plWordNet in Word Sense Disambiguation

Maciej Piasecki, Paweł Kędzia and Marlena Orlińska
Wrocław University of Technology, Poland
{Maciej.Piasecki,Pawel.Kedzia,Marlena.Orlinska}@pwr.edu.pl

Abstract

The paper explores the application of plWordNet, a very large wordnet of Polish, in weakly supervised Word Sense Disambiguation (WSD). Because plWordNet provides only partial descriptions by glosses and usage examples, and does not include sense-disambiguated glosses, PageRank-based WSD methods perform slightly worse than for English. However, we show that the use of weights for the relation types and the order in which lexical units have been added for sense re-ranking can significantly improve WSD precision. The evaluation was done on two Polish corpora (*KPWr* and *Składnica*) including manual WSD. We discuss the fundamental difference in the construction of both corpora and very different test results.

1 Introduction

Large wordnets are often treated as sense inventories that describe and enumerate word senses. If we want to process texts at the level of wordnet senses, a very useful operation, we first must map text words to those senses, i.e. to perform Word Sense Disambiguation (henceforth WSD). This is only trivial for monosemous words. WSD methods built upon supervised Machine Learning achieve good accuracy but are intrinsically impractical in their dependence on corpora that have been manually disambiguated with respect to word senses. Needless to say, such corpora are very laborious to annotate.

Weakly supervised WSD methods that use a wordnet as the basic knowledge source, but do not depend on a manually annotated corpus, can fully utilise wordnet senses, i.e. they can in theory assign any sense stored in a wordnet to words in text. So, in spite of their lower precision they seem to be noteworthy as a potentially practical solution. Most wordnet-based weakly supervised WSD methods are based on the idea of spreading activation in the wordnet graph, where the initial activation comes from the words in a textual context.

Several methods based on this general scheme were proposed. A short overview is presented in Section 2. Most such methods were developed and tested on Princeton WordNet (PWN) (Fellbaum, 1998) that is slightly different than plWordNet (Piasecki et al., 2009, Maziarz et al., 2013a), currently the world largest wordnet. First attempts to transfer the methods with good performance on PWN to plWordNet (Kędzia et al., 2015) were encouraging; the performance is relatively close to the performance of the supervised methods observed for Polish on limited test sets (Baś et al., 2008, Młodzki and Przepiórkowski, 2009). In addition to the differences between both wordnets, PWN has been enriched with various other resources in order to obtain better performance of unsupervised WSD. First of all, additional links between synsets were created on the basis of the manually disambiguated SemCore corpus (Miller et al., 1993). Such links have contributed significantly to the increase of WSD performance. There is no Polish corpus similar to SemCore.

The goal of the work presented here is to explore the structure and specific properties of plWordNet in order to improve the precision of the WSD methods based on the spreading activation in the wordnet graph, here the plWordNet graph.

In the rest of the paper, first we will briefly overview the existing wordnet-based unsupervised WSD methods, including their known applications to plWordNet. Next, the plWordNet model will be discussed and compared with PWN from the perspective of utilising different features in WSD method. On this basis, several possible versions of unsupervised WSD will be introduced. Finally, we will present data sets used in the evaluation and the results achieved for different settings used in WSD methods. Based on the results, we will analyse the the specific properties of plWordNet and its development process and its influence on wordnet-based unsupervised WSD methods for Polish.

2 Wordnet-based WSD

Unsupervised WSD methods (Pantel, 2003) use corpora to induce word senses and tune mechanisms for assignment of the induced senses to words. However, it is difficult to map the induced word senses to the wordnet. Weakly supervised WSD that are based on a wordnet as the knowledge base work directly on wordnet synsets and do not depend on manually disambiguated corpus.

Lesk's algorithm (Lesk, 1986) can be applied to textual definitions constructed on the basis on of synsets, e.g. from glosses, examples and synset members. The definitions are next compared with the occurrence contexts of words. Different similarity measures can be applied. The main problems are limited lenghts of the constructed definitions and high computational complexity, because many word sets must be compared.

Weakly supervised wordnet-based WSD algorithms assume that if we map words senses pertaining to a text fragment onto the wordnet graph, we can expect that the "hits" are located in short distances (in terms of paths) from each other in the wordnet graph. Moreover, we can use a kind of spreading activation algorithm in order to move this information along the wordnet graph, analyse the "hot" areas and identify word sense, i.e. lexical units (LUs),[1] located in them or close to them. Those LUs should be the most likely senses for words in the text. There are several parameters to set in this general scheme: the initial activation (text words vs LUs), spreading algorithm (topology and relations) and identification of association between "hot" areas and LUs to be chosen. Various methods propose a range of decisions.

Weakly supervised WSD methods are mostly based on the PageRank algorithm (Page et al., 1999) for spreading. Mihalcea et al. (2004) proposed application of the original PageRank to WSD called Static PageRank.

