Challenges behind the data-driven Bulgarian WordNet (BulTreeBank Bulgarian WordNet)

Petya Osenova and Kiril Simov

Institute of Information and Communication Technologies, BAS,
Akad. G. Bonchev. 25A, 1113 Sofia, Bulgaria
{petya,kivs}@bultreebank.org

Abstract. The paper presents our work towards the simultaneous creation of a data-driven WordNet for Bulgarian and a manually annotated treebank with semantic information. Such an approach requires synchronization of the word senses in both - syntactic and lexical resources, without limiting the WordNet senses to the corpus or vice versa. Our strategy focuses on the identification of senses used in BulTreeBank, but the missing senses of a lemma also have been covered through exploration of bigger corpora. The identified senses have been organized in synsets for the Bulgarian WordNet. Then they have been aligned to the Princeton WordNet synsets. Various types of mappings are considered between both resources in a cross-lingual aspect and with respect to ensuring maximum connectivity and potential for incorporating the language specific concepts. The mapping between the two WordNets (English and Bulgarian) is a basis for applications such as machine translation and multilingual information retrieval.

1 Introduction

There have been two prominent trends in language resources creation — compiling syntactically annotated resources (treebanks), on the one hand, and building lexical resources (WordNets), on the other. The former resources reflect the syntagmatic connectedness of the words, while the latter encode primarily the paradigmatic relations among words (via hierarchies). There are also works focused on the semantic annotation of corpora/treebanks, which apply the lexical knowledge onto real texts. Here we report on the challenges behind the construction of the BulTreeBank Bulgarian WordNet (BTBWN). BTBWN has been created in three different ways: (1) by manual translation of English synsets from Core WordNet subset of Princeton WordNet (PWN — [2])[1] into Bulgarian. This step ensures comparable coverage between the two WordNets on the most frequent senses; (2) by identification of senses used in BTB. The identified senses have been organized in synsets for the BulTreeBank Bulgarian WordNet. The newly created Bulgarian synsets are mapped onto the conceptual structure of

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

PWN. In this way, the BTBWN was extended with real usages of the words 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 activities: a) detection of the missing senses of processed lemmas in BulTreeBank and adding them to the BTBWN, and b) a semi-automatic extraction of information from the Bulgarian Wiktionary mapped to synsets from PWN and then manually checked[2] In this paper we present the second step of creating the BTBWN — simultaneous annotation of BTB with senses, the extension of BTBWN with these new synsets and their mapping to PWN.

The structure of the papers is as follows. Section 2 briefly discusses related work. The construction of BTBWN is presented in Section 3. Section 4 introduces the general principles of mapping. Section 5 presents two extensions of BTBWN in progress and future direction of developments. The last section concludes the paper.

2 Related Work

Concerning WordNets, many of them for the European languages have been created within EuroWordNet and BalkaNet projects (including BulNet for Bulgarian). However, also some of these WordNets are not publicly available (including BulNet). This motivated us to start our own WordNet creation endeavor, since we needed the lexico-semantic information in our work on Machine Translation, eLearning and Word Sense Disambiguation.

There are two main methods for building a WordNet as pointed out in [8]: the *expand* method and the *merge* method. The former relies on the translation of the synsets from the source into the target language, thus complying initially to the source hierarchy of concepts. The latter takes (also) into account the language specific resources. Different WordNet projects used the above mentioned methods alone or in combination with other strategies. For example, translation of English PWN into another language; data-driven approaches via identification of synsets within real texts; automatic extraction from existing lexical resources; various combinations of these. In our WordNet project we exploit all of these approaches at different stages of the resource development. They will be explained in more detail in the next sections.

Let us introduce briefly some best practices in the WordNet creation for specific languages. Most of them seem to go for the expand approach first. Some WordNets were created on the basis of publicly available resources. For example, the Open Dutch WordNet[3] (see [7]) was created by "removing the proprietary content from Cornetto[4], and by using open source resources to replace this proprietary content." For the Basque language [5] the construction approach relies on the joint development of WordNets and annotated corpora. The Basque

[2] We would like to thank Antoni Oliver Gonzalez who provided the automatic mapping from Bulgarian Wiktionary to PWN.

