Combining various degrees of supervision in PP-attachment disambiguation

Katarzyna Krasnowska, Adam Przepiórkowski

Institute of Computer Science, Polish Academy of Sciences
Jana Kazimierza 5, 01-248 Warsaw, Poland
kasia.krasnowska@gmail.com adamp@ipipan.waw.pl

Abstract

The aim of this paper is to present a novel approach to PP-attachment disambiguation where the features of a classifier include outcomes of two previously trained models: a standard supervised model and a robust partially (and very weakly) supervised model. In experiments performed on Polish data, we demonstrate that this hybrid model fares significantly better than the models used as features.

1. Introduction

The PP-attachment problem consists in identifying correct attachment sites for prepositional phrases occurring in natural language utterances. The problem was intensively investigated in 1990s, and while the major interest nowadays lies in more general approaches to probabilistic parsing, solutions to the distilled PP-attachment problem are still useful in some applications. The context of the current work is the manual development of large deep grammars of Polish within DCG (Warren and Pereira, 1980) and LFG (Dalrymple, 2001; Crouch et al., 2011) formalisms, where no treebanks large enough to train full probabilistic models exist yet, but sufficient resources are available to train PP-attachment disambiguation modules.

The typical formulation of a single instance of the PP-attachment problem is a quadruple $(v, n, p, n2)$, with the verb v and the noun n being two possible attachment sites for a phrase headed by the preposition p[1] whose nominal argument is headed by the noun $n2$. This paper presents a novel approach to PP-attachment disambiguation where the features of a classifier include outcomes of two previously trained models: a standard supervised model and a robust partially supervised model. We show that – in experiments based on Polish data – this hybrid model fares significantly better than either of the models used as features and that it approaches the human ceiling.

2. Previous experiments

This work builds directly on previous experiments leading to the construction of two PP-attachment classifiers for Polish.[2] One is fully supervised – referred to as FSC below – trained on KRZAKI,[3] a manually annotated dependency corpus of Polish of around 20,000 sentences (294,269 segments). A set of 5734 $(v, n, p, n2)$ tuples was extracted automatically from this treebank and the resulting dataset was divided into 3 sections: 50% training (KRZAKI-TRAIN), 25% development (KRZAKI-DEV) and 25% for final testing (KRZAKI-TEST). The best-performing supervised model was a backed-off model (Collins and Brooks, 1995) augmented with a wordnet-based similarity measure and WSD. Its best result on development data was 72.5%, but it was somewhat unstable between different parameter settings. For this reason, another variant was incorporated in the hybrid system described below, which used similarity lists[4] (resource created at Wrocław University of Technology using SuperMatrix tool; Broda and Piasecki 2013), required no parameter tuning and achieved a comparable accuracy of 72.3%.

The other classifier used the automatic morphosyntactic annotation in *books* and *press* sections of the the full 1.8-billion-segment National Corpus of Polish (NKJP; Przepiórkowski et al. 2012). It is supervised only to the extent that a tagger which produced this annotation was trained on a manually annotated 1-million-word subcorpus of NKJP and that two regular expressions over these annotations were manually constructed to find 18 million probable verb attachment examples and 3 million probable noun attachment examples. Then, a procedure similar to that described by Hindle and Rooth (1993) was implemented to estimate two probabilities for each preposition occurrence: that a verb is a governor and that a noun is a governor. The larger of these two numbers determined the choice of the governor. Additionally, in order to deal with data sparsity, the $n2$ lemma was generalised to a wordnet semantic category,[5] based on the Polish wordnet SŁOWOSIEĆ (Piasecki et al., 2009), thus introducing some further (but still very indirect) supervision. The best result of this classifier, tested again on KRZAKI-DEV, was 75.5% (a 2 percentage points improvement over the version which ignores $n2$ instead of generalising it to a wordnet semantic category). Note that the accuracy of this partially (and rather weakly) supervised classifier – referred to as PSC below – is higher than that of FSC.

[1] Polish has some prepositions which have the same surface form but different case requirements and different meanings (when used as semantic prepositions). Throughout this text, unless explicitly stated otherwise, by *preposition* we will mean its surface form together with the case it requires.

