Miller. Introduction to wordnet: An on-line lexical database. *International journal of lexicography*, 3(4):235–244, 1990.
- Agnieszka Mykowiecka and Malgorzata Marciniak. Combining wordnet and morphosyntactic information in terminology clustering. In *Proceedings of COLING 2012*, pages 1951–1962, 2012.
- Tomasz Naskret, Agnieszka Dziob, Maciej Piasecki, Chakaveh Saedi, and António Branco. Wordnetloom—a multilingual wordnet editing system focused on graph-based presentation. In *Proceedings of the 9th Global WordNet Conference (GWC2018)*, 2018.
- Ivan Obradović and Ranka Stanković. Wordnet development using a multifunctional tool. In *Proceedings of the International Workshop Computer Aided Language Processing (CALP) 2007*, pages 25–32, 2007.

- Karel Pala and Dana Hlaváčková. Derivational relations in czech wordnet. In *Proceedings of the Workshop on Balto-Slavonic Natural Language Processing*, pages 75–81, 2007.
- Siddharth Patwardhan and Ted Pedersen. Using wordnet-based context vectors to estimate the semantic relatedness of concepts. In *Proceedings of the Workshop on Making Sense of Sense: Bringing Psycholinguistics and Computational Linguistics Together*, 2006.
- Maciej Piasecki, Radosław Ramocki, and Marek Maziarz. Recognition of Polish Derivational Relations Based on Supervised Learning Scheme. In *LREC*, pages 916–922, 2012.
- Adam Przepiórkowski, Elżbieta Hajnicz, Agnieszka Patejuk, Marcin Woliński, Filip Skwarski, and Marek Świdziński. Walenty: Towards a comprehensive valence dictionary of Polish. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Hrafn Loftsson, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Ninth International Conference on Language Resources and Evaluation, LREC 2014*, pages 2785–2792. ELRA, 2014.
- Ewa Rudnicka, Marek Maziarz, Maciej Piasecki, and Stan Szpakowicz. A Strategy of Mapping Polish WordNet onto Princeton WordNet. In *Proc. COLING 2012, posters*, pages 1039–1048, 2012.
- Zygmunt Saloni. Podstawy teoretyczne „słownika gramatycznego języka polskiego”. *Warszawa, online: <http://sgjp.pl/static/pdf/Wst%20C4%2099p%20do%20II%20wydania%20SGJP.pdf>*, 1458313456, 2012.
- Zygmunt Saloni, Włodzimierz Gruszczyński, Marcin Woliński, and Robert Wołosz. Grammatical Dictionary of Polish. *Studies in Polish Linguistics*, 4:5–25, 2007.
- Radek Sedláček and Pavel Smrž. A new czech morphological analyser ajka. In *International Conference on Text, Speech and Dialogue*, pages 100–107. Springer, 2001.
- Milan Straka, Jana Straková, and Jan Hajič. Udpipeline at sigmorphon 2019: Contextualized embeddings, regularization with morphological categories, corpora merging. *arXiv preprint arXiv:1908.06931*, 2019.
- Ton Van der Wouden. Celex: Building a multifunctional polytheoretical lexical data base. *Proceedings of BudaLex*, 88:363–373, 1990.
- Piek Vossen. EuroWordNet General Document Version 3. Technical report, Univ. of Amsterdam, 2002.
- Wiktor Walentynowicz, Maciej Piasecki, and Marcin Oleksy. Tagger for polish computer mediated communication texts. In *Proceedings of the International Conference on Recent Advances in Natural Language Processing (RANLP 2019)*, pages 1295–1303, 2019.
- Jakub Waszczuk. Harnessing the crf complexity with domain-specific constraints. the case of morphosyntactic tagging of a highly inflected language. In *Proceedings of COLING 2012*, pages 2789–2804, 2012.
- Marcin Woliński. Morfeusz reloaded. In Nicoletta Calzolari, Khalid Choukri, Thierry Declerck, Hrafn Loftsson, Bente Maegaard, Joseph Mariani, Asuncion Moreno, Jan Odijk, and Stelios Piperidis, editors, *Proceedings of the Ninth International Conference on Language Resources and Evaluation, LREC 2014*, pages 1106–1111, Reykjavík, Iceland, 2014. European Language Resources Association (ELRA). ISBN 978-2-9517408-8-4. URL <http://www.lrec-conf.org/proceedings/lrec2014/index.html>.
- Krzysztof Wróbel. Krnnt: Polish recurrent neural network tagger. In *Proceedings of the 8th Language & Technology Conference: Human Language Technologies as a Challenge for Computer Science and Linguistics*, pages 386–391. Fundacja Uniwersytetu im. Adama Mickiewicza w Poznaniu, 2017.