Automatic Semantic Filtering of Morphosemantic Relations in WordNet

Svetlozara Leseva, Ivelina Stoyanova, Borislav Rizov,
Maria Todorova, Ekaterina Tarpomanova
Institute for Bulgarian Language – Bulgarian Academy of Sciences
{zarka,iva,boby,maria,katja}@dcl.bas.bg

Abstract

In this paper we present a method for automatic assignment of morphosemantic relations between derivationally related verb–noun pairs of synsets in the Bulgarian WordNet (BulNet) and for semantic filtering of those relations. The filtering process relies on the meaning of noun suffixes and the semantic compatibility of verb and noun taxonomic classes. We use the taxonomic labels assigned to all the synsets in the Princeton WordNet (PWN) – one label per synset – which denote their general semantic class.

In the first iteration we employ the pairs *<noun suffix : noun label>* to filter out part of the relations. In the second iteration, which uses as input the output of the first one, we apply a stronger semantic filter. It makes use of the taxonomic labels of the noun-verb synset pairs observed for a given morphosemantic relation. In this way we manage to reliably filter out impossible or unlikely combinations. The results of the performed experiment may be applied to enrich BulNet with morphosemantic relations and new synsets semi-automatically, while facilitating the manual work and reducing its cost.

1. The Morphosemantic Relations in WordNet

Morphosemantic relations are a type of semantic relations which have morphological expression in at least one language (Koeva, 2008), for instance through derivational means. Since these relations link concepts, they are universal and transferable across languages, as has been demonstrated successfully in the context of different initiatives within the WordNet community (Bilgin et al., 2004; Pala and Hlaváčková, 2007; Koeva, 2008; Koeva et al., 2008; Fellbaum et al., 2009; Barbu Mititelu, 2012; Piasecki et al., 2012a; Piasecki et al., 2012b; Dimitrova et al., 2014). The typology and the specifics of a language determine whether the lexemes that lexicalise the respective concepts will be derivationally related. The morphosemantic relations we deal with are the morphosemantic links encoded between derivationally related literals in verb–noun pairs of synsets in the Princeton WordNet – PWN (Fellbaum et al., 2009).

Currently a relatively small portion of the derivationally related synsets in the PWN are supplied with a semantic label. In the PWN 3.0 version used in this paper there are 36,142 pairs of derivationally related verb–noun synsets, with at least one pair of derivationally related literals in each pair of synsets, while morphosemantic links have been assigned to 17,740 pairs of literals.

These relations have been mapped from the stand-off file distributed with the PWN[1] to the corresponding synsets in the Bulgarian WordNet (Koeva, 2010) using the cross-language relation of equivalence between synsets (Vossen, 2004).

13 out of the 14 types of morphosemantic relations encoded in the PWN denote a relation between a predicate and a participant in its semantic representation and hence correspond to thematic roles. Those are: Agent, By-means-of (corresponding to inanimate Agents or Causes but also to Means), Instrument,

[1] http://wordnetcode.princeton.edu/standoff-files/morphosemantic-links.xls

Material, Body-part, Uses (function of purpose) Vehicle (means of transportation), Location, Result, State, Undergoer, Destination, Property[2]. The only exception is the relation Event, which links a verb to a deverbal noun denoting the same event.

Derivationally related verb–noun pairs may be obtained through direct or non-direct derivational paths. Moreover, the direction of the derivation is usually not taken into account. In the pair *programiram* ("to program") – *programa* ("computer program"), which is assigned the morphosemantic relation Result, the verb is produced from the noun via direct derivation. However, the relation (Agent) between *programiram* and *programist* ("computer programmer") results from non-direct derivation, since both words are derived independently from the noun *programa*.

The derivation between a pair of literals may involve one or more derivational steps. For example, the place noun *kovachnitsa* ("forge, smithy") is produced from the verb *kova* ("forge, hammer") in two steps: first an agentive noun *kovach* ("(black)smith") is formed with the suffix *-ach*, and then the place suffix *-nitsa* is attached to the agentive noun base.

We identified the following cases of derivationally related pairs in Bulgarian that remain unconnected by means of a morphosemantic relation after the automatic transfer from the PWN. For a derivationally related pair of synsets in Bulgarian: (i) the corresponding synsets in the PWN may not be derivationally related, e.g. *kova* ("hammer") – *kovach* ("blacksmith"); (ii) the English corresponding noun and/or verb may be compounds, and therefore – unrelated – e.g. *chakam* ("wait") – *chakalnya* ("waiting room"); (iii) there may be a derivational relation in the PWN but it is not assigned a morphosemantic link, e.g. *izvarsha* ("perpetrate") – *izvarshitel* ("perpetrator").