[2] Numerous variants of these experiments are described in (Krasnowska, 2014). This section only presents the best of the methodologically cleanest results.

[3] `http://zil.ipipan.waw.pl/Krzaki`.

[4] `http://nlp.pwr.wroc.pl/en/tools-and-resources/msr-list`

[5] By a *wordnet semantic category* we mean a coarse-grained wordnet class such as *place*, *animal* or *event*.

3. Improving the datasets

In order to verify and possibly increase the quality of data, as well as to examine the impact of such improvement on the results, the PP-attachment tuples extracted from KRZAKI were re-annotated. The re-annotation task was performed in two stages. First, two independent annotators were presented with the tuples. Each tuple was accompanied by the sentence from which it was extracted as a context. The annotators had four answers to choose from: verb attachment, noun attachment, “both possible” (i.e. genuine ambiguity) and “both incorrect”. The cases where the annotators opted for different answers were settled by one superannotator. The final decision of the superannotator was not restricted to the two original answers, but could be chosen from all four options. As a result, each tuple was assigned one of the four possible answers, either unanimously by the annotators or by the superannotator’s intervention. The tuples with “both possible” or “both incorrect” answers were rejected and the remaining ones were accepted, even if the original KRZAKI annotation was different than the result of the current re-annotation – this dataset is referred to below as ANN+SUPER. Experiments were also performed on the initial set of tuples automatically extracted from KRZAKI and on a subset of ANN+SUPER consisting of those tuples for which the two annotators agreed:

- AUTO — dataset automatically extracted from Krzaki,
- ANN+SUPER — the tuples accepted by the annotators or the superannotator,
- ANN–UNAN — the tuples accepted unanimously by the two annotators.

The split into -TRAIN, -DEV and -TEST of the original dataset AUTO was preserved in the ANN+SUPER and ANN–UNAN datasets: a tuple in, say, the -DEV part of the AUTO dataset was also in the -DEV part of ANN+SUPER and ANN–UNAN (if it was accepted to these two datasets at all). For the purpose of experiments presented in this work, the KRZAKI-DEV part of each dataset was further split into 10 approximately equally-sized folds. The split of all tuples into -TRAIN, -DEV and -TEST, as well as the division the -DEV datasets into the 10 folds for cross validation, were performed in such a way that all tuples from a given sentence were kept in one fold (to make the task more realistic and a little more difficult). The number of tuples in -TRAIN and -DEV datasets used in this work in each of the 3 variants is given in Table 1.

variant	-TRAIN	-DEV
AUTO	2810	1504
ANN+SUPER	2564	1391
ANN–UNAN	2247	1219

Table 1: Sizes of KRZAKI -TRAIN and -DEV datasets.

4. Using the models in a meta-classifier

In the presented work, the predictions of PSC and FSC models, together with additional features related to the tuple or the context in which it occurred, were used as input for a “meta-classifier” constructed using a standard machine learning classifier. In this section, a set of features is proposed for the “meta-classifier”, the conducted experiments are described and their results are discussed.