[3] `http://wordpress.let.vupr.nl/odwn/`

[4] `http://www2.let.vu.nl/oz/cltl/cornetto`

WordNet was developed within the EuroWordNet framework. First, a quick core Basque WordNet was developed through semi-automatic methods. The quality control included a concept-to-concept manual review. Afterwards, an additional word-to-word review was performed as a higher-level quality check. The Slovenian WordNet started as automatic translation from a closely-related language resource, namely — the Serbian WordNet with the help of a bilingual dictionary. Later on, manual correction has been performed [1]. The Croatian WordNet [8] also used the expand method, but it additionally explored monolingual dictionaries for incorporating language-specific relations into the resource. One of the few endevours for constructing a language-specific WordNet first, and then mapping it to some already existing one, such as the Princeton WordNet, is the Polish WordNet [9].

To sum up, there is no easy way to achieve typological consistency in building WordNets - if the expand method is chosen, the language resource suffers from lack of nativeness of the hierarchy and relations. If the merge method is followed, the language resource differs too much from other similar resources and it is time-consuming to map it back to them.

Now let us turn to the accompanying sense corpora. The usual way of annotating senses in treebanks is the following: there is a WordNet for the language in question, and then the treebank is annotated with senses from it. This is the case in the German Tuba/DZ treebank, the Italian treebank and the Polish treebank [3]. All of them use the WordNets they created in EuroWordNet Project for sense annotation of the treebanks. Thus, they bear also the restrictions that are presented in the so-called static lexical resources. This means the following: if we want to annotate our texts with senses, but some sense is missing in the lexical database, and we cannot control the WordNet resource to add it, then the sparseness of the sense coverage would be really problematic.

Our work differs from the above mentioned approaches in the fact that we first annotated the treebank with senses from an explanatory dictionary of Bulgarian [6] and then started the formation of synsets. They were then mapped to the PWN while keeping track of the various sense discrepancies by different mappings. We explain our motivation for such a decision below in a more contextually-bound manner. Here it can be only mentioned that in this way a wider sense coverage was achieved quickly for the purposes of Machine Translation, since our initial WordNet covered only the core concepts from PWN.

3 Construction of BTBWN

In this section we present the steps of construction of BTBWN also from a historical point of view. The creation of this resource started as an attempt to construct domain vocabularies for two domain ontologies: the domain of *Information Technology for End Users*, and the domain of *Home Textile* — see [11] and [10]. In both cases the domain ontologies were aligned to an upper ontology for the reasons of consistency and inheritance of general knowledge. The ontology and the aligned lexicons were used for several tasks: (1) semantic annotation of

domain documents; (2) multilingual search; (3) common conceptualization; and (4) interaction with the end users. Thus, the lexicon interrelated the concepts in the ontology to the lexical knowledge used by the grammar in order to recognize the realizations of the concepts in the text; and the lexicon represented the main interface between the user and the ontology. In order to achieve these goals the need of general lexica became apparent. Thus our next goal was to extend the domain lexicons to cover (at least) the most frequent senses in Bulgarian. We could not find any evaluation on the distribution of word senses in Bulgarian. Thus we decided to solve this problem in two steps: (1) by transferring the most frequent senses from another language to Bulgarian, assuming that European languages share substantial number of most frequent senses; and (2) by annotation of Bulgarian texts where we believed that the most frequent senses would be present. For the purposes of applications, such as word sense disambiguation, annotated texts were needed. So we decided to annotate the senses for all open class words in the texts.

Concerning the first step — transfer of most frequent senses from another language — we translated manually the English synsets from the Core WordNet subset of the Princeton WordNet into Bulgarian. The translation was done by two people with excellent knowledge of English. First, they formulated a Bulgarian definition reflecting the content of the concept represented by its correspondence to the English synset. Then they formed the Bulgarian synset recording the Bulgarian lemmas that have this meaning. Some of the lemmas might be multiword expressions. After this first phase a lexicographer checked both - the definition and the lemmas. The result from this work was published as part of the Open Multilingual WordNet[5] under CC BY 3.0 license[6].