Our goal is to discover derivationally related literals in verb–noun pairs of synsets in BulNet, such as the ones in (1-3), and to assign these pairs one or more morphosemantic relations using the semantics of the derivational means (focusing on suffixes). We assume that each morphosemantic relation corresponds to a distinct sense of a given suffix. Many suffixes express more than one morphosemantic relation. Usually, the knowledge about the semantics of the suffix is not sufficient alone to predict the morphosemantic relation unambiguously. We try to disambiguate fully or partially the possible morphosemantic relations for a given suffix by applying further semantic filtering. In this way we aim to facilitate the manual work on encoding and/or validating new instances of the morphosemantic relations. Once validated, they may be transferred to other languages.

The method uses a language-independent module – an inventory of morphosemantic relations obtained from the PWN automatically, and two language-dependent modules: (i) an inventory of suffixes and suffix variants; and (ii) mapping between suffix variants and suffix canonical forms. The former of the language-dependent modules is acquired automatically while the latter involves manual work. The method can be adapted relatively effortlessly to other sets of morphosemantic relations implemented in other wordnets, and to other languages. In the first case it would require an extension of the language-independent module (by transferring relations from another wordnet), and in the second case – an adaptation of the affix recognition algorithm and subsequent mapping to canonical forms.

2. Establishing Derivationally Related *Verb–Noun* Pairs and an Inventory of Affixes

After the assignment of the morphosemantic relations, an algorithm for recognising derivationally related pairs of verb–noun literals (Lv–Ln) was implemented (Dimitrova et al., 2014). The algorithm relies on string similarity and heuristic procedures. Similarity is established if at least one of the following conditions are met: (i) one of the literals is a substring of the other; (ii) the two literals have a common beginning (estimated to be at least half the length of the shorter literal); (iii) the two literals have a Levenshtein distance smaller than a given empirically determined value. This procedure resulted in linking 6,135 verb–noun pairs, each of which was validated manually.

2.1. Establishing an Inventory of Affixes

In order to establish the inventory of derivational patterns and the morphosemantic relations expressed by each of them, we extracted those 6,135 Lv–Ln pairs and identified the substrings which we assumed

[2] http://wordnetcode.princeton.edu/standoff-files/morphosemantic-links-README.txt

contained the affixes involved in the derivation. An expert linguist inspected the unique verb and noun beginnings and endings and associated each of them with a canonical form of the respective affix(es), suffixes in particular. This process was required because unlike prefixes, which usually concatenate with the stem, suffixes are realised by a number of morphophonemic variants due to the fact that their attachment to the stem may be accompanied by vowel and consonant changes in both the stem and the suffix. For example, the suffix *-nie* is also realised by the following variants: *-anie, -enie, -zhdenie, -zhenie, -zanie, -lenie, -sanie, -ovanie, -ovlenie, -shenie, -shlenie, -ovenie, -iyanie*. We identified 62 verb and 228 noun patterns of the type *canonical suffix > suffix variant*. Regular sound alternations resulting from assimilation and dissimilation processes, such as *k > ts, k > ch, g > h, sh > s*, etc. were also taken into consideration.

2.2. Suffix Normalisation in Detail

1. Given a pair of derivationally related literals Lv–Ln belonging to the synsets Sv and Sn, which are linked via a morphosemantic relation, we remove the vowels and find the longest similar substrings so that one of the substrings can be produced from the other. This is achieved using a dynamic programming algorithm. The vowel removal aims at reducing the phonetic alternations in the stems of the related forms caused by different linguistic phenomena, such as metathesis, e.g. *krav* ("blood") – *okarvavya* ("blood"); vowel mutation, e.g. *izbor* ("choice") – *izbera* ("choose"); elision, e.g. *bera* ("pick, pluck") – *brane* ("picking, plucking"), etc. In the examples the algorithm identifies the common strings – *krv*, *izbr* and *br* respectively. The common substring expanded by the vowels between the consonants is considered to be an approximation of the stem. The stem variants are generated by including/excluding the bordering vowels. For each stem, the remaining substring(s) that either precede(s) it (conditionally called a *prefix*), in the first example *o-*, or follow(s) it (conditionally called a *suffix*) – *-avya* in the same example), are established. They are subsequently checked against a list of prefix and suffix variants and the longest matches are selected.

