A New Version of the *Składnica* Treebank of Polish Harmonised with the *Walenty* Valency Dictionary

Marcin Woliński, Elżbieta Hajnicz, Tomasz Bartosiak
Institute of Computer Science, Polish Academy of Sciences
Jana Kazimierza 5, 01-248 Warsaw, Poland
{wolinski,hajnicz}@ipipan.waw.pl, tomasz.bartosiak@gmail.com

Abstract

This paper reports on developments in the Składnica treebank of Polish which were possible due to the switch to the Walenty valency dictionary. The change required several modifications in the Świgra parser, such as implementing unlike coordination, semantically motivated phrases, and non-standard case values. A procedure to upgrade manually disambiguated trees of Składnica was required as well. Modifications introduced in the treebank included systematic changes of notation and resolving ambiguity between semantically motivated phrases.

The procedure of confronting Składnica treebank with the trees generated with the new version of the Świgra parser using Walenty dictionary allowed us to check the consistency of all the resources. This resulted in several corrections introduced in both the treebank and the valence dictionary.

Keywords: treebank, constituency parsing, valency dictionary

1. Introduction

This article concerns adapting the Składnica treebank to the valence dictionary Walenty. Składnica predates Walenty, so initially the treebank was based on the valence dictionary of Świdziński (1994). Walenty surpasses this dictionary both in size and the number of linguistic phenomena being represented. Therefore, deploying Walenty was an obvious choice for the further development of Składnica.

The procedure to adapt Składnica to the new dictionary was to a large extent automatic. However, the differences between the resources made it necessary to manually correct some parse trees. In this article, we present the method of automatic mapping and the problems that needed manual intervention.

The article is organised as follows. First we present the resources – the treebank Składnica together with the parser Świgra used to generate the trees (Section 2) and the valence dictionary Walenty (Section 3). Next, we discuss adapting the parser to Walenty (Section 4) and then adapting the treebank (Section 5).

2. The treebank and the parser

Składnica is a treebank of Polish (Woliński et al., 2011) built on a sub-corpus sampled from the one million word manually annotated sub-corpus (NKJP1M) of the National Corpus of Polish – NKJP (Przepiórkowski et al., 2012). Corpus samples consist of a few sentences each, and they sum up to 20,000 sentences.

The primary form of the resource comprises constituency trees generated with the DCG (Pereira and Warren, 1980) parser Świgra (Woliński, 2004; Świdziński and Woliński, 2010) and then manually disambiguated and validated (Woliński, 2010). It was assumed as a construction rule for the treebank that all accepted trees have to be actually generated by the parser. The treebank annotators are not allowed to modify trees in any way nor to provide trees for sentences rejected by the parser. The representation of sentences without a proper parse tree has to be corrected by improving the parser's grammar. This leads to an iterative development of the grammar and the treebank. The grammar feeds the treebank and the treebank documents the coverage of the grammar.

The grammar used by Świgra stems from Świdziński's grammar (Świdziński, 1992), but it was deeply restructured. The trees generated by the present version are much simpler and more intuitive. The grammar was also extended in many ways, in particular to describe various forms of coordinated structures.

Figure 1 shows Składnica/Świgra annotation for the sentence:

(1) *Trzeba określić i zbadać rodzaj infekcji*
must determine and study kind infection
oraz co ją powoduje.
and what it causes

'It's necessary to determine and confirm the kind of infection and what causes it.'

Leaves of the tree correspond to terminals (represented with a form and a lemma in the picture). Internal nodes correspond to phrases. They are represented by the name of the non-terminal category in Fig. 1. The labels use abbreviations of Polish names, which are explained in Table 1. In fact, each node carries several attributes specifying its syntactic features (all such features of a fw node are shown in Fig. 2). Children of a given node are its constituents, as determined by some rule of the grammar being used. An important feature of Składnica trees is the fact that one of the constituents is marked as the syntactic head (marked with an edge with a thick grey background), which allows to convert constituency trees to dependency trees. Such conversion has in fact been performed resulting in a dependency version of Składnica (Wróblewska and Woliński, 2012), later on converted also to Universal Dependencies (Seddah et al., 2013). Non-terminals of the grammar fall into several types or layers in the tree (Świdziński and Woliński, 2010). From the bottom up, these are:

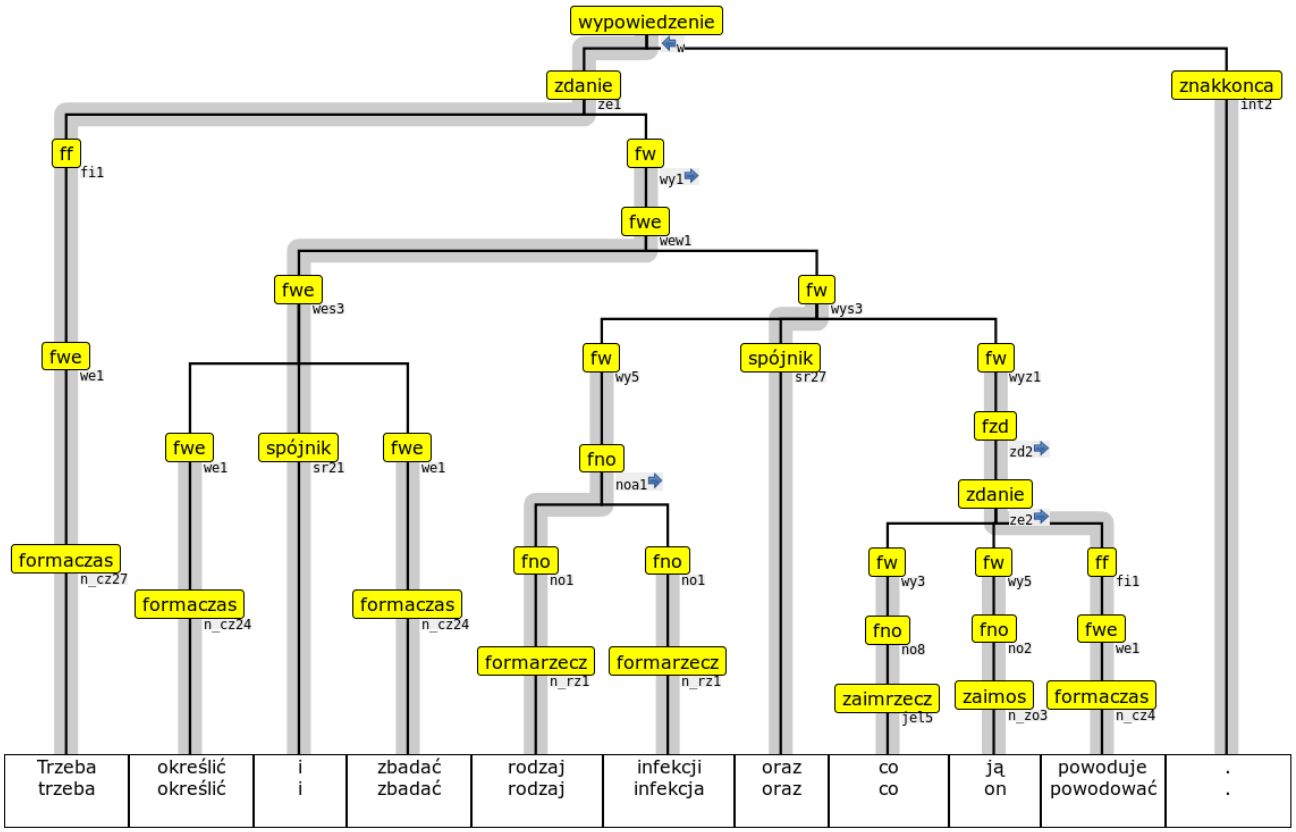


Figure 1: A Świgra parse tree for the example (1)

syntactic words	
formaczas	verbal form
formarzecz	nominal form
zaimrzecz	nominal pronoun
zaimos	personal pronoun
spójnik	conjunction
constituent phrases	
fno	nominal phrase
fwe	verbal phrase
ff	finite phrase (an fwe that can constitute a clause)
fzd	clausal phrase
valency phrases	
fw	required phrase (argument)
fl	free phrase (adjunct)
clauses	
zdanie	clause/sentence
wypowiedzenie	utterance