4.1. Features

	feature		description
NEW	$\mathrm{P}_{\mathrm{LEMMA}}$	C	p’s lemma
	$\mathrm{V}_{\mathrm{LEMMA}}$	C	v’s lemma if one of the 3 most frequent, x otherwise
	$\mathrm{N}_{\mathrm{LEMMA}}$	C	n’s lemma if equal to *to* ‘it, this’, x otherwise
	$\mathrm{N2}_{\mathrm{LEMMA}}$	C	$n2$’s lemma if equal to *to* ‘it, this’, x otherwise
	$\mathrm{N}_{\mathrm{NUM}}$	C	n’s morphological number
	$\mathrm{N}_{\mathrm{CASE}}$	C	n’s morphological case
	$\mathrm{N}_{\mathrm{GEND}}$	C	n’s morphological gender
	$\mathrm{N2}_{\mathrm{NUM}}$	C	n’s morphological number
	$\mathrm{N2}_{\mathrm{CASE}}$	C	n’s morphological case
	$\mathrm{N2}_{\mathrm{GEND}}$	C	n’s morphological gender
	$\mathrm{V}_{\mathrm{OFFSET}}$	N	distance between v and p
	$\mathrm{N}_{\mathrm{OFFSET}}$	N	distance between n and p
	$\mathrm{PP}_{\mathrm{OFFSET}}$	N	distance between p and $n2$
	P1	C	binary value: is n’s direct governor the v or a preposition?
	FRAMES	N	number of v’s valence frames in WALENTY
	$\mathrm{FRAMES}_{\mathrm{PP}}$	N	fraction of v’s valence frames with a matching prepnp slot
	$\mathrm{FRAMES}_{\mathrm{XP}}$	N	fraction of v’s valence frames with a matching xp slot
MODELS	$\mathrm{V}_{\mathrm{PROB}}$	N	$P(p\|v)$ from PSC
	$\mathrm{N}_{\mathrm{PROB}}$	N	$P(p\|n)$ from PSC
	$\mathrm{ATT}_{\mathrm{PS}}$	C	attachment chosen by PSC with semantic categories
	$\mathrm{ATT}_{\mathrm{FS}}$	C	attachment chosen by FSC
	LEVEL	N	data similarity level of FSC

Table 2: List of features; N – numerical, C – categorical.

The feature set used in the experiments is presented in Table 2; the features are split into those related to previously trained MODELS and other – NEW – features. The model features express the decisions of the two models, the level of back off in the supervised model, and probabilities of p given v and n estimated by the partially supervised model from its training data.

The other features express standard word-level properties of $(v, n, p, n2)$ tuples (lemmata and morphological information retrieved from the morphosyntactic layer of Krzaki treebank, which was, in turn, transferred from the manually annotated subcorpus of NKJP), but also some basic context properties going beyond what’s expressed by the tuple – linear distances between the elements of the tuple and one structural property of the sentence: whether n is governed by some preposition or whether it is a direct dependent of v (these properties were also taken from Krzaki; in the case of unseen text, they could be easily obtained using very shallow parsing techniques).

The lemmata of v, n and $n2$ are given only for the three most frequent verbs: *być* ‘be’, *mieć* ‘have’ and *zostać* ‘become’, and for the single (pro)noun *to* ‘it, this’, since lemmata for all verbs and nouns would create an excessively large and sparse feature space. Moreover, the lexical information carried by lemmata was already used by the partially supervised and backed-off methods and is therefore (in a very indirect way) present in their predictions included in the dataset. It is worth noting that the 4 lemmata correspond to lexemes with very broad senses and selectional preferences. Therefore, it could be beneficial for a model for PP-attachment to be able to learn a possibly different treatment of these words.

The final subset of the NEW features, perhaps the most interesting in such an application, refers to the information in the Polish valence dictionary WALENTY (Przepiórkowski et al., 2014). These features let the classifier know whether the preposition p may head an argument (an element of a valence frame) of the verb v. There are two basic types of valence slots in the dictionary that can be filled by a PP. First, `prepnp(p,c)`, represents a valence requirement for a prepositional phrase headed by preposition `p` (with case `c`). The second is `xp(t)`, where `t` is a semantic phrase type (locative, temporal, manner, etc.); many such phrase types may be morphosyntactically realised as prepositional phrases headed by specific prepositions. The three numerical valence features tell how many different valence frames are assigned to v and how many of them admit a p-headed argument of either type.