2. We map the *suffix* substrings found in the noun literals to the list of canonical noun suffixes on the basis of the patterns *canonical suffix > suffix variant*, looking for the longest match. We are interested in noun suffixes since they express the morphosemantic relations under consideration, while verb suffixes have mainly a grammatical meaning. Finally, the results were post-edited manually. 83 canonical noun suffixes were established.

The normalisation of affixes helps in two ways. It allows us (i) to identify more reliably the morphosemantic relations expressed by each affix; and (ii) to reduce data sparsity that arises from the morphophonemic variants.

3. Establishing an Inventory of Pairs <*Affix : Relations*>

For the literals containing a given canonical noun suffix, we calculated the types of morphosemantic relations with which it is associated and the number of instances for each relation. Out of the 83 noun suffixes, 32 are unambiguous (one morphosemantic relation per suffix). The largest portion of the unambiguous suffixes denote the relation Agent (13 suffixes), followed by Event (7), and the remaining 12 are distributed among several relations – Material, Result, Undergoer, Property, State, Instrument. The rest 51 suffixes are ambiguous. The number of senses for all the noun suffixes is 252. Not all the predictions are accurate since some <*affix : relations*> pairs are attested in few instances or not attested at all.

The senses expressed by a suffix are not arbitrary but clustered around a given relation which is the preferred reading for this suffix. Table 1 shows that for the most productive suffixes that express the relation Agent the majority of instances are cases of default reading, and the rest of the relations are represented by much fewer examples.

The other senses of the suffixes given in Table 1 also have agentive properties since they denote inanimate agents and causes, such as Instrument, Material, By-means-of, Vehicle. Certain agentive suffixes can also express the relation Undergoer when the verb is unergative, e.g., *rabotnik* ("worker" – a person who works at a particular occupation).

	Agent	Instrument	Material	Undergoer	Vehicle	By-means-of	other
-tel	169	13	17	1	-	6	1(Event),1(Uses)
-(y)ach	128	2	-	2	2	1	-
-(n)ik	87	2	1	4	-	-	3(Event)
-sht	83	-	-	4	-	-	-
-tor	42	15	12	-	-	8	3(Result),1(Uses)

Table 1: Distribution of senses of the top 5 agentive suffixes

So even though many of the suffixes are ambiguous, at least for a part of them the ambiguity is very predictable. The examination of the data shows that the different senses of a given suffix are to a great extent taxonomically distinct. For instance, nouns with the suffix *-(n)ik* which are Agents, are persons, Instruments are artifacts, Materials are substances. This works also for untypical suffix senses. For instance, the suffixes *-ne, -stvo, -tsiya* may express the relation Agent due to a metaphorical extension of the meaning of some eventive deverbal nouns to denote Agents. Since persons cannot be Events, these suffixes may be disambiguated on the basis of the noun semantics alone; in these cases the semantic (taxonomic) class of the noun is a very strong indicator for the relation.

4. Semantic Filtering

In order to (partially) filter out the possible combinations <*suffix : morphosemantic relation*> we explore the possibility of using the taxonomic restrictions imposed by each suffix as a semantic filter.

The taxonomic distinctions between the different senses of the suffixes largely correspond to natural semantic classes, such as persons, artifacts, locations, acts, etc. Being a linguistic taxonomy, WordNet distinguishes these classes.

The PWN synsets are organised in 45 lexicographer files (26 – nouns, 15 – verbs, 4 – for the other parts of speech) based on the syntactic category and the taxonomic class of a synset[3]. Nouns denoting people are found in the file *noun.person*, nouns denoting feelings and emotions – in the file *noun.feeling*, etc. This allows us to use the file names as taxonomic labels for the noun and verb synsets.

Given that (i) there is an algorithm that recognises the suffix of a word and associates it with its canonical form, and (ii) the taxonomic label and the morphosemantic relations associated with a canonical suffix can be obtained from the synsets, we can use those labels to filter the morphosemantic relations associated with a given suffix.