Table 1: Non-terminal categories of Fig. 1

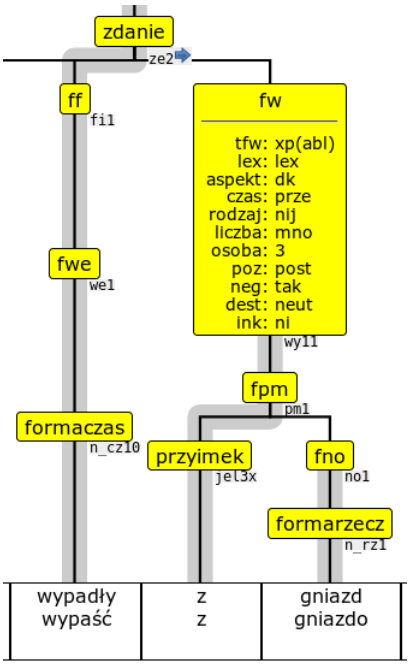
- *Syntactic words* form the syntactic counterpart of terminals. Typical examples are the units formaczas, formarzecz, zaimrzecz, and zaimos in Figure 1. However, units of this level can also represent multi-token verbal forms (e.g., analytical future forms of verbs *będzie czytać* 'will read') and other cases where one form, from the syntactic viewpoint, corresponds to several tokens in the NKJP tagset, e.g. two-word prepositions *wraz z* 'together with' and adverbs *po ciemku* 'in the dark'.

- *Constituent phrases* are used to describe the attachment of various modifiers to verbal, nominal, adjectival, and adverbial heads. Also at this level, prepositional-nominal phrases and subordinate clauses are formed. Constituent phrases can also be coordinated structures (with a conjunction as a head).

- *Valency phrases*, as proposed by Świdziński (1992), denote functions played by constituent phrases. These differentiate dependents into required phrases (arguments) fw and free phrases (adjuncts) fl. Thanks to their presence, the valency structure gets visible in the tree.

- The fourth layer comprises *clauses*. Simple clauses consist of a finite phrase and valency phrases. Coordinate clauses, based upon a conjunction as their head, have other clauses as their constituents.

As can be seen in Fig. 2, the first attribute of a required phrase fw is tfw – the 'type of required phrase'. This attribute shows the characteristic of the phrase assigned by the valence dictionary. Some of the other attributes of fw and fl are, somewhat counter-intuitively, shared with the head of the clause. This allows to confront some of the attributes of the given dependent with the attributes of the head, for example to require gender and number agreement of the verb and the subject.



wypadły *z* *gniazd*
fall from nests

'[they] have fallen out of nests'

Figure 2: Complete set of features for one required phrase fw of type xp(abl)

3. Valency dictionary Walenty

A valency dictionary specifies what types of arguments are possible for a given predicate. In Polish, the need of such information is most obvious for verbs, which differ widely in possible arguments, e.g., some verbs allow for a complement in the form of a verbal phrase in the infinitive and others don't.

Initially, Świgra used a valency dictionary based on (Świdziński, 1994). This dictionary was extended during the treebank development project. At the end of the project the dictionary consisted of 6400 schemata for 1450 Polish verbs (covering about 75% of verb occurrences in NKJP1M).

Later, this dictionary became a seed for a new one, which is currently being developed at ICS PAS. Walenty is a comprehensive valency dictionary of Polish based on corpus data (Hajnicz et al., 2016b; Przepiórkowski et al., 2014c; Przepiórkowski et al., 2014b; Przepiórkowski et al., 2014a; Hajnicz et al., 2016a). After a few years of development, Walenty contains 99,000 schemata for 18,100 predicates, which include 13,000 verbs, 4,000 nouns and 1100 adjectives and adverbs. Walenty covers 99,8% of occurrences of verbal forms in the 300 millions word balanced sub-corpus of NKJP. Moreover, Walenty is much more rich in information than the original Świgra dictionary.

The dictionary uses so-called structural case, i.e. case whose morphological realisation depends on the syntactic context. It is used in two contexts – as the case of nominal subjects and as the case of nominal phrases underlying the genitive of negation. For Świgra, a nominal subject is simply in the nominative (np(nom))[1]. For the structural case underlying the genitive of negation we use a mnemotechnical symbol np(accgen), since this type of phrase is realised in the accusative or in the genitive, depending on negation.

A similar mechanism is used to represent so called partitive nominal phrases, which can be realised in the accusative or in the genitive case with a slight difference in the meaning. This type of phrase is represented in Walenty with a special value part for the grammatical case.