Page Rank algorithm (henceforth PR) is an iterative method for ranking nodes in the graph G. In WSD the nodes in G represent synsets and the edges of G correspond to wordnet relations (between synsets and in other case between synsets and between LUs). The spreading is done iteratively in the following way:

$$\mathbf{P}^{(\mathbf{new})} = cM\mathbf{P}^{(\mathbf{old})} + (1-c)\mathbf{v} \qquad (1)$$

$M_{N\times N}$ ins the adjacency matrix of the wordnet graph with N nodes (synsets), where $m_{ij} = \frac{1}{d_i}$ if the edge from the node s_i to s_j exists, 0 otherwise; d_i is degree of the node s_i (representing the synset i); where c is the damping factor; $\mathbf{v_{N\times 1}}$ is the vector of the initial scores for nodes and $\mathbf{P_{N\times 1}}$ is a vector of node scores updated in every iteration. In Static PageRank (SPR) all values in $\mathbf{v}$ are equal $1/N$.

Agirre and Soroa (2009), Agirre et al. (2014) proposed a modified version called Personalised PageRank (PPR) in which the values in $\mathbf{v}$, called personalised vector, depends on the text context of the disambiguated word. The non-zero score values are assigned to those nodes which are contextually supported. In PPR all words from the context are disambiguated at once. The $\mathbf{v}$ values are equal to:

$$\mathbf{v}[i] = \frac{1}{\frac{CS}{NS(i)}}, \quad i = 1, 2, ..., N \qquad (2)$$

where CS is the number of different lemmas in the context, $NS(i)$ – the number of synsets sharing the same context lemma with the synset i.

Agirre and Soroa (2009), Stevenson et al. (2012) proposed a modified version of PPR called Personalised PageRank Word-to-Word (PPR_W2W), in which a word to be disambiguated is excluded from the occurrence contexts, i.e. all synsets of this word have initial scores in $\mathbf{v}$ set to zero. Thus, PPR_W2W cannot be run once for all ambiguous words in the context. The vector $\mathbf{v}$ must be initialised individually for each ambiguous word in the context – this is a disadvantage of PPR_W2W. A potential advantage is the removal of the effect of mutual amplification of the closely connected senses of the word being disambiguated. The best results (measured in recall) are obtained on the *Senseval-2* dataset for a graph built from WordNet 1.7 and eXtended WordNet (Harabagiu et al., 1999). For nouns the best results are obtained using PPR (recall 71.1%), for verbs and adjectives with PPR_W2W recall was between 38.9% and 58.3%. For adverbs SPR achieved the best result of 70.8%. The best result

[1] See Section 3 for more on LUs.

for nouns, 71.9%, was achieved by PPR_W2W on the basis of the combination of WordNet 3.0 with disambiguated glosses.

In (Kędzia et al., 2014), SPR algorithm for Polish was based on plWordNet 2.1. The graph consisted of synsets linked by edges representing a selected subset of the synset relations. The precision on nouns (43%) and verbs (28%) was low in comparison to the works for English. The algorithm was evaluated on the *KPWr* corpus of Polish discussed in Section 5. In the second version, a Measure of Semantic Relatedness was utilised to add links to plWordNet. The measure had been extracted automatically from a large corpus of 1.8 billion words. However, there was no improvement: the precision for nouns was 37% and 27% for verbs. Nevertheless, we observed that even a WSD method of limited precision can be helpful in improving the performance of text clustering.

Next we adapted several algorithms: SPR, PPR and PPR_W2W – to Polish resources Kędzia et al. (2015). plWordNet 2.2 was used with all synset relations for the edges. Due to the lack of word-sense disambiguation of glosses, no additional synset links could be added. The achieved precision (on *KPWr*) was in the range 42.79%-50.73% for nouns and 29.79%-32.94% for verbs. PPR_W2W produced the best results. We also tested different variants of combining plWordNet with the *Suggested Upper Merged Ontology* (SUMO) (Pease, 2011) on the basis of the mapping constructed in (Kędzia and Piasecki, 2014). All three PR-based algorithm were evaluated. A slight improvement of the precision for nouns up to 50.89% for PPR_W2W could be observed when the two joined graphs were treated as one large graph.

3 plWordNet properties

plWordNet is a very large wordnet built independently from PWN and expresses several unique features. Word senses are represented in plWordNet as *lexical units* (LUs), i.e. pairs: lemma[2] plus sense identifier. LUs are the basic building blocks of plWordNet, but one LU belongs to exactly one synset. plWordNet includes about 40 main types of lexico-semantic relation. Half of them links synsets, the rest directly link LUs (Piasecki et al., 2009, Maziarz et al., 2012, 2013a, Piasecki et al., 2013). Many relations, e.g. meronymy, have subtypes, so the total number of lexico-semantic relations in plWordNet 2.3 exceeds 90.