1. From the already validated noun literals and the synsets to which they belong we extract the pairs <*suffix: morphosemantic relation*>. For example, for the suffix *-(n)ik*, the following morphosemantic relations are licensed:

   -(n)ik: agent, *-(n)ik: undergoer*, *-(n)ik: instrument*, *-(n)ik: material*, *-(n)ik: event*

2. Given the pair <*suffix: morphosemantic relation*>, we rule out the taxonomically incompatible morphosemantic relations, that is, those relations that have not been attested for the pair <*suffix : taxonomic label*> in BulNet and obtain triples of the type <*suffix : taxonomic label : morphosemantic relation*>. For example, after applying this semantic filter, for the suffix *-(n)ik* we acquire the following triples:

   <*-(n)ik : noun.person : Agent, Undergoer*>

   <*-(n)ik : noun.artifact : Instrument*>

   <*-(n)ik : noun.substance : Material*>

   <*-(n)ik : noun.act : Event*>

[3]http://wordnet.princeton.edu/wordnet/man/lexnames.5WN.html

The triples represent the linguistic generalisations for the semantic restrictions on the senses of the suffixes. Those predictions are based on and therefore limited by the observed instances.

Since the algorithm which discovers derivationally related verbs and nouns links all the pairs that may be mapped by it, two types of problems arise: (i) erroneously linked unrelated words due to coincidental string similarity, such as in *slon* ("elephant") and *podslonya* ("to shelter"); (ii) overgeneration due to the lack of semantic restrictions on the verbs: for example, the noun *zaemane* ("loan") is connected not only to the verb *zaema* ("to loan"), but also to homonyms, such as *zaema* ("to assume a pose").

The first issue requires further improvement of the recognition algorithm, which we leave for future research. Overgeneration can be at least partially resolved by introducing additional semantic filters. To this end, we decided to explore further the potential of the taxonomic labels. For each instance of a morphosemantic relation we retrieve the taxonomic labels of the respective noun and verb synsets and calculate the frequency of occurrence of each triple <*morphosemantic relation : verb.label : noun.label*> in the PWN. For example, <*Agent : verb.communication : noun.person*> has 411 instances, or 15.83% of the instances of Agent, followed by <*Agent : verb.social : noun.person*> – 337 instances, or 12.98%. Certain patterns such as <*Agent : verb.change : noun.plant*> have few occurrences (1 instance or 0.01% of the occurrences of Agent). The low frequency of a pattern may indicate a specific or semantically restricted relation, compare *author* ("be the author of") – *author* ("writer"), as opposed to *tense* ("become tense, nervous, or uneasy") – *tensor* ("any of several muscles that cause an attached structure to become tense or firm"). The first pair illustrates a typical Agent relation between a verb of creation and a noun person, and the second one exemplifies a more specialised Agent-like relation (Body-part), which involves verbs and nouns from semantically restricted classes (bodily functions, movements, etc. and a part of the body that performs them respectively). Low frequency may also indicate semantically dubious or unlikely relations, such as the Agent relation between *titter* ("laugh nervously") and *titter* ("a nervous restrained laugh") assigned in the PWN. Although we filter out the combinations with low frequency, we consider including rarely seen legitimate patterns manually at a later stage (see Section 7.).

In order to test the application of semantic patterns to the task of semantic filtering, we set up an experiment, which we describe in the following Section.

5. Experimental Method

The experiment consists in: (1) identifying derivational pairs in BulNet that have not been assigned a morphosemantic relation, and predicting the probable morphosemantic relations for each of the pairs on the basis of information about the suffix senses and taxonomic classes; and (2) filtering out a part of these relations using semantic criteria. The main purpose of the method is to facilitate the manual validation of automatically assigned morphosemantic relations. Manual inspection is nevertheless necessary in order to ensure high-quality data that can be used for training various linguistic models and applications.

1. **Identification of potential derivational pairs Ln–Lv in BulNet.** This step requires two distinct procedures: (i) recognition of derivational pairs, and (ii) identification of the canonical suffix of the noun literal in the pairs.
   (a) **Recognition:**
      i. Given a noun in BulNet – Ln, look up its ending in the list of morphophonemic variants of the noun suffixes.
      ii. If the ending is found in the list, remove it from the word.
      iii. If the remaining string is at least 4 characters long, attach to it a verb suffix from the list of the morphophonemic variants of verb suffixes.
      iv. If the resulting word is a legitimate verb in BulNet – Lv, find all the verb synsets in which Lv occurs.
   (b) **Mapping:**
      i. Given a pair Ln–Lv recognised at the previous stage, map the morphophonemic variant of the suffix of Ln to its canonical form. In this way we acquire all the instances of a given suffix, regardless of the morphophonemic environment.

   ii. For a given pair Ln–Lv, retrieve all the synsets Sn and Sv in which they are found.