Our next step for extending the BTBWN was the manual annotation of running Bulgarian texts. Here our goals were: (1) to extend the coverage of BTBWN to really frequent Bulgarian words; (2) to have a corpus of semantically annotated texts which to be used for experiments with tasks like Word Sense Disambiguation; and (3) to check how many of the English most frequent senses are frequent also in Bulgarian. The actual annotation of the treebank was done in the following way: (1) for each lemma of the open class word forms in the treebank a concordance was created; (2) each lemma in the concordance was annotated with all possible senses from the Core WordNet version of BTBWN as well as from an explanatory dictionary of Bulgarian; (3) the annotators selected the appropriate sense for each example, if available. If there was no appropriate sense, or there was no available senses for a given lemma, the annotator had the possibility to create a new sense (definition). After the completion of this initial annotation the result was turned into lexical entries which contain the lemmas, selected in the text, the chosen definitions and the examples.

The following step was to manually map each new lexical entry to an appropriate synset in PWN. Thus we achieved several goals: (1) different lemmas with similar senses were grouped together and in this way the lexical entries

[5] http://compling.hss.ntu.edu.sg/omw/

[6] https://creativecommons.org/licenses/by/3.0/

for synonyms were recorded in the corresponding synsets; (2) the mapping to PWN allowed the execution of various bilingual applications; (3) mapping to WordNets of other languages. The annotation was checked by a second person and validated by a judicator. After the completion of the annotation, BTBWN contained about 11000 synsets. From them about 1800 synsets are from the Core WordNet version of BTBWN. In this way we empirically showed that the most frequent senses in the texts of BulTreeBank correspond roughly to one third of the English Core WordNet.

The next extension of BTBWN was performed by a semi-automatic addition from Bulgarian Wiktionary mapped to synsets from PWN and then manually checked. Behind this extension we added new senses for the words that have been already included in synsets of BTBWN. The idea is that each word is represented with all its senses.

The extensions on the basis of text annotation and the existing lexicons exhibit however the sparseness problem: not all synonyms appear in the annotated texts and the lexical entries. For that reason, we performed checks on the completeness of the synsets with respect to the missing synonyms. The checks have been performed with respect to the available monolingual synonymic dictionaries of Bulgarian. Special attention was paid to the aspect variation of verbs. In many of the synsets it turned out that for one of the verbs in the aspect pair there were only few real examples or no examples at all in the data. Thus, we started searching for examples from bigger corpora or on the web. Our goal is to have at least five examples for each synset. Ideally, examples are expected to be included for each lemma in the synsets.

4 Types of Mapping

The synchronization of word senses in BTBWN and the word senses in PWN is complicated by the fact that many senses in BTBWN originate from BulTreeBank, where the words reflect Bulgarian lexicalization of concepts which differ from PWN, which provides the English-specific view on the lexical relations.

From the annotator perspective, the mapping of the word from the text starts with its translation into English and is followed by a search through the corresponding lemmas in the PWN. Factors of importance for the adequate mapping are the following: the Bulgarian definition and the matching examples from the treebank. The provision of examples plays a crucial rule for the specification of the correct definition as well as the English description and accompanying examples. The PWN examples themselves can also help in indicating the matching concept, since the Bulgarian definition and the English one can be phrased in different ways and might reflect various granularity of conceptualizations.

Several types of correspondences have been attested during the mapping process: full correspondence (one-to-one); partial correspondence (one-to-many or many-to-one); forced connectivity (re-design of Bulgarian definition); common general meaning; resolving metonymies; incorrect and extended correspondences. Needless to say, these are not novel at all. However, they are still valuable, be-

cause they provide feedback for the typologically-based and the resource-oriented similarities and differences between Bulgarian and English, thus opening the path to comparisons with other languages as well.

4.1 Full Correspondence

The ideal case in the mapping is when equal concepts are encountered, i.e. the concepts in the two languages map one-to-one. That is, the Bulgarian concept matches the one in the Princeton WordNet. For example, the Bulgarian "сигурност", "sigurnost" and English "safety" both mean in short 'lack of danger'. If a Bulgarian definition corresponds equally well to more than one definition in the Princeton WordNet, then all these definitions are mapped to the Bulgarian one, using a special separator. For example, English "answer" and "response" map to Bulgarian "отговор", "otgovor".

4.2 Partial Correspondence

In many cases, however, the concepts differ in terms of specificity in both language directions. In the first case, the Bulgarian definition is more specific than the English one. In this case, it is mapped to a more general English one, but it is also marked with a specificity label. The most frequent cases here are the following ones: (i) regular polysemy — for example, in Bulgarian "прокуратура", "prokuratura", is given also as the building, while in English it is the institution, the group of people and the act; (ii) restrictive in Bulgarian vs. general in English definitions — for example, "дирекция", "direkcia", in the meaning of "director's office" in Bulgarian is mapped to the more general concept "office" in English with the meaning of "place of business".