The detailed description of the model underlying plWordNet can be found in (Maziarz et al., 2013b), below we present only a concise overview due to the space limit. LUs that share a set of constitutive lexico-semantic relations are grouped into *synsets* that are considered to consists of *near synonyms*. Synset relations are notational abbreviations for the relations shared between LUs from the linked synsets. The relations are the basic means of describing word senses. Different types of relations express different semantic associations, and provide different semantic information. This properties can be explored in WSD to improve the use of knowledge during spreading activation in the graph.

plWordNet provides as well some additional means of semantic description: *stylistic registers*, *glosses* and *use examples*. Stylistic registers signal pragmatic constraints on the use of LUs. However, such subtle differences are difficult to explore in WSD methods, so we have not done it. Glosses in plWordNet are comments to the LUs (not to synsets like in PWN) provided for a human reader in order to explain the motivation behind the given word sense and clarify its difference from other senses of the same lemma. Glosses are short descriptions but they are not proper lexicographic definitions and are much less elaborated from the point of view of their application in Lesk's algorithm (Lesk, 1986). Glosses are intended to be secondary and additional to the lexico-semantic relations that are the primary tool for the description of the lexical meanings in plWordNet, e.g. the genus information is expressed by hypernymy and should not be provided in a gloss. As such they have been added only to a subset of LUs. In addition to glosses, LU can be described by one or more use examples. They are also focused on human readers, but they can be used in WSD as an additional source of information. There have been not attempts so far to disambiguate word senses in the plWordNet glosses and examples.

plWordNet has been automatically mapped onto SUMO with high precision. The extended graph, plWordNet plus SUMO, has been already used in WSD with positive signals, discussed in Section 2.

plWordNet LUs are not clustered into semantic

[2] A lemma is a basic morphological form representing a group of word forms that have the same meaning but differ in the values of the morphological categories.

domains, but only into PWN-like, i.e. domains that correspond to the lexicographer files introduced in early stages of PWN development (Fellbaum, 1998). They do not seem to provide important knowledge for WSD.

Finally, there is no information about the frequency or salience of LUs, e.g. in comparison to other LUs of the same lemma. Numerical identifiers of LUs and the order of synsets in the plWordNet database mostly originate from the order in which editors introduced them into the database.

4 Exploring plWordNet in WSD

Taking as a starting point the work of Kędzia et al. (2015) and the observations in the previous section, we explored several ways of using the knowledge present in plWordNet to improve WSD performance.

4.1 Glosses and Examples

As the number of glosses and examples has been increased in the version 2.3 of plWordNet[3] we can apply Lesk's algorithm in a straightforward way – further on called basic Lesk's:

1. For a word w to be disambiguated, we select all synsets s_i that include LUs with lemma identical to the lemma of w.

2. Description sets $D(s_i)$ encompass all lemmas that are included in glosses and examples describing LUs from s_i, as well lemmas from s_i.

3. For each occurrence of w a context set $C(w)$ is collected, such that it contains all lemmas from the fixed size context of the w occurrence.

4. s_i such that the set $D(s_i)$ that have the maximal intersection with $C(w)$ is selected as the sense of the given occurrence of w.

The results obtained with the basic Lesk's algorithm are presented in Table 5.

4.2 Structural Description

In all experiments presented in (Kędzia et al., 2015) the wordnet graph was treated as a direct but uniform graph, i.e. every relation link was represented in the same way independent of the relation type. In order to increase the density of the graph LU relations were mapped on the synset level, i.e. if there was a link between LUs, then a link between their synsets was added. However, different relations represent different types of semantic association and provide different descriptions for the elements (synsets or LUs) they are attached to. On the basis of preliminary experiments, we assumed that synset relations and LU relations convey information of different importance for WSD and we assigned different weights to both types of links: $w_{LU} = 0.3$ for LU relations and $w_S = 0.7$ for synset relation[4]. The assigned weights can be next used in the spreading activation algorithm.

4.3 Sense order

In the case of highly polysemous words, some word senses located close to each other in the word graph are difficult to be distinguished. However, for practical applications, sometimes there is no need to differentiate such closely related word senses. So, we also tested partial WSD in which the top-ranked LUs within the range of $k = 30\%$ of the maximal score from the WSD algorithm were selected as a joint result. In a natural way, this relaxation of the task resulted in significantly improved precision.

It is well known that the most frequent sense baseline is difficult to be beaten by WSD. This is due to the mostly skewed distribution of word senses, in which one or few senses dominate among occurrences. Having LUs ordered according to their frequency in plWordNet, we could use this information to boots WSD performance. However, both Polish corpora annotated with word senses are much too small to provide such data. Regardless, LUs are numbered in plWordNet according to the order in which they have been added for the given lemma. The detailed guidelines for plWordNet editors say nothing about the order in which LUs should be defined[5], and our null hypothesis was that this would be almost a random factor from the point of WSD, i.e. the use of this information should not have any positive effect on the WSD performance. Nevertheless, we suspected that the null hypothesis does not match the

[3] However, most glosses take the form of short comments that are several words long.