4.2. Experiments

Different algorithms implemented in the Weka machine learning toolkit (Hall et al., 2009) were evaluated using 10-fold cross-validation. The division into folds mentioned above was used. The default settings were used for each algorithm. The tested algorithms – and their abbreviations used in Table 3 – are: `LibSVM2` (SVM; Chang and Lin 2011), `J48` (decision tree; DT; Quinlan 1993), `Random Forest` (RF; Breiman 2001), `ClassificationViaRegression` (CVR; Frank et al. 1998), `RandomSubSpace` (RSS; Ho 1998). The experiments were performed on data passed through different Weka filters: NONE – no filtering; BIN – all categorical features were binarised (replaced with sets of binary, 0/1-valued numeric features, each corresponding to one value of the original categorical feature); e.g. the $\mathrm{N_{GEND}}$ feature with 5 values: `m1`, `m2`, `m3`, `f`, `n` would be replaced with 5 binary features $\mathrm{N_{GEND}}$=`m1`, ..., $\mathrm{N_{GEND}}$=`n`, exactly one of them equal 1 for any data instance; BIN+NORM – after binarising, all features were scaled to the 0–1 range; BIN+STAND – after binarising, all features were scaled to mean 0 and variance 1.

4.3. Results

Table 3 contains the accuracies (i.e. overall percentage of correctly classified examples in all 10 folds) for all experiments, and Table 4 summarises the results by showing the best accuracy result obtained for each KRZAKI variant and for each of the 3 tested feature sets: NEW – features unrelated to the two previous models, MODELS (abbreviated to MDLS in Table 3) – features related to the two models, ALL – the sum of NEW and MODELS.

Note that the results presented above are not directly comparable to the previous experiments with the “base” PP-attachment models, reported in Section 2., where all the examples from KRZAKI-DEV were test instances and other data served as training material. Here, there are two “stages” of training: training the partially supervised and backed-off method on NKJP data and KRZAKI-TRAIN, respectively, and then training one of Weka’s algorithms on part of the new, KRZAKI-DEV-based dataset in each cross-validation fold. Therefore, two new factors appear in this experiment: (1) an additional level of training with new features and (2) additional training data (the 9 folds of KRZAKI-DEV).

In order to meaningfully compare the performance of this hybrid method with the results cited in Section 2., the base methods were tested on each fold of KRZAKI-DEV, with the remaining 9 folds available as additional training data, and the result was a weighted average from the 10 folds. In case of the partially supervised method, this amounts to replicating the previous experiment since the model is, by design, trained on data different from NKJP and cannot use the tuples extracted from KRZAKI. Moreover, the weighted average accuracy over the 10 folds is equivalent to accuracy calculated in a single run on whole KRZAKI-DEV dataset. However, the backed-off model could effectively use each 9 folds as additional training material, just like the supervised component of the combined method. Since two new variants of KRZAKI (ANN+SUPER and ANN–UNAN) were introduced in this work, we conducted a total of 6 experiments, one for each variant of the data and each base method. The results are presented in Table 5.

When only the NEW features are used, the performance is the worst. It is however worth pointing out that the best results obtained using these features are similar to (albeit slightly worse than) those yielded by the FSC base model. This is interesting in the light of the fact that the NEW features contain almost no lexical information apart from the lemmata of prepositions and just a couple of the most frequent verbs and nouns. Instead, they are mostly morphosyntactic, with addition of some valence information and some very basic context information that is easy to obtain without any additional resources. The use of a machine learning algorithm of choice with these features makes for a rather “lightweight” model, as opposed to the two base methods, which required retrieving a large number of occurrence counts (particularly massive in the partially supervised case) from the respective training data and storing them for the use of the model.

Using only the MODELS features as a basis for classification, the hybrid classifiers manage to achieve an improvement of 1.9%–2.6% over the better-performing among the two base models (PSC). An even more substantial improvement is attained when both NEW and MODELS are used in the ALL feature set: the results are 4.1%–4.9% higher than those of PSC.

The accuracy increases steadily when dataset variant is set to AUTO, ANN+SUPER and ANN–UNAN. One exception is the hybrid method using NEW features which, for some classifier/filter combinations, performs worse on ANN+SUPER than on AUTO data. Apart from this irregularity, the general tendency is intuitive. The AUTO dataset

