Handling Synset Overgeneration: Sense Merging in BTB-WN

Zara Kancheva
IICT
Bulgarian Academy of Science
zara@bultreebank.org

Ivaylo Radev
IICT
Bulgarian Academy of Science
radev@bultreebank.org

Abstract

The paper reports on an effort to reconsider the representation of some cases of derivational paradigm patterns in Bulgarian. The new treatment implemented within BulTreeBank-WordNet (BTB-WN), a wordnet for Bulgarian, is the grouping together of related words that have a common main meaning in the same synset while the nuances in sense are to be encoded within the synset as a modification functions over the main meaning. In this way, we can solve the following challenges: (1) to avoid the influence of English Wordnet (EWN) synset distinctions over Bulgarian that was a result from the translation of some of the synsets from Core WordNet; (2) to represent the common meaning of such derivation patterns just once and to improve the management of BTB-WN, and (3) to encode idiosyncratic usages locally to the corresponding synsets instead of introducing new semantic relations.

1 Introduction

Since the first release of Princeton WordNet (PWN) (Fellbaum, 1998) — at the beginning of 1990 — it has become one of the most popular and used language resources in the field of NLP. It models lexical knowledge as a lexico-semantic network, containing groups of word senses with associated set of synonyms, called synsets, which are related via lexical and semantic relations between them. The need for machine-readable dictionaries made the wordnets one of the most valuable resources for lexical and semantic knowledge for tackling NLP tasks such as word sense disambiguation (WSD), relation extraction, named entity recognition and linking, multi-word expression (MWE) handling, machine translation, etc. PWN also became a model for similar semantic lexicons for many other languages. Currently, the Global WordNet Association hosts information for wordnets for more than 50 languages.[1]

These subsequent wordnets are usually created by two approaches called the *merge model* and the *expand model* (Vossen, 1998). In the *merge model* the monolingual wordnet for the given language is created from scratch — building the set of synsets for the language and interlinking them via set of semantic and lexical relations, without any influence from any previously existing wordnet (for English or other language). If necessary this monolingual wordnet is later mapped to the PWN. Examples for wordnets constructed according to this approach (among others) are GermaNet (Hamp and Feldweg, 1997), Polish WordNet (Derwojedowa et al., 2008), Norwegian WordNet NorNet (Fjeld and Nygaard, 2009) and Danish WordNet DanNet (Pedersen et al., 2009).

In the *expand model* a new wordnet for the given language is created by transferring lexical knowledge from PWN via translation of synsets, their glosses and semantic relations in a manual or semi-automated way. The expand approach rests on the assumption of the universality of the human mental lexicon and the accuracy of the specific language structure of the internal relation network may be arguable when transferred to another language. Examples of wordnets created in this way are MultiWordNet (Bentivogli and Pianta, 2004), AsianWordNet (Robkop et al., 2009), IndoWordNet (Sinha et al., 2006), Open Dutch WordNet (Postma et al., 2016), sloWNet for Slovenian (Fišer and Sagot, 2015), FinnWordNet (Lindén and Carlson, 2010) and the French WordNet WOLF (Sagot and Fišer, 2008).

In both cases if the mapping between a wordnet in a given language to PWN has to be supported, it suffers from two major problems: (1) the conceptual difference between languages — having differ-

---

[1] http://globalwordnet.org/resources/wordnets-in-the-world/

ent ways of concept lexicalization in the different languages; and (2) the granularity of the senses encoded within the wordnets. The usual way to cope with the first problem is to prepare a rich inventory of interlingual semantic relations. The second is more problematic. In this paper we present a linguistic approach to restructure Bulgarian BulTreeBank WordNet (BTB-WN) synsets in order to escape from the high granularity of senses in PWN. The granularity problem of PWN was reported in (Mihalcea and Moldovan, 2001). They discover that some word senses are semantically so close or too fine-grained that even humans have difficulties to make a distinction. The paper proposes a solution to this problem by collapsing similar synsets and removing rarely used senses. This encoding of too fine-grained distinctions was considered to be detrimental for tasks such as WSD and even led to attempts to reduce the granularity of the WordNet sense inventory (Navigli, 2006). In this paper we propose a linguistically motivated way to join some related senses in BTB-WN that are representationally separated.