[4] The highest weight of 1.0 was implicitly assigned to the synonymy relation that was not present in the graph structure but was expressed by synsets. The synsets collected activations from the occurrence of their members in the contexts of disambiguation.

[5] In fact it would be very difficult to define this in guidelines in a way resulting in consistent decisions of editors.

data and that the order of LUs identifiers is not accidental. We assumed that LUs with the highest identifiers represent the most salient senses of lemmas. Thus, selecting them should bring us closer to selecting the most frequent sense.

The relatively good results, presented in Section 5, seem to be in favour of rejecting the null hypothesis. They give some insights into the work of plWordNet editors, see Section 5.2.

5 Results and evaluation

Evaluation was based on applying the analysed algorithms to a corpus with manually disambiguated LUs (word senses). As a main criterion for evaluation we used the precision, calculated by comparing the LUs assigned by annotators and the algorithms, see Equation (3):

$$Pr = \frac{t}{t+f} \quad (3)$$

- t: the number of correctly disambiguated instances,
- f: the number of incorrectly disambiguated instances.

5.1 Experimental settings

Two corpora including disambiguated assignment of LUs to words were used during the evaluation. They have different character and were built by two independent teams but both are based on plWordNet, so that seems to be an interesting opportunity for evaluation.

The *KPWr* corpus (Corpus of the Wrocław University of Technology) (Broda et al., 2012), available under the Creative Commons license,[6] contains 1,127 documents (≈250,000 tokens) divided into 11 thematic categories. *KPWr* has been manually annotated and disambiguated at several levels: morpho-syntactic, syntactic relations, semantic relations, Named Entities. The documents are also described with manually assigned keywords and meta-information, like genre, author, etc.

In the case of 88 different lemmas, all their occurrences have been manually described with LUs from plWordNet by two annotators plus a super-annotator, who was responsible for solving conflicts. In the case of all lemmas annotated, their descriptions in plWordNet have been verified according to the defined set of LUs and the information provided for them, i.e. relation links, glosses and usage examples. In the case of lacking LUs (missing word senses), they have been added. If for some LU of one of the 88 lemmas there was no usage examples in *KPWr* or the number was very small, *KPWr* was expanded with some new texts. The WSD part of *KPWr* has been built in two stages, and in the second stage all previous annotations have been verified.

The WSD lemma set includes 58 different nouns and 30 verbs, see the statistics in Table 1. The lemmas were not selected randomly, but were chosen by linguists in such a way that all the lemmas are polysemous and represent different types of homonymy and polysemy. Moreover they vary according to numbers of possible lexical meanings, i.e. possible LUs. From the very beginning this set of WSD annotations was meant to be a gold standard for the evaluation of WSD methods.

	Total	**Nouns**	**Verbs**
Tagged words	88	58	30
Tagged instances	6048	3846	2202

Table 1: Statistic of WSD annotations in *KPWr*.

For 58 nouns and 30 verbs, the average number of word senses per word are 5.98 and 7.50 respectively. The standard deviation is 4.30 for nouns and 3.96 for verbs. The median of number of senses for the nouns is 5; 4 nouns have the number of senses equal to the median. 28 nouns have more senses than the median, and 26 have fewer. The median number of senses for the verbs is 6; 5 verbs have a number of senses equal to the median. 12 verbs have fewer senses than the median, and 13 have more. Thus, the annotated words are quite diversified and challenging for WSD.

Składnica (Hajnicz, 2014a), a treebank of Polish, is the second test set used during the evaluation. It includes 20,000 sentences among which more than 8,200 have manually assigned parse trees. For all these sentences, nouns, verbs and adjectives occurring in them have been manually mapped to LUs from plWordNet 1.6 (Hajnicz, 2014b). Proper Names included in them have been marked and semantically classified. Lemmas or word senses not found in plWordNet have been marked. *Składnica* includes sentences randomly selected from the open part of NKPJ (National Corpus of Polish) (Przepiórkowski et al., 2009). All sentences are described by identifiers and links to the original paragraphs, so it is possible to use

[6] http://nlp.pwr.edu.pl/kpwr

the whole paragraphs as contexts for WSD. *Składnica* differs significantly from *KPWr* with respect to words disambiguated with word senses: the selection was made at the level of sentences, so in the case of most lemmas only selected senses are covered. In *KPWr* all senses of every selected word are represented. Moreover, the *KPWr* builders paid attention to acquiring as many usage examples as possible for every senses, including those that are infrequent.