The dictionary provides semantic classification of some adverbial-like arguments (e.g., ablative and adlative), denoted in Walenty as xp(...). Such valency positions can be filled mainly with adverbs and prepositional phrases, but the subtype of xp explicitly specifies a semantically motivated set of allowed phrase types. For example xp(abl) – ablative phrase, marking departure point of a motion – can be realised (among others) by adverbs *stąd* 'from here', *znikąd* 'out of nowhere', or prepnp(z,gen) – phrases with the preposition *z* 'from'. Adlative phrases xp(adl) denote point of arrival: *tutaj* 'here', *naprzód* 'forward', prepnp(do,gen) – *do* 'towards', complex preposition comprepnp(w kierunku) 'in the direction of', or even clauses, e.g. cp(rel[dokąd;gdzie]) – a relative clause limited to two relative pronouns *dokąd* 'where to' and *gdzie* 'where'.

In total, there are 10 specific subtypes of xp – expressing time, duration, place, starting or ending point, path, tool, manner, cause, or aim.

The following example depicts one of syntactic schemata for the verb określić 'specify/determine' that is used in the tree of Figure 1:

subj	obj		
np(nom)	np(accgen) cp(int) cp(że) ncp(accgen,int) ncp(accgen,że)	prepnp(w,loc)	xp(mod)

Each column represents a single syntactic position. Two positions are labelled: subject subj (the argument in this position influences morphological features of the finite verb) and passivable object obj (the argument in this position turns into a subject in passive voice).

Phrases that can fill a given position are specified by their non-terminal category and selected grammatical features. In the example, the subject position can be realised by a nominal phrase in the nominative case np(nom). The object position can be realised by a nominal phrase in a structural case np(accgen) or an interrogative clause cp(int) or a clause with the complementizer że 'that' – cp(że), and two other. (For a full list of available phrase types see (Hajnicz et al., 2016b).)

Walenty describes coordination of syntactically different arguments within a single syntactic position (so called unlike coordination). The fact that given phrase type specifications are listed within a single position is to mean that arguments of these types can be coordinated. An example of such coordination can be seen in Figure 1, where a nominal

[1] Świgra does not follow Przepiórkowski's concept of numeral subjects being in the accusative (Przepiórkowski, 2004).

phrase *rodzaj infekcji* 'kind of infection' gets coordinated with a clause *co ją powoduje* 'what causes it'. If the coordination was not possible, separate schemata with respective phrase types would be given.
The following table shows a schema for the other verb occurring in the example sentence — ZBADAĆ 'examine':

subj	obj	
np(nom)	np(accgen) ncp(accgen,int) cp(int)	xp(instr)

The third position in this schema can be realised by a phrase of type xp and subtype instr – instrument/tool.
Other phenomena represented in Walenty include syntactic control and raising, but these are not implemented in Świgra yet.
Walenty includes a rich phraseology component (Przepiórkowski et al., 2014a; Hajnicz et al., 2016b). It aims at precise representation of the structure of lexicalised arguments. In particular, it is used to represent all acceptable dependants of complex prepositions.
In Walenty, due to the free word order of Polish, the order of positions within a schema and the order of argument types within a position is not important.
Valency schemata given by Walenty are maximal – the dictionary does not list possible sub-schemata of a given schema. In Polish most of arguments are optional (in particular subjects are often omitted).
The syntactic layer of Walenty is being currently complemented with semantic frames (Hajnicz et al., 2016a).