As reported in (Osenova and Simov, 2018) Bulgarian BulTreeBank WordNet (BTB-WN) has been created by a combination of the merge and expand methods in three different stages: (1) by manual translation of English synsets from the Core WordNet subset of PWN [2] into Bulgarian. This step ensures comparable coverage between the two wordnets on the most frequent senses; (2) by identification of senses used in the Bulgarian Treebank BulTreeBank (BTB). Then these identified senses have been put into newly created synsets for the BulTreeBank WordNet and at the same time being mapped onto the conceptual structure of PWN. In this way, the BTB-WN was extended with real usages of the word meanings in texts. Also, the coverage of the core and base concepts for Princeton WordNet has been evaluated over a Bulgarian syntactic corpus; (3) by sense extension, which includes two tasks: a) detection of the missing senses of processed lemmas in BulTreeBank and adding them to the BTB-WN, and b) a semi-automatic extraction of information from the Bulgarian Wiktionary and later from the Bulgarian Wikipedia mapped to synsets from PWN and then manually checked. In all these steps relation between synsets in BTB-WN and PWN have been maintained. There are

several reasons to do this, but the two main reasons are: the transfer of information from PWN to BTB-WN such as lexical and semantic relations and other; and the support of multilingual applications of BTB-WN.

The same paper shows the use of several types of correspondence used to ensure that the differences between the two languages are preserved during the mapping between BTB-WN and PWN. The full correspondence is observed when the Bulgarian and English concepts match perfectly. The next type of correspondence is the partial kind which can be divided into two subtypes depending on the direction. The first one is observed when the Bulgarian concept is more specific than the English one. In this case the Bulgarian term is mapped to a more general English. The second one is observed when the English concept is more specific. In this case the Bulgarian concept is mapped as hypernym of the English.

Currently BTB-WN is under extension and revision. We are doing reconsideration of the granularity of Bulgarian synsets in order to represent Bulgarian conceptualization of lexical knowledge in a better way and also to establish more appropriate mappings to PWN while reducing the impact that PWN has over the Bulgarian. Here we are reporting on exploiting derivational morphology in order to establish better definition of Bulgarian synsets.

The paper is structured as follows: in section 2 some related works discussing similar difficulties in the mapping of wordnets are shown; section 3 presents the strategy for sense merging and some typical cases; and section 4 concludes the paper.

2 Related Work

In this section we present some previous works (beside the ones sketched above in the introduction) on mappings between different wordnets and PWN. Our goal is to stress the main problems related to the maintenance of mapping between wordnets. When comparing two monolingual wordnets the major challenges are connected to language and cultural idiosyncrasies that affect the granularity of the represented synsets.

The most prominent example of supporting mappings between languages with different conceptualizations (sometimes based on differences in the grammatical structure) is the mapping between Polish Wordnet (PlWN) and Princeton WordNet. In

---

[2] The Core WordNet contains the 5000 most frequent synsets of PWN. http://wordnetcode.princeton.edu/standoff-files/core-wordnet.txt

(Rudnicka et al., 2021) the authors explore the problems comparing the two wordnets. The first challenge is related to the different approaches towards the construction of PlWN and PWN. The Polish synset is rigidly defined as a set of lexical units sharing a set of constitutive relations which makes the PlWN more fine-grained than PWN, where synset members are more arbitrary. The second challenge lies in the substantial cross-linguistic differences between English and Polish which also reflect on the relation structures of the two wordnets. We identified similar issues in our own work on BTB-WN with respect to cross-linguistic differences between English and Polish. This became obvious when we translated the CoreWordNet into Bulgarian. It was necessary, for example, to relate some Bulgarian adjectives to English nouns, or to provide a Bulgarian syntactic phrase to an English lexicalized verb in these languages, etc.

We differ from the Polish approach to creation of a wordnet in the way of handling the granularity. We would like to minimize the number of semantic and lexical relations introduced by the different kinds of synsets. We believe that such relations become idiosyncratic and do not reflect the richness of sense usages. This motivates the work reported here.

To deal with these differences new types of lexical unit relations were added to PlWN, such as diminutivity or cross-categorial synonymy (Noun-Adj for example). The effect on relation structures was that they are partially or wholly different leading to the usage of different types of interlingual relations going far beyond interlingual synonymy defined as simple equivalence by (Vossen, 2002) plus the addition of some new ones.