		AUTO			ANN+SUPER			ANN−UNAN		
filter	classifier	NEW	MDLS	ALL	NEW	MDLS	ALL	NEW	MDLS	ALL
NONE	SVM	67.6%	77.3%	76.9%	67.0%	78.3%	79.1%	68.5%	80.7%	81.8%
	DT	**73.7%**	77.3%	76.7%	70.3%	79.9%	80.5%	73.9%	80.4%	83.1%
	RF	71.3%	74.4%	77.5%	69.7%	77.2%	79.6%	75.4%	79.5%	83.4%
	CVR	71.2%	**78.1%**	78.8%	71.8%	79.6%	82.0%	74.3%	**82.4%**	85.1%
	RSS	73.4%	76.4%	77.8%	71.8%	77.6%	81.0%	73.8%	81.6%	84.4%
BIN	SVM	67.6%	77.3%	76.9%	67.0%	78.3%	79.1%	68.5%	80.7%	81.8%
	DT	71.5%	77.3%	76.3%	70.4%	79.9%	78.2%	74.0%	80.4%	83.3%
	RF	69.4%	75.5%	77.5%	71.9%	78.6%	79.9%	73.2%	78.9%	84.6%
	CVR	70.8%	**78.1%**	**79.6%**	69.7%	79.6%	81.7%	74.1%	**82.4%**	83.0%
	RSS	71.6%	77.7%	79.3%	71.4%	79.6%	81.4%	74.9%	80.6%	84.8%
BIN+NORM	SVM	67.2%	77.7%	75.7%	66.3%	76.6%	78.7%	69.5%	80.3%	81.3%
	DT	71.5%	77.3%	76.3%	70.4%	79.9%	78.2%	74.0%	80.4%	83.3%
	RF	70.6%	74.4%	77.4%	71.3%	77.9%	81.5%	73.6%	80.4%	83.7%
	CVR	70.8%	**78.1%**	**79.6%**	69.7%	79.6%	81.7%	74.1%	**82.4%**	83.0%
	RSS	72.4%	77.3%	79.2%	**72.5%**	78.2%	81.8%	75.2%	81.9%	**85.4%**
BIN+STAND	SVM	72.6%	78.1%	78.5%	73.0%	**80.4%**	**82.7%**	**75.9%**	81.4%	85.1%
	DT	71.5%	77.3%	76.3%	70.4%	79.9%	78.2%	74.0%	80.4%	83.3%
	RF	70.4%	73.1%	77.3%	72.0%	77.3%	81.9%	73.0%	79.8%	84.3%
	CVR	70.8%	**78.1%**	**79.6%**	69.7%	79.6%	81.7%	74.1%	**82.4%**	83.0%
	RSS	72.3%	76.6%	78.4%	72.4%	78.2%	81.3%	74.1%	81.1%	84.9%

Table 3: Accuracy of selected algorithms from Weka on KRZAKI-DEV. The best result obtained for each feature set and each dataset (i.e. the best result in each column) is given in boldface.

dataset	NEW	MODELS	ALL
AUTO	73.7%	78.1%	79.6%
ANN+SUPER	72.5%	80.4%	82.7%
ANN−UNAN	75.9%	82.4%	85.4%

Table 4: Best hybrid accuracy on each variant of KRZAKI-DEV for different feature sets.

dataset	FSC	PSC
AUTO	73.6%	75.5%
ANN+SUPER	75.1%	77.9%
ANN−UNAN	76.5%	80.5%

Table 5: Base method accuracies on each variant of KRZAKI-DEV.

is very probably the most noisy one, making it harder to learn from as well as to predict its (possibly partially erroneous) attachment sites. The ANN−UNAN data should be the easiest for the classifiers to cope with since it contains the cases which apparently posed no serious difficulty to human annotators. The results on ANN+SUPER being inferior to those on ANN−UNAN is not surprising since the former contain examples which needed a superannotator's intervention, therefore are expected to be harder. On the other hand, they represent a much more realistic scenario when the data, although thoroughly annotated, contain some problematic examples.