2. **Semantic filtering.** The semantic filtering is performed in two steps. The output of Step 1 serves as input for Step 2.
   (a) **Step 1.**
      i. For each Ln, retrieve all the morphosemantic relations licensed by the combination of the suffix and the taxonomic label of the synset in which Ln is found by intersecting the possible pairs <*canonical suffix : noun.label*> with the possible pairs <*canonical suffix : morphosemantic relation*>.
      ii. Assign all the morphosemantic relations licensed by Ln to the pairs Sn–Sv, such that Sn contains Ln and Sv contains Lv.
   (b) **Step 2.**
      i. Given the frequency of occurrence of a triple <*morphosemantic relation : verb label : noun label*> in the PWN, estimate the probability of each triple in the output of Step 1.
      ii. For a given probability threshold, filter out the Ln–Lv pairs that are below the threshold. By varying the threshold we can obtain balance between precision and coverage in accordance with the particular purposes – a lower threshold means a larger number of assigned relations and more manual work on their validation but higher recall, and vice versa. We determine the threshold empirically on randomly selected samples of the data (see Section 6.).

6. Results

57,771 derivationally related literal pairs were identified at Step 1, out of which 7,601 pairs could not be assigned a morphosemantic relation because the particular semantic pattern <*canonical suffix : taxonomic label*> had not been observed previously in the manually validated literals. These pairs need to be examined systematically so that we can extend the already discovered combinations with new attested patterns, such as <*-iya : noun.body*>, which was found in the pair *anatomiya* (“anatomy, a human body”) – *anatomiziram* (“anatomize”). The remaining pairs were assigned one or more relations (up to 8) out of the 14 morphosemantic relations, which amounted to a total of 219,597 relations assigned.

At Step 2, in order to determine the threshold, we experimented with several values from 0.1 to 0.9 set apart by 0.1, by observing the proportion of assigned relations, on the one hand (Table 2), and by evaluating the precision and recall on random samples, on the other. The samples included (i) 100 automatically assigned relations, and (ii) 100 discarded relations for each threshold value[4] (Figure 1). Each threshold is evaluated using $F_{0.5}$ measure, where precision is given twice as much weight as recall, although results were consistent for other F_β measures for $0 < \beta < 1$. The highest $F_{0.5}$ measure of 0.882 was achieved for a threshold of 0.7.

The performance of the method with the selected threshold of 0.7 was evaluated using the following set of criteria:

- Efficiency – it was evaluated in terms of the reduction in the number of assigned morphosemantic relations as a measure of the feasibility of further manual validation. The total number of 219,597 relations was reduced to 26,766 (12.19% of the total). Moreover, the number of highly ambiguous cases of initial relation assignment was markedly decreased by an average factor of 6.76. As a result, the manual validation of the semantic filtering is rendered much more tractable.

- Precision and recall – the precision and recall were estimated based on a different set of samples of 100 assigned relations (above the threshold) and 100 discarded relations (below the threshold), using the formulae:

$$\text{Precision} = \frac{\text{correctly assigned}}{\text{all assigned}}, \qquad \text{Recall} = \frac{\text{correctly assigned}}{\text{correctly assigned} + \text{incorrectly discarded}}$$

[4]Assuming that all the possible morphosemantic relations were identified in advance, precision can be calculated as the percentage of the correctly assigned relations out of all the assigned relations, and recall – as the percentage of all the correctly assigned relations out of all the correct relations (assigned and discarded).