Another example for merging two wordnets is the case of the Danish WordNet — DanNet compared to Princeton WordNet in (Pedersen et al., 2019). The authors present the linking between core concepts in the direction English to Danish language, where the majority of links are direct links in terms of manually evaluated exact synonym relations. In the opposite direction when Danish is the starting point for the linking process it becomes more complex. In these cases the human annotators had to look for a Danish hypernym, and search for candidates for mapping among the related PWN hyponyms or search directly for a potential PWN hyponym. Regarding the challenges caused by different sense granularities in the two

lexical resources, the Danish lexicographers responsible for the mapping of the core concepts reached the conclusion that the sense inventory of PWN is more fine-grained than the one of DanNet. The directionality problem applies also to BTB-WN when mapped to PWN, but it is outside the scope of this paper.

Similar problems are found also in the mapping of the Romanian WordNet (ROWN) to Princeton WordNet (Cristea et al., 2004). The paper reports on the difficulties encountered during the alignment of synsets between English and Romanian. The authors define the reasons for these difficulties as the inconsistencies and the difference in criteria for senses recording found both in PWN and ROWN and the inherent differences in the lexicalization of concepts in the two languages. The Romanian wordnet uses a mixed merge-expand model where the Romanian senses were collected from different sources and linked to PWN synsets and some Romanian synsets were modified to match the linguistic concepts of PWN. The two major challenges reported in this work are: (1) the difference in the granularity between the second language (Romanian in this case) and the English language, and (2) the conceptual gaps in one or the other language.

The paper also puts forward a rarely discussed problem — that mapping a non-English wordnet to PWN is often done by non-native English speakers, who encounter difficulties in correctly perceiving the English senses especially in the fine-grained structure of the PWN. Similar observations were made also by us. Thus, further development in mapping might require involving native English speakers.

The granularity problem appears in the case of the Turkish WordNet — KeNet (Bakay et al., 2021). It is built as a monolingual resource on the base of the latest contemporary dictionary of Turkish with its semantic relations suggested by PWN, but manually checked and only the correct relations added to the KeNet. Before the addition of sense relations the Turkish synsets underwent a merge and a split processes. In the merge process "different synsets that should be grouped together were identified and grouped as a single synset". During that process three important rules were established: (1) each synset has to have one single and unique definition, (2) only true synonyms are valid synset members, and (3) a representative lemma for the whole synset must be a first member in each synset. In the split

156

process wrongly grouped senses were identified and separated into new synsets. This allowed the creators of KeNet to control the synset members and identify which synsets had missing members or had extra members.

We use a similar approach to reconsider the structure of BTB-WN and its mapping to PWN. The main difference is that we try to be consistent with respect to whole sets of synsets in Bulgarian motivated by linguistic knowledge about related senses as it will be clear in the next section.

3 Linguistically Motivated Sense Merging

Currently BTB-WN is being developed in two directions: (1) refinement of the legacy synsets from the previous versions by editing synset definitions, synset members (lemmas) and sense relations, and (2) extending the word coverage of the BTB-WN with lemmas from a Bulgarian explanatory dictionary — (Popov, 1994) containing about 56,000 lemmas grouped in 46,000 entries. We consider this number of lemmas and entries as a medium-size lexicon of Bulgarian. In its current state the BTB-WN consists of nearly 23,400 synsets and 50,000 lemmas. For comparison at the start of this process of refinement and extension the BTB-WN had nearly 19,000 synsets and 38,300 lemmas. The current result shows that our strategy for reducing the granularity of senses is working well and the number of lemmas is growing faster than the number of synsets.

During the manual work on refinement the experts met the above reported problem with the granularity of the English WordNet (EWN), where every nuance in meaning is reflected, resulting in many similar synsets. For example the synset for "over and over again" with definition *repeatedly* and no relations and the synset for "repeatedly" with definition *several times* and one pertainym relation or the case with the lemma "marquess" where the EWN makes a distinction between a *nobleman (in various countries) ranking above a count* and *a British peer ranking below a duke and above an earl*, but the Bulgarian concept consider it just as *a nobleman in various countries with this title*. A decision had to be made to completely follow the structure of EWN thus making the BTB-WN an English language-dependent wordnet which would make it very difficult to map it with wordnets in other languages or to go in the direction of monolin-

gual wordnets and take the approach of consolidation of the hierarchy, and therefore — unification of very similar meanings, supported by detailed definitions and diversified examples. Making the BTB-WN English-dependant does not correspond with the usage and the goals BTB-WN needs to achieve. This led to the only possible solution — diverging from the expand model of creation and going closer to the merge model.