	Total	MN	PN	MV	PV
Tag. words	6309	1717	2424	684	1484
Tag. instances	15342	3560	6610	1307	3865

Table 2: Statistics of WSD annotations in *Składnica*.

WSD annotations in *Składnica* has been provided not only for polysemous words, but also for monosemous – in Table 2 the column *MN* contains statistics for monosemous nouns, *PN* for polysemous nouns, *MV* for monosemous verbs, *PV* polysemous verbs.

5.2 Results

5.2.1 Baseline PageRank approaches

As a baseline, we repeated experiments from (Kędzia et al., 2015) using plWordNet 2.2 as originally, but also version 2.3 as a basis for the WSD algorithm. All tests were performed on *KPWr* ; the results are shown in Table 3. The columns grouped under the label *PPR* include results achieved by the application of the *Personalized PageRank* algorithm, while the joint label *Static* signals the application of *Static PageRank*. The description of the tested combinations (algorithm parameters and the wordnet version) could make the table too large, so the combinations have been encoded as follows:

C1 the results achieved on plWordNet 2.2,

	PPR			**Static**		
	V	**N**	**All**	**V**	**N**	**All**
C1	28.64	47.25	40.45	28.14	43	37.57
C2	33.70	50.23	44.58	34.11	44.17	40.73
C3	29.57	48.06	37.57	29.79	42.79	38.05
C4	32.61	52.22	45.52	32.19	44.63	40.38

Table 3: Comparison of disambiguation precision using PLWN 2.2 and PLWN 2.3 evaluated on *KPWr*

	KPWr			***Składnica***		
	V	**N**	**All**	**V**	**N**	**All**
C5	34.11	44.17	40.73	47.08	57.37	53.37
C6	33.70	50.23	44.58	42.05	54.15	49.44
C7	32.19	44.63	40.38	47.00	57.97	53.70
C8	32.61	52.22	45.52	41.99	55.40	50.17

Table 4: Precision of disambiguation achieved on *KPWr* and *Składnica*.

C2 as above, but for plWordNet 2.3,

C3 and C4 the results achieved on plWordNet versions 2.2 and 2.3, respectively, merged with the SUMO ontology; in both only nodes belonging to plWordNet are initialised (i.e. receive non-zero values in the initial vector).

In Table 3 we can observe that the increasing size of plWordNet affects positively the precision when the same configuration of the algorithm is applied. This effect can be caused by the increasing number of text words covered by the wordnet that results in the increasing number of initially activated nodes in the PR graph. Moreover, in plWordNet 2.3 the number of adjectives and relation links between adjectives and nouns have been increased significantly. Thus cross-categorial connections have been improved, facilitating the activation flow in PR-based algorithms.

Next, we performed similar tests but using both data sets, i.e. *KPWr* and *Składnica*. Once again algorithms and parameters from (Kędzia et al., 2015) were applied, but this time we concentrated only on plWordNet 2.3. This resulted in better precision in the experiments presented above. Table 4 contains the results achieved for the following configuration of the algorithms:

C5 *Static* algorithm, only plWordNet 2.3 synset graph used,

C6 *PPR* algorithm, only plWordNet 2.3 synsets,

C7 *Static* algorithm, plWordNet 2.3 synset graph merged with *SUMO* ontology, but only nodes from plWordNet are initialised,

C8 *PPR* algorithm, as above, plWordNet 2.3 synset graph merged with *SUMO* ontology, but only nodes from plWordNet are initialised for disambiguation.

Results on *Składnica* are higher and close to the results obtained for English. The precision is

	KPWr			Składnica		
	V	N	All	V	N	All
Lesk	16.80	18.80	18.12	39.34	38.56	38.87

Table 5: Simple Lesk algorithm run on *KPWr* and *Składnica*

	KPWr			*Składnica*		
	V	N	All	V	N	All
C8	32.61	**52.22**	45.52	49.02	64.02	58.48
C9	**42.66**	47.91	**46.12**	**47.51**	**61.67**	**56.16**

Table 6: Static PageRank WSD algorithm based on the weighted plWordNet graph (**C9**) in comparison to the PPR algorithm.

clearly boosted by the monosemous words, while monosemous words are not annotated *KPWr*. However this influence is too small to be the only reason for the difference, e.g. in Tab. 6 in the case of *Składnica* only polysemous words were evaluated, i.e. for polysemous and monosemous words the precision of **C9** is: 69.08% for nouns, 53.86% for verbs and 63.46% for all. The higher precision on *Składnica* can be also caused by the different way of selecting words for WSD annotation. In *Składnica* they come from the running text and we can expect some bias towards most frequent LUs (word senses), while the authors of *KPWr* tried to cover in WSD annotation all LUs for the selected lemmas, so less frequent LUs received more occurrences than we could expect in a text sample. Tests on *KPWr* illustrate the ability of the algorithms to distinguish between all possible senses, while tests on *Składnica* are a better picture of average precision we can expect in practical applications (especially when monosemous words are included in the result).