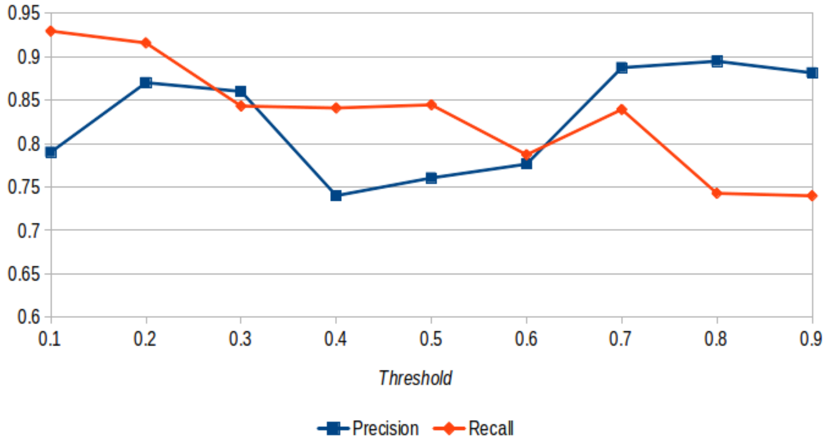

Figure 1: Precision and recall for various thresholds

Threshold	Assigned,#	Assigned,%
0	219,597	100.00
0.1	77,987	35.51
0.2	54,409	24.78
0.3	46,673	21.25
0.4	40,495	18.44
0.5	34,338	15.64
0.6	30,376	13.83
0.7	26,766	12.19
0.8	26,195	11.93
0.9	24,461	11.14

Table 2: Decrease in the number of relations using various thresholds

With a precision of 0.90, the results are promising and justify the application of the method for semi-automatic expansion of WordNet with morphosemantic relations. The relations that were filtered out were predominantly invalid, resulting in recall of 0.84. This result leads us to the conclusion that the implemented semantic filtering largely preserves the coverage of the morphosemantic relations.

7. Discussion

In order to improve the method, we performed a preliminary error analysis, focusing on 2,000 pairs of literals that had been assigned a single morphosemantic relation at Step 1. Three types of errors were identified: (i) the pair of words is wrongly recognised due to coincidence of symbol strings; (ii) the words in a pair are derivationally related but none of the defined morphosemantic relations is appropriate; (iii) the words have a derivational relation but the assigned morphosemantic label is wrong. With respect to the third type of errors we draw two directions for further improvement.

1. **Enriching the semantic description of suffixes.** In the collection of automatically assigned morphosemantic relations we observed valid suffix senses unattested in the synsets related through morphosemantic relations in the PWN. For instance, the suffixes *-er/-ier/-ur* and *-in* had not been attested with the meaning of Undergoer and when such cases were discovered in the automatically assigned pairs: *pensioner* ("pensioner, retired person") – *pensioniram* ("superannuate, retire") and *grazhdanin* ("citizen") – *pograzhdanyavam* ("urbanise"), the nouns were incorrectly recognised as Agents. The systematic exploration of falsely assigned relations will make it possible for us to draw a full description of the semantics of the suffixes and to make more precise predictions.

2. **Enriching the semantic restrictions imposed by the taxonomic labels.** In analysing the pairs

that were falsely assigned a morphosemantic relation, we observed new semantic restrictions. For instance, the agentive suffix *-ach/-yach* had been found in the synset {povdigach:1; levator:1} ("a muscle that serves to lift some body part") which has the taxonomic label *noun.body*. We looked up for other synsets with the same taxonomic tag that also contain a noun with the suffix *-ach/ -yach* but were assigned a morphosemantic relation. We found such an example – {obtegach:1; tensor:1} ("any of several muscles that cause an attached structure to become tense or firm") – which was assigned the relation Body-part. As a result, we acquired the following generalisation *<-ach/ -yach : noun.body : Body-part>*. Although this pattern has a low frequency, it shows very distinct semantic properties so it can be safely included in the list of semantic restrictions. Respectively, the suffix senses also need to be updated. This line of research is directed towards increasing the coverage of morphosemantic relations.

8. Related Work

The task of recognising and/or generating derivatives from existing words in a wordnet is explicitly or implicitly directed towards the expansion of a wordnet with new synsets and relations, and/or the transfer of those synsets and relations to other wordnets (Bilgin et al., 2004; Pala and Hlaváčková, 2007; Koeva, 2008; Koeva et al., 2008; Piasecki et al., 2012a; Stoyanova et al., 2013).

We focus on the task of assigning new instances of the morphosemantic relations and proposing an algorithm for (partial) disambiguation by means of semantic filters. In a similar vein, Piasecki et al. (2012a) use a bigger inventory of relations which include the morphosemantic relations in the PWN to the end of training a tool to discover derivational pairs of words and to suggest derived words missing in the Polish WordNet. The authors discuss the possibility of using semantic information obtained from WordNet, such as upper-level hypernyms and semantic domains, to filter erroneous pairs. Piasecki et al. (2012b) propose a method for semantic classification of verb–noun derivational relations using supervised machine learning. Their approach uses context features of the derivationally related pairs observed in a huge corpus to disambiguate the derivational relations, whereas our method employs semantic patterns observed in the Princeton WordNet. Our proposal is closest in spirit to the work of Stoyanova et al. (2013), who suggest filtering morphosemantic relations assigned automatically to derivationally related pairs of synsets by means of a semantic filter based on the taxonomic labels in WordNet. The results of their experiment have not been reported in detail. Drawing on their idea, we further expand on and test the hypothesis that together with the semantics of suffixes verb–noun taxonomic labels are a reliable semantic filter for morphosemantic relations of the type discussed herein.