The following subsections will outline several cases in which sense merging is the preferred choice instead of creating several very similar in meaning synsets. As we stressed earlier we use derivational patterns in Bulgarian for this purpose.

3.1 Diminutives

Diminutives are a specific linguistic category considered by scholars either as distinct lexical units or as forms of other, 'general' form (Krastev, 1976), because a diminutive form could be derived from almost all parts of speech and additionally it denotes the same object as the non-diminutive forms, but with some difference in the appearance or with a particular attitude towards this object. (Dimitrova, 1959) considers diminutives as an intermediate category between separate words and forms of general words, (Zidarova, 2004) outlines their unclear status. Anyway, there is an aspect of the diminutives that is not controversial — that their meaning is not limited only to the traditional smallness, but they can also express positive or negative attitude. Similar contradictions of the nature of diminutives in English are observed for example in (Jurafsky, 1996), (Bagasheva, 2018), (Schneider, 2013). Depending on the notion of diminutives, there can be observed several approaches of their presentation in dictionaries.

(Avramova, 2016) compiles three approaches for dictionary articles of diminutives: structural definition, semantic definition and structural-semantic definition. The structural definition says that the given term is a diminutive for the general term. It could be considered the poorest in terms of information given. The semantic definition is used mostly for diminutives for people and animals and it describes the diminutive as a young person/animal. The third type — structural-semantic — provides information that the term is diminutive, but also presents all of its possible meanings — it could express youngness, smallness, weakness, positive or negative attitude.

In the current work on BTB-WN we will take into consideration only the diminutives for nouns; such forms for adjectives could be part of future work. The approach towards the diminutives' definition will follow the idea of the structural-semantic definitions in the dictionaries with one difference — the diminutive forms of the nouns will be in one synset with the general form, but will have a special lemma marker. If we presume that a diminutive noun usually has around three meanings (for something small, for expressing affection or for expressing ridicule towards somebody or something), we can imagine how many synsets a wordnet could generate only from this category of words. The lemma markers that we will use would contain information about the different meanings of the given diminutive, but will prevent the wordnet to have great number of very similar synsets. Analogous markers will be used to present dialectical, archaic, slang and vulgar meanings on the level of lemmas. The same approach could be applied for the augmentative meanings.

A good example for the variety of meanings that a diminutive can bear is the noun даскалче "daskalche" (teacher-diminutive) — it could be interpreted as a young teacher, a rather short or thin teacher, it could express negative attitude towards the teacher, but it can also be a expression of affection.

3.2 Nouns vs. Human Gender

One of the productive derivational patterns in language are the masculine and feminine lemmas of nouns for professions, nationalities, etc. Similar is the case for some animals. Usually wordnets encode this difference in the lemmas as separate synsets. At the beginning in BTB-WN we have started to encode it in the same way. Later we discovered that in many cases this distinction is quite insignificant and not really important for the differences in the meaning.

With respect to the mapping to PWN the differences between Bulgarian and English became obvious. Only a few English nouns have a separate lemma for feminine. As a result many Bulgarian synsets did not have corresponding English synset to be mapped to. A decision was made to always unite the forms in one synset. Occupations like "actor" and "actress", "waiter" and "waitress" are divided in separate synsets, but many similar are combined, for example the synsets for "physician"

or "teacher".

Another problem with this type of synsets is that when the synsets for male and female terms are separated, the female form is hyponym of the male, which is not the most precise relation, and when they are both put in one synset in BTB-WN it is very difficult to find the appropriate position in the hierarchy.

Other examples of that problem are the terms for nationality — in English often there are no distinct nouns denoting men, women and children (only one genderless form), but in Bulgarian such derivative nouns are productive, thus an effort was made to include them in one synset — българин, българка, българче "balgarin, balgarka, balgarche" (Bulgarian man, Bulgarian woman, Bulgarian child).

In order to solve the problem in Bulgarian and to encode the information we decided to add such lemmas to the same synset sharing the main meaning in the synset and the gender differences to be encoded with respect to the different lemmas in the synset.