A second scenario is possible, where the Bulgarian definition is more general and subsumes one or more synsets from PWN. In this case, the following approach has been taken — the common definition in Bulgarian is mapped once to the more specific English definitions (with relation *specificity*) and a second time to their hypernyms (with relation *subsumption*). For instance, in Bulgarian "режисьор", "rezhisyor", "director" has only one definition: *The lead person in the making of a theater play, film, TV program, etc.* However, in PWN there exist two synsets that can be related to it: *director as someone who supervises the actors and directs the action in the production of a show* (with a hypernym "supervisor" as one who supervises or has charge and direction of) and *director as the person who directs the making of a film* (with a hypernym "film maker" as a producer of motion pictures). In order to preserve both — the more abstract concept in Bulgarian as well as the hierarchical structure of PWN — the Bulgarian definition is mapped to all four ones of these synsets — with relation *specificity* to the specific ones, and with relation *subsumption* to their hypernyms. These mappings are presented in Fig. 1. Some more explanations are presented below.

Ensuring a One-to-One Mapping. In some cases of mismatch the one-to-one mapping can be achieved through re-working the Bulgarian definitions.

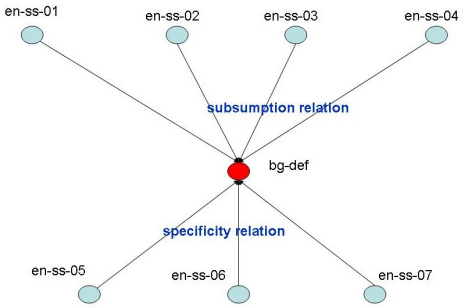

Fig. 1. Classification of a Bulgarian definition with respect to English synsets in Princeton WordNet hierarchy. We use the relation *subsumption* to map Bulgarian concept (definition) to more general synsets in Princeton WordNet, and the relation specificity to map it to more specific English synsets.

This often means dividing the Bulgarian definition into two separate ones. For example, the word "седмица", "sedmica", "week", has the following definition: seven consecutive days, usually counted from Monday to Sunday. All examples correspond to this definition. There are two synsets in English: "week" as any period of seven consecutive days, and "week" as a period of seven consecutive days starting on Sunday. Such a division in nouns referring to the passing of time has been done in Bulgarian for the concept of "month". Thus it can be implemented for the "week" as well. Since the Bulgarian definition has been mapped to the second synset in English, it can remain as it is, while a second definition is introduced (Seven consecutive days), which is mapped to the first synset; the examples are correspondingly divided between the two definitions.

Searching for a More General Meaning. There is another group of examples to which no equivalent sense can be detected. In this case the strategy is to find a more general one. Usually this applies to the cases of regular polysemy:

1. Types of institutions, buildings, people. Often there is no node in PWN corresponding to a given institution. Thus a mapping is made to the general node for an institution, company, establishment, etc.
2. Words like "'цар", "car", "king", etc. in Bulgarian refer to both concepts — a person and a title. In PWN there is a definition only for a person, therefore there is no word sense corresponding to the title meaning, as in "The title of the Bulgarian and Russian monarchs." Therefore this Bulgarian definition is mapped to the more general concept for title. On the other hand, a second definition in Bulgarian is added to reflect persons.
3. The Bulgarian concepts that are more specific than the English ones are treated in a special way, too. For instance, "чичо", "chicho", and "вуйчо", "vujcho", mean different things in Bulgarian. The former is *brother of a person's mother* or *husband to a sister of the father or the mother*. The latter is *brother of a person's father*. In this case the relations are mapped to the more general "uncle" concept by specificity relation.

As it was mentioned, since it is important to preserve access to the PWN hierarchy, it is necessary to align the concepts by introducing a non-lexicalized definition in the Bulgarian lexicon, namely "Brother of a person's mother or father", which corresponds to the English one. We should note again that this step is not done at the cost of losing language specific concepts. In this way a conditional connection is established, which will be made more complex in the Bulgarian concept hierarchy, because the relevant definitions of "чичо", "chicho", and "вуйчо", "vujcho", will denote subcategories of the newly created one.