5. Related work and conclusion

The approach described in this paper follows the long line of publications describing various ways of combining fully supervised and partially (or not at all) supervised methods for resolving PP-attachment, starting with the manual combination of such models (Wu and Furugori, 1996; Volk, 2002), enriching them with WordNet data (Bharathi et al., 2005), features calculated from unlabelled data (Kawahara and Kurohashi, 2005; Coppola et al., 2011; Roh et al., 2011), and subcategorisation and FrameNet information (Olteanu and Moldovan, 2005). Also somewhat related to our approach is that independently proposed for Modern Greek by Nalmpantis et al. (2012), where a number of classifiers were trained on features similar to our NEW features and a meta-classifier was trained with these base classifiers as features. Various improvements stemming from the combination of supervised and partially or not supervised methods were reported; they are not cited here as – given that they were reported for languages different than Polish and for datasets very different from ours – they are not directly comparable to our results. To the best of our knowledge, there is no other work on the (isolated) problem of PP-attachment disambiguation in Polish – or any other Slavic language – that our approach could be compared to; in general, most work on hybrid approaches to PP-attachment disambiguation focuses on English, with some work on other Germanic and Romance languages.

Compared to previous work, our approach is novel in various respects. First of all, it combines different degrees of supervision via the use of a meta-classifier, some of whose features express the decisions of base models. This combination, its evaluation and comparison to other results is methodologically careful, to ensure that all compared systems – whether they are base classifiers requiring a single round of training or meta-classifiers requiring two rounds – are trained on comparable data and not tested on data even very indirectly used for training.

Second, the system described here combines such model features with valence features, taken from a manually-constructed valence dictionary, which indicate the possibility that the preposition heads an argument of the verb; the evaluation of feature importance (not reported here for lack of space) shows that such valence features are among the most important features increasing the quality of the meta-classifier above the quality of a manual combination of the two base classifiers. Perhaps surprisingly, the simple verbal lemma feature, which only provides the lemma of the three most frequent verbs and distinguishes them from other verbs, also turned out to be one of the most important features. On the other hand, morphosyntactic features and features expressing some basic context properties going beyond the information in a PP-attachment tuple turned out to be relatively unimportant.

Third, the work reported here is concerned with a language different from English, which provides the empirical basis for the vast majority of experiments on PP-attachment disambiguation. Working on English certainly facilitates comparison (e.g. by the use of the standard IBM data) and, hence, also increases the publishability of the results. However, it also constitutes a methodological danger, as methods reported in the literature become fine-tuned to a language with poor morphology and fixed word order. In our view it is important to develop methods well-suited for languages which are morphologically richer and linearly freer than not just English, but also German. Polish, together with a majority of Balto-Slavic languages, is just such a language.

Acknowledgements The paper is partially founded by the European Union from resources of the European Social Fund. Project PO KL “Information technologies: Research and their interdisciplinary applications”, and partially by the CLARIN-PL project.

References

Bharathi, Akshar, U. Rohini, P. Vishnu, S.M. Bendre, and Rajeev Sangal, 2005. A Hybrid Approach to Single and Multiple PP Attachment Using Wordnet. In *Second International Joint Conference on Natural Language Processing: Full Papers*.

Breiman, Leo, 2001. Random Forests. *Machine Learning*, 45(1):5–32.

Broda, Bartosz and Maciej Piasecki, 2013. Parallel, Massive Processing in SuperMatrix – a General Tool for Distributional Semantic Analysis of Corpora. *International Journal of Data Mining, Modelling and Management*, 5(1).

Chang, Chih-Chung and Chih-Jen Lin, 2011. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, 2:27:1–27:27. Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.

Collins, Michael and James Brooks, 1995. Prepositional Phrase Attachment through a Backed-Off Model. In *Proceedings of the Third Workshop on Very Large Corpora*.

Coppola, Gregory F., Alexandra Birch, Tejaswini Deoskar, and Mark Steedman, 2011. Simple Semi-supervised Learning for Prepositional Phrase Attachment. In *Proceedings of the 12th International Conference on Parsing Technologies*, IWPT ’11. Stroudsburg, PA, USA: Association for Computational Linguistics.

Crouch, Dick, Mary Dalrymple, Ron Kaplan, Tracy King, John Maxwell, and Paula Newman, 2011. XLE Documentation. `http://www2.parc.com/isl/groups/nltt/xle/doc/xle_toc.html`.