There is a similar case with the English synsets for animals. Often the domestic animals are divided into two synsets (or sometimes more — including the young of this animal) — one for the meaning of female animal, and one for the male (like "hen" and "cock", "male horse" and "mare", "billy" and "nanny", "cow" and "bull"), but for wild animals there is usually one synset (wolf) in comparison to the Bulgarian where we have вълк and вълчица "valk and valchica" (male wolf and female wolf). In this cases we encode all the lemmas for the same animal in the same synset.

3.3 Variation in Verb Prefixation

Bulgarian verbal paradigm contains a great amount of verbs with prefixes that bear the semantics of start, end, duration, repetition, etc. of the action. (Todor Boyadzhiev and Penchev, 1998) state that "this type of prefix verbs do not denote new action, but the same action of their derivative verb — they only specify, fix this action in space, time and in quantitative terms". Synsets with such verbs do not have equal English synsets, so the decision is to map them to the synset for the general meaning of the verb and to use lemma markers which contain the specific semantic features of the distinct meanings. There is a variety of prefixes that compose such verbs, their semantics are thoroughly analysed

in (Ivanova, 1974) (Velkova, 2019), but for the purposes of this research only the most common senses of the prefixes are outlined:

- въз- "vaz-" (renewing the action; bringing the action to result; producing effect; bringing the action to a higher or upper level)

- до- "do-" (bringing the action to result, to completion)

- за- "za-" (start of action; bringing to action; bringing to a new state; performing the action continuously and to a high degree)

- из- "iz-" (bringing to result, to completion; bringing something to a new state; moving the action outward)

- на- "na-" (bringing to result; bringing to new state; performing the action to a larger extent and with accumulation)

- над- "nad-" (bringing to result; performing the action beyond, over the area of the object; bringing the action to a higher level or better state (than someone/something else))

- по- "po-" (performing the action to a lesser degree; bringing to a new state; start of action; performing the action shortly and intermittently)

- под- "pod-" (develop the indicated action under the object or in its lower part; bringing the action to result; performing the action secretly, covertly; performing the action in a rather weak, slight, soft manner)

- пре- "pre-" (renewal or repetition of the action, sometimes in new way or with the inclusion with one more action; performing the action to a greater extent;)

- пред- "pred-" (performing the action before other action, preceding something)

- при- "pri-" (bringing something closer to some place or object; start of action; bringing something to result; renewal of the action; lower intensity of the action; parallel performance of two actions)

- про- "pro-" (bringing to a new state; performing the action shortly and intermittently; start of action)

- о- "o-" (bringing to result; spreading the action on the entire surface of the object)

- об- "ob-" (bringing to result; spreading the action on the entire surface of the object)

- раз- "raz-" (performing the action to a significant extent; bringing to result; setting the action in multiple directions; dividing something in parts by the action)

- с- "s-" (bringing to result; bringing to new state; performing the action suddenly and with great force)

- съ- "sa-" (performing the action by reducing (destroying) the distance between two or more objects within its range; parallel performance of two actions)

- у- "u-" (directing the action in depth or to the center of the object; directing to result)

- в- "v-" (bringing to a new state; bringing to result)

- въ- "va-" (bringing to a new state; bringing to result)

- от- "ot-" (taking out subject or the object from the definition area of the action by performing the action; end of the action)

Needless to say that not every verb has derivatives with all these prefixes, but still they are very productive.

For example the verbs пиша "pisha" (with meaning 'to write'), написвам "napisvam", напиша "napisha" ('to write (something) completely, till the end'), изписвам "izpisvam", изпиша "izpisha" ('to completely fill something with writing, text), записвам "zapisvam", запиша "zapisha" ('start to write'), дописвам "dopisvam", допиша "dopisha" ('write (something) to the end, write everything that is necessary') and прописвам "propisvam", пропиша "propisha" ('learn how and start to write') will be united in one synset; the verbs with prefixes will have lemma markers that further describe their meaning.

There is also a rather big group of poliprefixed verbs (with two or more prefixes — for example изпонапрепиша, изпонапреписвам "izponaprepisha", "izponaprepisvam" ('to completely write something down from a source')) in Bulgarian, but poliprefixation will be part further research.