Metonymic Usages. If a word is used with its metonymic sense, then the metonymic sense is mapped to the appropriate sense in the PWN. For example, in Bulgarian the word "армеец", "armeec", can be used in two senses: literal and rare (soldier), and metonymic and more frequent (member of a specific football team). When used in the latter sense, it must not be mapped to the concept of soldier, but to the concept of footballer. In this way, the specific features of the figurative language are kept in the lexicon.

Extended Mapping. With regard to Bulgarian, derived nouns with a special suffix and an ending "-ка", "-ka", mark the feminine gender and denote female persons. Thus, an addition is needed of a definition indicating the reference to a female. For example, "баскетболист", "basketbolist", is a basketball player, but "баскетболистка", "basketbolistka", is a a female basketball player. Accordingly, the definition remains mapped to the more general synset in English (provided that no equivalent-level definition is available). The two definitions in Bulgarian are related in a one-to-one manner, without indicating that one is a subclass of the other.

5 Current and Future Developments

Here we discuss briefly: the treatment of MultiWord Expressions (MWEs) as specific cases of lexicalization; and extensions of the current version with new words. We also present some directions of future developments.

5.1 Treatment of MultiWord Expressions in BTBWN

Currently, we include MWEs as strings of several words separated by spaces. They are represented in their standard form: lemmatized (where possible) and reflecting the canonical word order. However, in this representation we lose information about possible word order variations of the MWE elements and their potential for morphosyntactic variation and modification. In order to add this information we rely on the notion of *catena*.

The notion of catena (chain) was introduced in [4] as a mechanism for representing the syntactic structure of idioms. He shows that for this task a definition of syntactic patterns is needed that does not coincide with constituents. He defines the catena in the following way: *The words A, B, and C (order irrelevant) form a chain if and only if A immediately dominates B and C, or if and only if A immediately dominates B and B immediately dominates C.* In

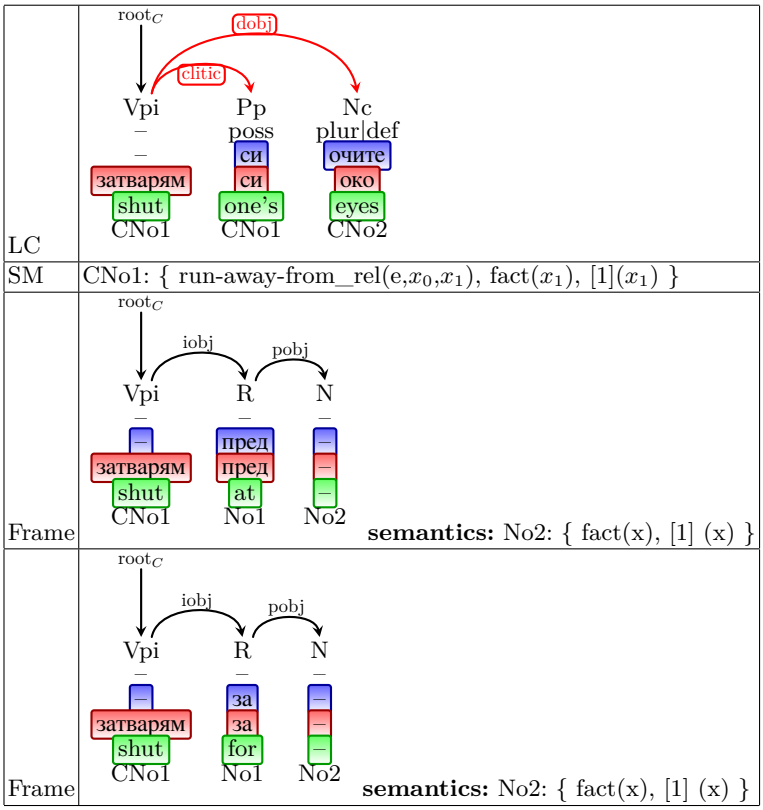

Fig. 2. Lexical entry for затварям си очите, "zatvaryam si ochite", 'I close my eyes'.