4. Adapting Świgra to Walenty

Adopting Walenty was an obvious step in the development of Świgra but it meant that some changes needed to be introduced in the grammar to take advantage of the more detailed description. Simple changes included translating the symbols used in the old dictionary, which were based on Polish abbreviations for some grammatical categories, to those used in Walenty (Latin/English based).
A much more fundamental change was the introduction of coordination within syntactic positions. An example of such coordination can be seen in Fig. 1. The phrase *rodzaj infekcji* 'kind of infection' is analysed as a nominal phrase fno in the accusative, which turns into a required phrase fw of type np(accgen). The sentence *co ją powoduje* 'what causes it' becomes a clausal phrase fzd of type int – interrogative (*co* 'what' is an interrogative pronoun), and then a required phrase fw of type cp(int). These two required phrases get coordinated to become a phrase of type [np(accgen),cp(int)].
A mechanism was introduced that checks that this composite type is a subset of the appropriate position in some schema for the given verb. As can be seen, it is the case with the quoted schemata of the verbs OKREŚLIĆ 'determine' and ZBADAĆ 'study'.
What makes the example even more interesting, the two verbs are also coordinated and form a complex verbal phrase *określić i zbadać*. Syntactic schemata for both verbs differ and even the respective obj positions differ. Nonetheless, both schemata contain a position that is a superset of the type [np(accgen),cp(int)], which allows to accept the sentence. As can be seen, the use of Walenty's schemata can be quite complicated, which means an efficient way of using those had to be developed (Woliński, 2015).
Walenty applies the label subj not only to nominal phrases in the nominative, but also to some other phrases, e.g. cp(że). It was decided to interpret subjects in the same way in Świgra, which means new rules had to be added for those realisations. As a result, much fewer verbs are inherently subjectless in this new interpretation.
New rules had also been added to implement special types of arguments present in Walenty: nonch, refl, cp(żeby2) and complex prepositions.
Semantically motivated phrase types xp(...) had also to be implemented. The old dictionary used a much less precise general type advp, so respective rules had to be replaced with ones defining the possible subtypes.
To use lexicalised schemata from Walenty it was necessary to make the lemma of the head of each phrase available. In DCG information is only available "locally" – a grammar rule can only access the category and the information available as attributes of a given node. So it was necessary to add attributes that carry the information on the lexical head along the 'head branch' of each subtree. With these changes Świgra now uses phraseological schemata of Walenty (although the complete analysis of embedded modifiers of lexicalised items is not yet performed).

5. Adapting Składnica to Walenty

The core reason to use Walenty in Świgra was to introduce its rich information to the Składnica treebank. But that required some non-trivial operations to be performed on the treebank.
The system used to manually disambiguate trees generated with Świgra (Woliński, 2010) includes a module to automatically re-annotate a parse forest generated with a changed grammar preserving the tree previously chosen by annotators. However, in the form previously implemented, the system sought for a tree literally identical to the previously selected one. Due to new features from Walenty, some systematic changes had to be allowed between the old and the new trees. An algorithm was implemented that accepts the tree as matching if it differs only in the pre-specified way.
To make the upgrade procedure easier to manage, the changes required to adopt Walenty were split into a few sets of independent changes, which were applied incrementally. Each set of changes was tested against the treebank and necessary corrections were performed. The corrections involved the rules of the grammar, valency schemata of Walenty, or arguments selected for particular sentences in the treebank. This way all three resources were tested against each other.
In the first step, Walenty was mapped to a form close to the original dictionary with the intention to detect incompatible changes in valency schemata. At this stage Polish names of grammatical categories were used; all xp(...) phrases were mapped to generic advp; and lexical heads were introduced in the grammar and confronted with lexicalised schemata of Walenty. After re-parsing of the corpus, schemata from

Walenty have been confronted with arguments selected by annotators.

At the beginning of procedure there were 10,673 accepted trees in Składnica. The tree previously accepted by the annotators was found among new parses in 10,193 cases (95.5%). For the remaining 480 sentences (4.5%) the parser using Walenty did not produce a compatible tree (in 255 cases (2.4%) the new parse forest was empty). Preliminary analysis has shown that these sentences exhibit several problems including errors both in Składnica and in Walenty. In particular, for some verbs the two dictionaries differ whether a given dependent should be considered a complement or an adjunct. We have decided to upgrade the rest of the treebank and present those problematic sentences for a new assessment of treebank annotators.

In following steps, which were mostly automatic, the symbols used for types of phrases were made consistent with Walenty and the subj label was added to respective phrases. The last step was devoted to the introduction of semantically motivated xp(...) phrases. The advp specification in the old dictionary was very general: this type of phrase could be realised by any adverbial phrase or any prepositional-nominal phrase prepnp. The annotators were free to decide whether a particular prepositional phrase can be interpreted as advp in a given context. We expected many problems in matching these types.

It turned out that in about 300 sentences some of the advp phrases in the old trees did not match any subtype of xp in the new ones. The list of sentences with this problem was analysed and the problems resolved in one of the following ways:

- the old advp was replaced in the treebank with a specific prepositional phrase in accordance with a schema present in Walenty,
- a schema of Walenty needed to be amended by a particular subtype of xp or prepnp phrase,
- the offending phrase was changed from required fw to free fl (adjunct) in the treebank,
- a new realisation for some subtype of xp had to be added.