Dalrymple, Mary, 2001. *Lexical Functional Grammar*. San Diego, CA: Academic Press.

Frank, E., Y. Wang, S. Inglis, G. Holmes, and I.H. Witten, 1998. Using model trees for classification. *Machine Learning*, 32(1):63–76.

Hall, Mark, Eibe Frank, Geoffrey Holmes, Bernhard Pfahringer, Peter Reutemann, and Ian H. Witten, 2009. The WEKA Data Mining Software: An Update. *SIGKDD Explor. Newsl.*, 11(1):10–18.

Hindle, Donald and Mats Rooth, 1993. Structural Ambiguity and Lexical Relations. *Computational Linguistics*, 19(1):103–120.

Ho, Tin Kam, 1998. The Random Subspace Method for Constructing Decision Forests. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 20(8):832–844.

Kawahara, Daisuke and Sadao Kurohashi, 2005. PP-attachment Disambiguation Boosted by a Gigantic Volume of Unambiguous Examples. In *Proceedings of the Second International Joint Conference on Natural Language Processing*, IJCNLP’05. Berlin, Heidelberg: Springer-Verlag.

Krasnowska, Katarzyna, 2014. Different approaches to the PP-attachment problem in Polish. In Verena Henrich, Erhard Hinrichs, Daniël de Kok, Petya Osenova, and Adam Przepiórkowski (eds.), *Proceedings of the Thirteenth International Workshop on Treebanks and Linguistic Theories (TLT 13)*. Tübingen, Germany: Department of Linguistics (SfS), University of Tübingen.

Nalmpantis, Pavlos, Romanos Kalamatianos, Konstantinos Kordas, and Katia Kermanidis, 2012. Ensemble Learning for Low Resources Prepositional Phrase Attachment. *International Journal of Information Science*, 2(2):1–7.

Olteanu, Marian and Dan Moldovan, 2005. PP-attachment Disambiguation Using Large Context. In *Proceedings of the Conference on Human Language Technology and Empirical Methods in Natural Language Processing*, HLT ’05. Stroudsburg, PA, USA: Association for Computational Linguistics.

Piasecki, Maciej, Stanisław Szpakowicz, and Bartosz Broda, 2009. *A Wordnet from the Ground Up*. Wrocław: Oficyna Wydawnicza Politechniki Wroclawskiej.

Przepiórkowski, Adam, Mirosław Bańko, Rafał L. Górski, and Barbara Lewandowska-Tomaszczyk (eds.), 2012. *Narodowy Korpus Języka Polskiego*. Warsaw: Wydawnictwo Naukowe PWN.

Przepiórkowski, Adam, Elżbieta Hajnicz, Agnieszka Patejuk, Marcin Woliński, Filip Skwarski, and Marek Świdziński, 2014. 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 (eds.), *Proceedings of the Ninth International Conference on Language Resources and Evaluation, LREC 2014*. Reykjavík, Iceland: ELRA.

Quinlan, Ross, 1993. *C4.5: Programs for Machine Learning*. San Mateo, CA: Morgan Kaufmann Publishers.

Roh, Yoon-Hyung, Ki-Young Lee, and Young-Gil Kim, 2011. Improving PP Attachment Disambiguation in a Rule-based Parser. In *Proceedings of the 25th Pacific Asia Conference on Language, Information and Computation*.

Volk, Martin, 2002. Combining Unsupervised and Supervised Methods for PP Attachment Disambiguation. In *Proceedings of the 19th International Conference on Computational Linguistics - Volume 1*, COLING ’02. Stroudsburg, PA, USA: Association for Computational Linguistics.

Warren, D. H. D. and Fernando C. N. Pereira, 1980. Definite Clause Grammars for Language Analysis — A Survey of the Formalism and a Comparison with Augmented Transition Networks. *Artificial Intelligence*, 13:231–278.

Wu, Haodong and Teiji Furugori, 1996. Prepositional phrase attachment through a hybrid disambiguation model. In *COLING 1996 Volume 2: The 16th International Conference on Computational Linguistics*.