our work on BTBWN we convert MWEs into a representation defined in [12] and [13] in which the catena is depicted as a dependency tree fragment with appropriate grammatical and semantic information. Here we demonstrate the model by just one lexical entry for the Bulgarian MWE: затварям си очите, "zatvaryam si ochite", "I close my eyes". The lexical entry uses the following format: a **lexicon-catena**, **semantics** (SM) and **valency** (Frame). The lexicon-catena for the MWEs is stored in its canonical form. The realization of the catena in a sentence has to obey the rules of the grammar. In this way the possible word order is managed. The semantics of a lexical entry specifies the list of elementary predicates contributed by the lexical item. When the MWE allows for some modification (including adjunction) of its elements, i.e. modifiers of a noun, the lexical entry in the lexicon needs to specify the role of these modifiers. For example, the MWE represented in Fig. 2 'затварям си очите'.[7] The valency frame contains two alternative elements for indirect object introduced by two

[7] The grammatical features are: 'poss' for possessive pronoun, 'plur' for plural number and 'def' for definite noun.

different prepositions. The situation that the two descriptions are alternatives follows from the fact that the verb has no more than one indirect object. If there is also a direct object then the valency set will contain elements for it as well. The semantic contribution of the indirect object is specified for each valency element. This semantic contribution is added to the semantic contribution of the lexical entry when the valency element is realized. In the dependency tree fragments also grammatical features and lemmas are represented. The catena for the frame and for the whole lexical entry are unified on the basis of nodes with the same names. In this case CNo1. Within BTBWN the semantic contribution will depend also on the corresponding synsets to which the MWEs belong.

5.2 Extensions of BTBWN

The extension of the BTBWN coverage is a constant task. The selection of new lexical entries, including new synonyms, new senses for words that are already in synsets in BTBWN, new words with corresponding new senses and synsets is an on-going activity. To perform this task we use the following approaches: (1) a task-based approach; (2) a dictionary-based approach; (3) a corpus-based approach; and (4) a linguistically-based approach. All of these approaches are used by different language groups for the construction of the corresponding wordnets. We briefly present each of them.

The *task-based approach* concerns with the coverage of BTBWN for a concrete application. For example, the construction of lexicons for domain ontologies. In this case we identify concepts for the task to be performed and construct an aligned lexicon. The synsets in the domain lexicon are also aligned to the rest of BTBWN. The identification of the domain concepts frequently is based on the annotation of appropriate domain texts.

The *dictionary-based approach* is performed by comparing senses for each word that is already in BTBWN with senses of the same word in a given dictionary. We did this in two ways comparing with senses registered in the Bulgarian Wiktionary and with senses in the Bulgarian explanatory dictionary. In many cases we reformulated the identified uncovered senses on the basis of the existing senses in BTBWN and with the requirement for a better mapping to the English PWN.

The *corpus-based approach* uses several mechanisms for identification of new words and senses to be added to BTBWN. They include at least the following ones: (1) annotation of new texts; (2) clustering of word forms in a large corpus on the basis of their contexts (Polish WordNet and some others); and (3) checking the coverage of BTBWN over a frequency list compiled from a large corpus. Currently, we perform point three over a frequency list compiled over a 7- million-word corpus covering different types of text. Our goal is to include all the words that appear in the corpus at least 100 times. Any word that is not presented in BTBWN is lemmatized in all possible ways and then it is included by each possible lemma and each possible sense. For example, the Bulgarian word form “поет”, “poet”, is lemmatized as a noun “поет”, “poet”, and a verb “поема”, “poema”, “take”. In this way we cover all frequent word forms in the corpus

with their relevant senses, independently from the context. The people who add the new words and senses are free to search for usages of the corresponding lemmas not only within the corpus, but also on the web.

The *linguistically-based approach* exploits productive phenomena within the language. We mainly exploit derivation patterns with clear new semantics. For example, the names of citizens of a given location is such a case: from "New York" to form "New Yorker". This pattern is easy to recognize in the corpus and the definition and mapping to the rest of BTBWN is thus predictable.

Besides these two current activities we plan to perform also the following two tasks: (1) addition of relational structure over BTBWN; and (2) including the synsets that are not mapped exactly to synsets in PWN to the Collaborative Interlingual Index (CILI) — [14]. For the latter task it is necessary to write appropriate definitions in English. For the former task we will exploit the mapping to English WordNet and additionally the mapping from the English WordNet to the Polish Wordnet. In this way we will be able to transfer relations between synsets in English and Polish WordNets to Bulgarian. Thus, we expect to impose a reliable relation structure over BTBWN. We manually will check the cases of lexical relations like *antonymy* and *derivation* and the cases where English and Polish Wordnets disagree with each other.