The three cases of derivational relations presented above show the approach we took in BTB-WN — joining in the same synsets many of the derivatives that do not alternate on a large scale the main meaning of the lemma from which all the rest are derived. The borderline between modified sense and separate sense is a challenge on its own and we implement manual checks in several Bulgarian dictionaries to make this distinction. Then the differences are encoded as small modifications of the main meaning — modifying some characteristics of this meaning like the gender of the referent, the size, other features, etc. With this approach we have succeeded to reduce the number of synsets. Another benefit of this approach is that given meaning is represented just in one synset and any modification and the addition of new knowledge are in one place. In this way it becomes manageable in a better way.

4 Conclusion and Future Work

The paper presents an attempt at dealing with sense merging in the process of mapping two wordnets with a focus on the treatment of synsets for diminutives, verb forms with different ways of action and female and male forms of nouns for professions, roles, nationalities and animals. The work reported here is based on the notion that making the coverage of BTB-WN based on number of lemmas (synset members) instead on number of synsets will be beneficial when applying wordnet data to the tasks like WSD. As a secondary result of this work several suggestions for edits in the English WordNet were made.

In the future we hope to achieve even larger coverage of lemmas by including dialectal, archaic, slang, diminutive and vulgar words as new synsets or new synset members. One task still remaining is the ordering of the synset members relying on the notion that the first member has to be the exemplar.

We have reported some works on the mapping mainly of wordnets for languages from Indo-European language families to English PWN. It would be interesting to extend the study for mapping between different languages within the family (for example Bulgarian to Polish or Bulgarian to German) and also between wordnets for languages from different language families. We are aware of the even greater problems with the mapping between Asian wordnets and the English one.

References

Tsvetanka Avramova. 2016. Deminutivite kato rechnikovi edinitsi (Diminutives as dictionary units). pages 89–97, Sofia, Bulgaria. Prof. Marin Drinov Academic Publishing House.

Mariya Bagasheva. 2018. Ways of expressing the category of diminutiveness in english. volume 56, pages 460–469. Plovdivski universitet "Paisii Hilendarski (University of Plovdiv "Paisii Hilendarski").

Özge Bakay, Özlem Ergelen, Elif Sarmış, Selin Yıldırım, Bilge Nas Arıcan, Atilla Kocabalcıoğlu, Merve Özçelik, Ezgi Sanıyar, Oğuzhan Kuyrukçu, Begüm Avar, and Olcay Taner Yıldız. 2021. Turkish WordNet KeNet. In *Proceedings of the 11th Global Wordnet Conference*.

Luisa Bentivogli and Emanuele Pianta. 2004. Extending wordnet with syntagmatic information. In *Proceedings of second global WordNet conference*, pages 47–53.

D. Cristea, Cătălin Mihăilă, C. Forascu, Diana Trandabat, Maria Husarciuc, Gabriela Haja, and Oana Postolache. 2004. Mapping Princeton WordNet synsets onto Romanian wordnet synsets.

Magdalena Derwojedowa, Maciej Piasecki, Stan Szpakowicz, Magdalena Zawisławska, and Bartosz Broda. 2008. Words, concepts and relations in the construction of Polish WordNet. pages 162–177.

Milka Dimitrova. 1959. Umalitelnite imena v knizhovniya ezik (The Diminutives in the Literary Bulgarian Language). volume 6, page 263–319, Sofia, Bulgaria. Akademichno izdatelstvo "Prof. Marin Drinov" ("Prof. Marin Drinov" Academic Publishing House).

Christiane Fellbaum. 1998. *WordNet: An Electronic Lexical Database*. MIT Press.

Darja Fišer and Benoît Sagot. 2015. Constructing a poor man's wordnet in a resource-rich world. *Language Resources and Evaluation*, 49(3):601–635.

R. Fjeld and Lars Nygaard. 2009. Nornet — a monolingual wordnet of modern Norwegian. In *In NODAL-IDA 2009 workshop: WordNets and other Lexical Semantic Resources - between Lexical Semantics, Lexicography, Terminology and Formal Ontologies*, pages 13–16.

Birgit Hamp and Helmut Feldweg. 1997. GermaNet - a lexical-semantic net for German. In *Automatic Information Extraction and Building of Lexical Semantic Resources for NLP Applications*.