9. Future Directions

The methodology reported in this paper gives promising results. Future work will be focused on exploring the possibilities of mutually disambiguating the suffixes of words from the same synset on the basis of their senses and the semantic restrictions imposed by them both in a monolingual and in a multilingual setting. As suggested in the previous Section, the analysis of the errors and the cases where no relation is assigned will be further employed to identify and collect new semantic restrictions imposed by suffixes and possibly new suffix senses. The application of additional semantic filters, such as upper-level hypernyms, will also be explored. Another line of research that is worth investigating is the application of the method to enriching WordNet with new synsets on the basis of morphosemantic relations.

Acknowledgements

The present paper was prepared within the project *Integrating New Practices and Knowledge in Undergraduate and Graduate Courses in Computational Linguistics* (BG051PO001-3.3.06-0022) implemented with the financial support of the Human Resources Development Operational Programme 2007-2013 co-financed by the European Social Fund of the European Union. The authors take full responsibility for the content of the present paper.

References

Barbu Mititelu, V. (2012). Adding Morpho-semantic Relations to the Romanian Wordnet. In *Proceedings of the Eight International Conference on Language Resources and Evaluation (LREC 2012)*, pages 2596–2601.

Bilgin, O., Cetinoglu, O., and Oflazer, K. (2004). Morphosemantic Relations in and across Wordnets – A Study Based on Turkish. In *Proceedings of the Second Global Wordnet Conference (GWC 2004)*, pages 60–66.

Dimitrova, T., Tarpomanova, E., and Rizov, B. (2014). Coping with Derivation in the Bulgarian WordNet. In *Proceedings of the Seventh Global Wordnet Conference (GWC 2014)*, pages 109–117.

Fellbaum, C., Osherson, A., and Clark, P. (2009). Putting Semantics into WordNet's "Morphosemantic" Links. In *Proceedings of the Third Language and Technology Conference, Poznan, Poland. [Reprinted in: Responding to Information Society Challenges: New Advances in Human Language Technologies. Springer Lecture Notes in Informatics]*, volume 5603, pages 350–358.

Koeva, S., Krstev, C., and Vitas, D. (2008). Morpho-semantic Relations in Wordnet – A Case Study for two Slavic Languages. In *Proceedings of the Fourth Global WordNet Conference (GWC 2008)*, pages 239–254.

Koeva, S. (2008). Derivational and Morphosemantic Relations in Bulgarian Wordnet. *Intelligent Information Systems*, pages 359–368.

Koeva, S. (2010). Bulgarian Wordnet - Current State, Applications and Prospects. In *Bulgarian-American Dialogues*, pages 120–132. Sofia: Prof. M. Drinov Academic Publishing House.

Pala, K. and Hlaváčková, D. (2007). Derivational Relations in Czech Wordnet. In *Proceedings of the Workshop on Balto-Slavonic Natural Language Processing*, pages 75–81.

Piasecki, M., Ramocki, R., and Maziarz, M. (2012a). Automated Generation of Derivative Relations in the Wordnet Expansion Perspective. In *Proceedings of the 6th Global Wordnet Conference (GWC 2012)*, pages 273–280.

Piasecki, M., Ramocki, R., and Minda, P. (2012b). Corpus-based Semantic Filtering in Discovering Derivational Relations. In Ramsay, A. and Agre, G. (eds.), *Applications – 15th International Conference, AIMSA 2012, Varna, Bulgaria, September 12-15, 2012. Proceedings. LNCS 7557*, pages 14–22. Springer.

Stoyanova, I., Koeva, S., and Leseva, S. (2013). WordNet-based Cross-language Identification of Semantic Relations. In *Proceedings of the 4th Biennal International Workshop on Balto-Slavonic Natural Language Processing*, pages 119–128.

Vossen, P. (2004). EuroWordNet: A Multilingual Database of Autonomous and Language-Specific Wordnets Connected via an Inter-Lingual Index. *International Journal of Lexicography*, 17(3):161–173.