Another type of a problem, that showed up in the process, was ambiguity of the advp specification. Some phrases can be interpreted as xp of various subtypes. For example *gdzieś* 'somewhere' can be xp(loc) – locative or xp(adl) – ablative. The phrases *przez most* 'through a bridge' and *przez godzinę* 'during one hour' both are prepnp(przez,acc) in Polish, so they both qualify as realisations of xp(perl) or xp(dur), but only the first is really perlative (expresses a path of a movement) and the second – durative. The list of about 200 sentences containing such ambiguities was given to an expert, who decided for each of them, which interpretation to choose.

We are aware that some problems remain after the update procedure. Nominal phrases are not typical realisations of xp phrases. The only exception is np(inst), which is a possible realisation of xp(dur) (*czekać godzinami* 'to wait for hours') and xp(perl) (*jechać drogą* 'to drive along the road'). Such realisations were absent in the old valency dictionary, so such phrases were considered adjuncts in the treebank. Moreover, for some verbs of movement, which allow for an xp(perl) argument, the schemata of Walenty contain both xp(perl) and np(inst) (*jechać samochodem* 'to drive a car'). Only the np(inst) argument was present in the old dictionary, and could be used for both types of arguments. Unfortunately, occurrences of these problems could not be detected automatically. To make the annotation consistent with Walenty some more manual corrections will be needed.

In total, Świgra with the changes described in Section 4 was able to accept 14,103 sentences of the Składnica corpus (70.5%), while the version with the old dictionary accepted 13,194 (66%). These newly accepted sentences include many sentences with verbs missing in the old dictionary, but also interesting examples of various coordinated phrases. These sentences are currently being assessed by the annotators.

6. Conclusions

Składnica is the first treebank of Polish of a considerable size. The resource is now coupled with an independently developed valency dictionary, which marks an important turning point in its development. The fact that Walenty is actively maintained makes further development of the parser easier. From the other point of view Składnica provides verification for schemata of Walenty.

The current version of Składnica can be downloaded from the address `http://zil.ipipan.waw.pl/Składnica`. The new version is also already available in the treebank search engine: `http://treebank.nlp.ipipan.waw.pl/`.

As said above, the treebank is being currently enlarged with the sentences accepted thanks to Walenty. We hope to reach the level of 70% analysed sentences in a few months.

Future plans for Świgra and Składnica include deployment of the non-verbal part of Walenty (this is are relatively easy task since non-verbal schemata are simpler and use the same types of phrases as verbs). The new version of Składnica will also be converted to the dependency form and used for training dependency parsers. An interesting question is whether the new features of the treebank (in particular types of xp phrases) can help in training statistical disambiguation tools and parsers. Another direction of development is to use the semantic layer of Walenty to generate predicate-argument structures using semantic role labels. In fact, the work in this direction has already started.

Acknowledgements

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

Bibliographical References

Hajnicz, E., Andrzejczuk, A., and Bartosiak, T. (2016a). Semantic layer of the valence dictionary of Polish *Walenty*. In Nicoletta Calzolari, et al., editors, *Proceedings of the Tenth International Conference on Language*

Resources and Evaluation, LREC 2016, pages 2625–2632, Portorož, Slovenia. ELRA, European Language Resources Association (ELRA).

Hajnicz, E., Patejuk, A., Przepiórkowski, A., and Woliński, M. (2016b). Walenty: słownik walencyjny języka polskiego z bogatym komponentem frazeologicznym. In Karolina Skwarska et al., editors, *Výzkum slovesné valence ve slovanských zemích*, pages 71–102. Slovanský ústav AV ČR, Prague.

Pereira, F. and Warren, D. H. D. (1980). Definite clause grammars for language analysis–a survey of the formalism and a comparison with augmented transition networks. *Artificial Intelligence*, 13:231–278.

Przepiórkowski, A., Bańko, M., Górski, R.L., and Lewandowska-Tomaszczyk, B., editors. (2012). *Narodowy Korpus Języka Polskiego*. Wydawnictwo Naukowe PWN, Warsaw.

Przepiórkowski, A., Hajnicz, E., Patejuk, A., and Woliński, M. (2014a). Extended phraseological information in a valence dictionary for NLP applications. In *Proceedings of the Workshop on Lexical and Grammatical Resources for Language Processing (LG-LP 2014)*, pages 83–91, Dublin, Ireland. Association for Computational Linguistics and Dublin City University.

Przepiórkowski, A., Hajnicz, E., Patejuk, A., Woliński, M., Skwarski, F., and Świdziński, M. (2014b). Walenty: Towards a comprehensive valence dictionary of Polish. In Nicoletta Calzolari, et al., editors, *Proceedings of the Ninth International Conference on Language Resources and Evaluation, LREC 2014*, pages 2785–2792, Reykjavík, Iceland. ELRA.