6 Conclusion

The paper discussed our strategy for the mapping of the word senses in a treebank to the WordNet ones in the context of the overall construction of the BTBWN. In the presented approach the resource annotation does not rely on pre-created WordNet, but rather on an explanatory dictionary of Bulgarian. Later on, these senses have been mapped to the PWN 3.0, while keeping the language specific concepts through the introduction of special markings. The adopted strategy allowed for dense connectivity between the resources, and at the same time it leaves room for the further creation of a language-specific hierarchy. Currently BTBWN contains 12,147 synsets equivalent to the synsets in PWN, and about 2500 additional synsets mapped as described in the paper.

These mappings have been exploited actively for knowledge-based word sense disambiguation of Bulgarian by using the English WordNet as a knowledge graph that transfers the linguistic relations to the Bulgarian lemmas. Also, they will determine the language-specific hierarchy of concepts over the Bulgarian definitions.

Acknowledgements

This research has received partial support by the grant 02/12 — *Deep Models of Semantic Knowledge (DemoSem)*, funded by the Bulgarian National Science Fund in 2017–2019. We are grateful to the anonymous reviewers for their remarks, comments, and suggestions. All errors remain our own responsibility.

References

1. Erjavec, T., Fišer, D.: Building Slovene WordNet. In: Proceedings of the 5th Intl. Conf. on Language Resources and Evaluations, LREC 2006 22 - 28 May 2006. pp. 1678–1683. Genoa, Italy (2006)
2. Fellbaum, C.: WordNet: An Electronic Lexical Database. MIT Press (1998)
3. Hajnicz, E.: The Procedure of Lexico-Semantic Annotation of Składnica Treebank. In: Proceedings of LREC-2014. pp. 2290–2297 (2014)
4. O'Grady, W.: The syntax of idioms. Natural Language and Linguistic Theory 16, 279–312 (1998)
5. Pociello, E., Agirre, E., Aldezabal, I.: Methodology and construction of the basque wordnet. Language Resources and Evaluation 45(2), 121–142 (2011)
6. Popov, A., Kancheva, S., Manova, S., Radev, I., Simov, K., Osenova, P.: The Sense Annotation of BulTreeBank. In: Proceedings of TLT13. pp. 127–136 (2014)
7. Postma, M., van Miltenburg, E., Segers, R., Schoen, A., Vossen, P.: Open Dutch WordNet. In: Proceedings of the Eight Global Wordnet Conference. Bucharest, Romania (2016)
8. Raffaelli, I., Tadić, M., Bekavac, B., Željko Agić: Building Croatian WordNet. In: Proceedings of the 4th Global WordNet Conference. pp. 349–359 (2008)
9. Rudnicka, E., Maziarz, M., Piasecki, M., Szpakowicz, S.: A strategy of Mapping Polish WordNet onto Princeton WordNet. In: Proceedings of COLING 2012: Poster. pp. 1039—1048 (2012)
10. Simov, K.: Ontology-based lexicon of bulgarian. Journal for Language Technology and Computational Linguistics 24(2), 40–55 (2009)
11. Simov, K., Osenova, P.: Language resources and tools for ontology-based semantic annotation. In: Proceeding of OntoLex 2008 Workshop at LREC 2008. pp. 9–13 (2008)
12. Simov, K., Osenova, P.: Formalizing multiwords as catenae in a treebank and in a lexicon. In: Verena Henrich, Erhard Hinrichs, Daniël de Kok, Petya Osenova, Adam Przepiórkowski (eds.) Proceedings of the Thirteenth International Workshop on Treebanks and Linguistic Theories (TLT13). pp. 198–207 (2014)
13. Simov, K., Osenova, P.: Catena operations for unified dependency analysis. In: Proceedings of the Third International Conference on Dependency Linguistics (Depling 2015). pp. 320–329. Uppsala University, Uppsala, Sweden, Uppsala, Sweden (August 2015), http://www.aclweb.org/anthology/W15-2135
14. Vossen, P., Bond, F., McCrae, J.P., Fellbaum, C.: CILI: the Collaborative Interlingual Index. In: Eighth meeting of the Global WordNet Conference (GWC 2016). pp. 50–57. Bucharest, Rumania (2016)