5.2.2 Glosses and Examples

The results of the simple Lesk's algorithm based on plWordNet 2.3 run on both corpora are presented in Tab. 5, where the precision is given for verbs and nouns in percentage points. This algorithm can be treated as the second baselines. The results illustrate the amount of disambiguating information included in the textual descriptions of plWordNet. They are much lower than obtained by PageRank-based algorithms, that explore the rich structure of plWordNet relations

5.2.3 Structural Description

Tab. 6 presents a comparison of the best baseline configuration for *KPWr*, namely **C8** with the approach using the information about the relation types called **C9**. In **C9** *Static* algorithm based on plWordNet 2.3 was used, but synset relations were assigned weights equal to 0.7 and LU relations weights equal to 0.3. Moreover, the top-scoring LUs within the range of 10% from the best score (according to the WSD algorithm) are re-ranked according to their order (i.e. their identifiers) in the plWordNet database. The re-ranking is limited to those cases in which the values from WSD are very close and the differences can be insignificant.

On *KPWr*, the use of weighting gave improvement only for verbs. Verbs have a higher ratio of LU relations in comparison to synset relations than nouns, so this supports the intuition that synset relations provide more information for WSD. However, a more in-depth analysis of different weights for different relations is needed. Such an optimisation would need larger training-testing WSD data sets. The situation was completely different in tests on *Składnica* – here in all cases a significant improvement can be observed. It seems that the higher weights for synset relations and synonymy (the weight 1.0) favour the most frequent senses.

	KPWr			*Składnica*		
	V	N	All	V	N	All
C10	38.57	43.20	41.62	48.77	**61.74**	**56.69**
C11	39.76	39.30	39.46	**49.28**	61.12	56.51

Table 7: PageRank-based WSD algorithms supported by re-ranking based on the synset order in plWordNet.

5.2.4 Sense order

Finally, we tested the use of the order of adding LUs to plWordNet for a given lemma as an additional source of knowledge for WSD algorithms. In all cases this knowledge was used for post-re-ranking. Two configurations were tested:

C10 *Static* algorithm, plWordNet 2.3 synset graph only, WSD results post-processed by re-ranking of the top highest scored LUs within the range of $k = 30\%$ of the maximal score, the re-ranking is based on LUs numbers in plWordNet.

C11 Similar to **C10**, but re-ranking is limited to $k = 40\%$ of the maximal score.

The results obtained with the help of **C10** and **C11** are presented in Tab. 7. In comparison to the

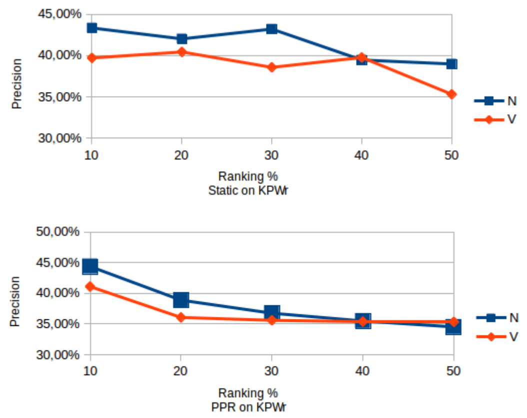

Figure 1: Influence of ranking % on precision evaluated on *KPWr* with Static and PPR.

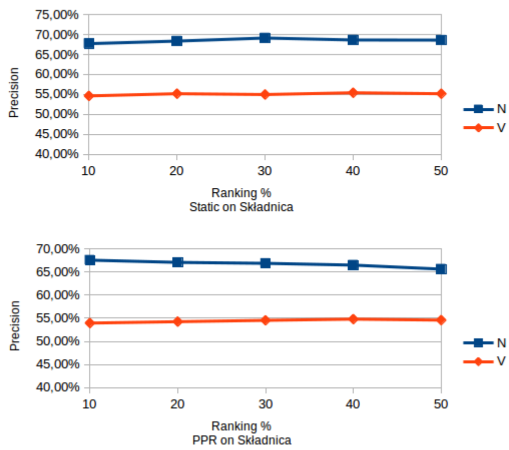

Figure 2: Influence of ranking % on precision evaluated on *Składnica* with Static and PPR.

baselines shown in Tab. 4, we can notice that re-ranking brought significant improvement in tests on *Składnica* for both configurations. The situation is different for *KPWr*. *KPWr* includes more occurrences of less frequent senses, while *Składnica* has a bias towards more frequent senses as built on randomly selected sentences. This difference supports our assumptions that LU numbers in plWordNet are correlated with their frequency in corpora. This correlation is next transferred to re-ranking. This observation is important for practical applications. Thus, we guess that the wordnet editors share some notion of the word sense saliency or their frequency. For a new lemma being edited, they seem to add to the plWordNet its more prominent and more frequent senses first. plWordNet 1.6 noun synsets were automatically ordered according to the estimated frequency of the word senses they represent (McCarthy et al., 2004, 2007). However, this method is of limited accuracy and all synsets added later (a large number, the majority) were not ordered in this way.