Kalina Ivanova. 1974. *Nachini na glagolnoto deystvie v savremenniya balgarski ezik (Ways of verb action in contemporary Bulgarian language)*. Akademichno izdatelstvo "Prof. Marin Drinov" (Prof. Marin Drinov Academic Publishing House), Sofia, Bulgaria.

Daniel Jurafsky. 1996. Universal tendencies in the semantics of the diminutive. *Language*, 72(3):533–578.

Borimir Krastev. 1976. Umalitelnostta v balgarskiya ezik (The diminutiveness in Bulgarian language). Sofia, Bulgaria. Narodna prosveta.

Krister Lindén and Lauri Carlson. 2010. FinnWordNet–Finnish WordNet by translation. *LexicoNordica–Nordic Journal of Lexicography*, 17:119–140.

Rada Mihalcea and Dan I Moldovan. 2001. Automatic generation of a coarse grained wordnet.

Roberto Navigli. 2006. Reducing the granularity of a computational lexicon via an automatic mapping to a coarse-grained sense inventory. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)*, Genoa, Italy. European Language Resources Association (ELRA).

Petya Osenova and Kiril Simov. 2018. The Data-driven Bulgarian WordNet: BTBWN. *Cognitive Studies — Études cognitives*, 18(1713).

Bolette S. Pedersen, Sanni Nimb, Ida Rørmann Olsen, and Sussi Olsen. 2019. Merging DanNet with Princeton Word Net. In *GWC*.

Bolette Sandford Pedersen, Sanni Nimb, Jørg Asmussen, Nicolai Hartvig Sørensen, Lars Trap-Jensen, and Henrik Lorentzen. 2009. DanNet: the challenge of compiling a wordnet for Danish by reusing a monolingual dictionary. *Language resources and evaluation*, 43(3):269–299.

Dimitar Popov. 1994. *Balgarski talkoven rechnik. Chetvarto izdanie, preraboteno i dopalneno (Bulgarian Explanatory Dictionary. Forth revised and extended edition)*. Nauka i izkustvo (Science and art), Sofia, Bulgaria.

Marten Postma, Emiel van Miltenburg, Roxane Segers, Anneleen Schoen, and Piek Vossen. 2016. Open Dutch WordNet. In *Proceedings of the 8th Global WordNet Conference (GWC)*, pages 302–310.

Kergrit Robkop, Sareewan Thoongsup, T. Charoenporn, V. Sornlertlamvanich, and H. Isahara. 2009. Wnms : Connecting the distributed wordnet in the case of Asian WordNet.

Ewa Rudnicka, Wojciech Witkowski, and Maciej Piasecki. 2021. A (non)-perfect match: Mapping plWordNet onto PrincetonWordNet. In *Proceedings of the 11th Global Wordnet Conference*, pages 137–146, University of South Africa (UNISA). Global Wordnet Association.

Benoît Sagot and Darja Fišer. 2008. Building a free French wordnet from multilingual resources. In *OntoLex*, Marrakech, Morocco.

Klaus P. Schneider. 2013. The truth about diminutives, and how we can find it: Some theoretical and methodological considerations.

Manish Sinha, Mahesh Reddy, and Pushpak Bhattacharyya. 2006. An approach towards construction and application of multilingual indo-wordnet. In *3rd Global Wordnet Conference (GWC 06), Jeju Island, Korea*. Citeseer.

Ivan Kutsarov Todor Boyadzhiev and Yordan Penchev. 1998. Savremenen balgarski ezik (Contemporary Bulgarian language).

Yordanka Velkova. 2019. Predtsavkite kato markeri na determinatsiya (Prefixes as markers of determination). *Materiali ot Mezhdunarodnata konfenetsiya "Balgaristichni ezikovedski cheteniya (Proceedings of The International Conference "Bulgarian Linguistic Readings"), 19–20.11.2018*, pages 209–217.

Piek Vossen. 1998. EuroWordNet: A multilingual database with lexical semantic networks.

Piek Vossen. 2002. WordNet, EuroWordNet and Global WordNet. *Revue Française de Linguistique Appliquée*, 7.

Vanya Zidarova. 2004. Deminutivi i leksikalna nominatsiya (Diminutives and lexical nomination). pages 114–122, Sofia, Bulgaria. Universitetsko izdatelstvo "St. Kliment Ohridski" (University press "St. Kliment Ohridski").