Przepiórkowski, A., Skwarski, F., Hajnicz, E., Patejuk, A., Świdziński, M., and Woliński, M. (2014c). Modelowanie własności składniowych czasowników w nowym słowniku walencyjnym języka polskiego. *Polonica*, XXXIII:159–178.

Przepiórkowski, A. (2004). O wartości przypadka podmiotów liczebnikowych. *Biuletyn Polskiego Towarzystwa Językoznawczego*, LX:133–143.

Seddah, D., Tsarfaty, R., Kübler, S., Candito, M., Choi, J. D., Farkas, R., Foster, J., Goenaga, I., Gojenola Galletebeitia, K., Goldberg, Y., Green, S., Habash, N., Kuhlmann, M., Maier, W., Marton, Y., Nivre, J., Przepiórkowski, A., Roth, R., Seeker, W., Versley, Y., Vincze, V., Woliński, M., and Wróblewska, A. (2013). Overview of the SPMRL 2013 shared task: A cross-framework evaluation of parsing morphologically rich languages. In *Proceedings of the Fourth Workshop on Statistical Parsing of Morphologically-Rich Languages*, pages 146–182, Seattle, WA. Association for Computational Linguistics.

Świdziński, M. and Woliński, M. (2010). Towards a bank of constituent parse trees for Polish. In Petr Sojka, et al., editors, *Text, Speech and Dialogue: 13th International Conference, TSD 2010, Brno, Czech Republic*, volume 6231 of *Lecture Notes in Artificial Intelligence*, pages 197–204, Heidelberg. Springer-Verlag.

Świdziński, M. (1992). *Gramatyka formalna języka polskiego*. Rozprawy Uniwersytetu Warszawskiego. Wydawnictwa Uniwersytetu Warszawskiego, Warszawa.

Świdziński, M. (1994). Syntactic dictionary of Polish verbs. Manuscript, Uniwersytet Warszawski i Universiteit van Amsterdam.

Woliński, M., Głowińska, K., and Świdziński, M. (2011). A preliminary version of Składnica—a treebank of Polish. In Zygmunt Vetulani, editor, *Proceedings of the 5th Language & Technology Conference: Human Language Technologies as a Challenge for Computer Science and Linguistics*, pages 299–303, Poznań, Poland.

Woliński, M. (2004). *Komputerowa weryfikacja gramatyki Świdzińskiego*. Ph.D. dissertation, Institute of Computer Science, Polish Academy of Sciences, Warsaw.

Woliński, M. (2010). Dendrarium – an open source tool for treebank building. In Mieczysław A. Kłopotek, et al., editors, *Proceedings of IIS'2010*, pages 193–204. Wydawnictwo Akademii Podlaskiej.

Woliński, M. (2015). Deploying the new valency dictionary Walenty in a DCG parser of Polish. In Markus Dickinson, et al., editors, *Proceedings of the Fourteenth International Workshop on Treebanks and Linguistic Theories (TLT 14)*, pages 221–229, Warsaw. Institute of Computer Science, Polish Academy of Sciences.

Wróblewska, A. and Woliński, M. (2012). Preliminary experiments in Polish dependency parsing. In Pascal Bouvry, et al., editors, *Security and Intelligent Information Systems: International Joint Conference, SIIS 2011, Warsaw, Poland, June 13-14, 2011, Revised Selected Papers*, volume 7053 of *Lecture Notes in Computer Science*, pages 279–292. Springer-Verlag.

Language Resource References

Elżbieta Hajnicz *et al.* (2017). *Walenty valency dictionary of Polish*. Institute of Computer Science, Polish Academy of Sciences, `http://walenty.ipipan.waw.pl`, v. 1.2.

Marcin Woliński. (2011). *Składnica – a constituency and dependency treebank of Polish*. Institute of Computer Science, Polish Academy of Sciences, `http://zil.ipipan.waw.pl/Składnica`, v. 0.5.

Marcin Woliński. (2017a). *Składnica – a constituency and dependency treebank of Polish*. Institute of Computer Science, Polish Academy of Sciences, v. 0.8.

Marcin Woliński. (2017b). *Świgra 2 – a constituency parser of Polish*. Institute of Computer Science, Polish Academy of Sciences, v. 1.9.