In Tab. 1 and 2 the analysis of the relation between the re-ranking threshold and precision is presented. In the case of *KPWr* the best results were obtained for the 10% re-ranking threshold. However, in the case of *Składnica* the highest results are concentrated around the threshold 30% and decrease beyond it, so scores produced by the WSD algorithm are at least useful in selecting the most likely LUs for a given word.

6 Conclusions

Weakly supervised WSD methods based on plWordNet have slightly lower precision in tests on Polish WSD corpora than similar PWN-based methods. However, plWordNet does not provide glosses for all LUs and the existing glosses are not disambiguated. Instead we looked into utilisation of other features. We showed that except glosses and examples, we can explore relation types by weighting them for the needs of WSD and the order in which LUs have been added to plWordNet. Both resulted in the increased precision of WSD on one of the test corpora – the one that seems to be closer to the practical applications. While the positive influence of the relations weights on PageRank-based WSD algorithm had been expected, the positive influence of the LUs adding order is a surprise, as the wordnet editors were not asked to use any specific order in introducing new LUs into plWordNet. Thus they have to share some idea of the salience or frequency of the individual LUs for the given lemma. This effect may not be visible when we analyse lists of LUs of individual lemmas, but it seems to be the most probable explanation for the results WSD algorithms using this order as a knowledge source. In future work we plan to develop more sophisticated system of weights assigned to relations for WSD and to work on combining different knowledge sources in one complex WSD algorithm.

Acknowledgment

Work supported by the Polish Ministry of Education and Science, Project CLARIN-PL, the European Innovative Economy Programme project POIG.01.01.02-14-013/09, and by the EU's 7FP under grant agreement No. 316097 [ENGINE].

References

Eneko Agirre and Aitor Soroa. Personalizing PageRank for Word Sense Disambiguation. In *Proceedings of the 12th Conference of the European Chapter of the Association for Computational Linguistics*, EACL '09, pages 33–41, Stroudsburg, PA, USA, 2009. Association for Computational Linguistics. URL `http://dl.acm.org/citation.cfm?id=1609067.1609070`.

Eneko Agirre, Oier Lopez de Lacalle, and Aitor Soroa. Random walks for Knowledge-Based Word Sense Disambiguation. *Computational Linguistics*, 40(1):57–84, 2014.

Dominik Baś, Bartosz Broda, and Maciej Piasecki. Towards Word Sense Disambiguation of Polish. In *Proceedings of the International Multiconference on Computer Science and Information Technology — 3rd International Symposium Advances in Artificial Intelligence and Applications (AAIA'08)*, pages 65–71, 2008. URL `http://www.proceedings2008.imcsit.org/pliks/162.pdf`.

Bartosz Broda, Michał Marcińczuk, Marek Maziarz, Adam Radziszewski, and Adam Wardyński. KPWr: Towards a free corpus of Polish. In *Proceedings of LREC'12*, Istanbul, Turkey, 2012. ELRA.

Christiane Fellbaum, editor. *WordNet: An Electronic Lexical Database (Language, Speech, and Communication)*. The MIT Press, May 1998. ISBN 026206197X.

Elżbieta Hajnicz. The procedure of lexico-semantic annotation of Składnica treebank. In Nicoletta Calzolari (Conference Chair), 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'14)*, Reykjavik, Iceland, may 2014a. European Language Resources Association (ELRA). ISBN 978-2-9517408-8-4.

Elżbieta Hajnicz. Lexico-semantic annotation of *składnica* treebank by means of PLWN lexical units. In *Proceedings of the 7th International WordNet Conference*, pages 23–31, 2014b.

Sanda M. Harabagiu, George A. Miller, and Dan I. Moldovan. WordNet 2 - a morphologically and semantically enhanced resource. In *SIGLEX99: Standardizing Lexical Resources*, pages 1–8, 1999. URL `http://www.aclweb.org/anthology/W99-0501`.

Paweł Kędzia and Maciej Piasecki. Ruled-based, interlingual motivated mapping of plWordNet onto SUMO ontology. In Nicoletta Calzolari (Conference Chair), 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'14)*, Reykjavik, Iceland, may 2014. European Language Resources Association (ELRA). ISBN 978-2-9517408-8-4.

Paweł Kędzia, Maciej Piasecki, Jan Kocoń, and Agnieszka Indyka-Piasecka. Distributionally extended network-based Word Sense Disambiguation in semantic clustering of Polish texts. *IERI Procedia*, 10(Complete):38–44, 2014. doi: 10.1016/j.ieri.2014.09.073.

Paweł Kędzia, Maciej Piasecki, and Marlena J. Orlińska. Word sense disambiguation based on large scale Polish CLARIN heterogeneous lexical resources. *Cognitive Studies*, 14(To appear), 2015.

Michael Lesk. Automatic sense disambiguation using machine readable dictionaries: how to tell a pine cone from an ice cream cone. In *Proceedings SIGDOC '86 Proceedings of the 5th annual international conference on Systems documentation*, pages 24–26. ACM, 1986.

Marek Maziarz, Maciej Piasecki, and Stanisław Szpakowicz. Approaching plWordNet 2.0. In *Proceedings of the 6th Global Wordnet Conference*, Matsue, Japan, January 2012.

Marek Maziarz, Maciej Piasecki, Ewa Rudnicka, and Stanisław Szpakowicz. Beyond the transfer-and-merge wordnet construction: plWordNet and a comparison with WordNet. In *Recent Advances in Natural Language Processing, RANLP 2013, 9-11 September, 2013, Hissar, Bulgaria*, pages 443–452, 2013a. URL `http://aclweb.org/anthology/R/R13/R13-1058.pdf`.

Marek Maziarz, Maciej Piasecki, and Stanisław Szpakowicz. The chicken-and-egg problem in wordnet design: Synonymy, synsets and constitutive relations. *Language Resources and Eval-*

uation, 47(3):769–796, 2013b. doi: 10.1007/s10579-012-9209-9.

Diana McCarthy, Rob Koeling, Julie Weeds, and John Carroll. Finding predominant word senses in untagged text. In *Proceedings of the 42Nd Annual Meeting on Association for Computational Linguistics*, ACL '04, Stroudsburg, PA, USA, 2004. Association for Computational Linguistics. doi: 10.3115/1218955.1218991. URL `http://dx.doi.org/10.3115/1218955.1218991`.

Diana McCarthy, Rob Koeling, Julie Weeds, and John Carroll. Unsupervised acquisition of predominant word senses. *Comput. Linguist.*, 33(4):553–590, December 2007. ISSN 0891-2017. doi: 10.1162/coli.2007.33.4.553. URL `http://dx.doi.org/10.1162/coli.2007.33.4.553`.

Rada Mihalcea, Paul Tarau, and Elizabeth Figa. PageRank on semantic networks, with application to Word Sense Disambiguation. In *Proceedings of the 20th International Conference on Computational Linguistics*, COLING '04, Stroudsburg, PA, USA, 2004. Association for Computational Linguistics. doi: 10.3115/1220355.1220517. URL `http://dx.doi.org/10.3115/1220355.1220517`.

George A. Miller, Claudia Leacock, Randee Tengi, and Ross T. Bunker. A semantic concordance. In *Proceedings of the Workshop on Human Language Technology*, HLT '93, pages 303–308, Stroudsburg, PA, USA, 1993. Association for Computational Linguistics. ISBN 1-55860-324-7. doi: 10.3115/1075671.1075742. URL `http://dx.doi.org/10.3115/1075671.1075742`.

Rafał Młodzki and Adam Przepiórkowski. The WSD development environment. In Zygmunt Vetulani, editor, *LTC*, volume 6562 of *Lecture Notes in Computer Science*, pages 224–233. Springer, 2009. ISBN 978-3-642-20094-6. URL `http://dblp.uni-trier.de/db/conf/ltconf/ltconf2009.html#MlodzkiP09`.

Lawrence Page, Sergey Brin, Rajeev Motwani, and Terry Winograd. The PageRank citation ranking: Bringing order to the Web, 1999.

Patrick A. Pantel. *Clustering by Committee*. PhD thesis, University of Alberta Edmonton, Alta., Canada, 2003.

Adam Pease. *Ontology: A Practical Guide*. 2011.

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

Maciej Piasecki, Stan Szpakowicz, Christiane Fellbaum, and Bolette Sandford Pedersen. Introduction to the special issue: On wordnets and relations. *Language Resources and Evaluation*, 47(3):757–767, 2013. ISSN 1574-020X. doi: 10.1007/s10579-013-9247-y. URL `http://dx.doi.org/10.1007/s10579-013-9247-y`.

Adam Przepiórkowski, Rafał Górski, Barbara Lewandowska-Tomaszczyk, and Marek Łaziński. Narodowy Korpus Języka Polskiego, 2009.

Mark Stevenson, Eneko Agirre, and Aitor Soroa. Exploiting domain information for Word Sense Disambiguation of medical documents. *JAMIA*, 19(2):235–240, 2012. URL `http://dblp.uni-trier.de/db/journals/jamia/jamia19.html#StevensonAS12`.