Recent Developments in BulTreeBank-WordNet (BTB-WN)

Kiril Simov and **Petya Osenova**
Institute of Information and Communication technologies
Bulgarian Academy of Sciences
`kivs|petya@bultreebank.org`

Abstract

The paper reports on recent developments in Bulgarian BTB-WordNet (BTB-WN). This resource is viewed as playing a central role with respect to the integration and interlinking of various language resources such as: e-dictionaries (morphological, terminological, bilingual, orthographic, etymological and explanatory, etc., including editions from previous periods); corpora (coming from outside or being internal - like the corpus of definitions as well as the corpus of examples to synset meanings); ontologies (such as CIDOC-CRM, DBpedia, etc.); sources of world knowledge (such as information from the Bulgarian Encyclopedia, Wikipedia, etc.). The paper also gives information about a number of applications built on BTB-WN. These are: the Bulgaria-centered knowledge graph, the *All about word* application as well as some education-oriented exercises.

1 Introduction

In this paper we report on the developments of the Bulgarian BTB-WordNet (BTB-WN) during the last three years (2020, 2021, 2022). The development of BTB-WN goes back to the times when an Ontology-based lexicon for Bulgarian was initially constructed (Simov and Osenova, 2010). Here we started with the concept set from the upper ontology DOLCE[1]. Then it was extended with concepts selected from the OntoWordNet (Gangemi et al., 2003), which correspond to Core WordNet and EuroWordNet Base concepts[2]. The construction of the Ontology-based lexicon - that later on evolved into the BTB-WN – was driven by the need of such a resource for some NLP applications like domain ontology text annotation, word sense disambiguation, co-reference resolution, machine translation and others. However, it turned out

that each of these applications required not only available resources but also appropriate integration among them. The interface between the lexical semantics and grammar, between the lexicons and corpora has been extensively discussed from various points of view: linguistic, typological, formal, implementational, etc. Either starting point causes problems – the lexicalist-centric and the grammar-centric ones. Here we support the point of view in which the grammar is born in the lexicon, i.e. the lexicalist-centric one, without lowering the role of the grammar at all. This view is on a par with the linguistic theories that are constraint-based (such as HPSG and LFG) or are word-based (dependency theories). It is also in line with the ideas behind the flagship project in eLexicography – ELEXIS[3]. The result from this project is a roadmap in eLexicography where all the steps in the various life cycles of producing a dictionary have been studies, documented, implemented and tested. The interested reader is forwarded to (Tiberius et al., 2021).

It is well-known that WordNets are thesauri and despite providing the meaning of words grouping them within synsets and providing relations among these synsets, they are still very static, self-contained and often do not cover all parts of speech. At the same time, they are good candidates for playing a central role – like a hub – for linking grammar, other lexical data and world knowledge. Our ultimate goal however would be that users could customise their own dictionaries, examples or other material through these interlinked resources. For that reason, along with cleaning the meanings and relations within BTB-WN, we started also other tasks such as: linking lemmas to their morphosyntactic characteristics through a rich tagset and morphological/inflectional dictionary of Bulgarian; linking meanings to examples from corpora; constructing a corpus of definitions, annotated with senses from BTB-WN; adding domain

---

[1] http://www.loa.istc.cnr.it/dolce/overview.html
[2] http://globalWordNet.org/resources/gwa-base-concepts/
[3] https://elex.is/

terms; adding dictionaries from previous times with their specific spellings; constructing a Bulgaria-centric knowledge graph as an extension of BTB-WN; aligning different ontologies with respect to BTB-WN; using the lexical chains over the BTB-WN graph for generating correct sense detection drills for Bulgarian learners.

The extension of BTB-WN with information from the Bulgarian Wikipedia has been enhanced in three directions: adding concepts, adding instances, and adding properties. The idea behind this approach is to support the mapping of the ontology with the vocabulary of BTB-WN as well as the mapping of the BTB-WN relations to knowledge graphs created on the basis of Wikipedia, DBpedia, Wikidata, etc. Such mappings would also facilitate the knowledge extraction from the wiki media themselves. This endeavour is in line with works like (McCrae and Cillessen, 2021), where a method is presented for linking English Wordnet with Wikidata.

The paper is structured as follows: the next section outlines some related works from different perspectives and thus is not exhaustive. Section 3 describes the linking of BTB-WN with in-house and external resources. Section 4 focuses on some BTB-WN based applications. Section 5 concludes the paper.

2 Related works

It is difficult to refer to the great number of publications that discuss various parameters of integration and usage of WordNets. Also, here we do not focus on the integration and representation of WordNets through formatting standards like LEMON[4], LMF[5], etc. but rather on resource integration where the WordNet plays the main role. For that reason, only some of the many works are cited here with the aim to illustrate our work in the context of the existing research.

A lot of works have been devoted to the usage of language specific and multilingual resources such as monolingual and bilingual dictionaries for the quicker and less expensive construction of Word-Nets. For example, (Siegel and Bond, 2021) report on the construction of the German WordNet called OdeNet and (Fišer and Sagot, 2015) report about the creation of the Slovene WordNet. Our BTB-WN was constructed semi-automatically with the

combination of both established methods - *expand* and *merge*. The automatic part was used when extracting data from Bulgarian resources and for merging it before being validated by a human.

(Bentivogli et al., 2004) share their experience on how to incorporate domain lexica in their Word-Net. As expected, the main reported problems were in the synchronization of the hierarchies between the WordNet and the specialized thesaurus in the domain of architecture. In our case the inclusion of domain terms still follows the WordNet hierarchy.

(Ahmadi et al., 2020) present a method for an automatic alignment between the senses of the same lemma across two monolingual Danish dictionaries that come from two periods - modern and historic. In our case we have performed automatic lemma-based alignments among a contemporary dictionary of Bulgarian and an older one. The spellings in both resources differ. A sense alignment has not been performed yet but it is envisaged as a future task.

(Laparra et al., 2009) present a graph-based Word Sense Disambiguation algorithm for integrating WordNet with FrameNet[6]. In our case the integration considers VerbNet[7] first – through the customized mapping with the Bulgarian Valency Dictionary (BVD) (Osenova et al., 2012). FrameNet is incorporated through the inclusion of the event-evoking verbs within the Bulgarian Event Corpus (Osenova et al., 2022). These events have been annotated with named entities, roles and relations adapted from FrameNet and CIDOC-CRM ontology[8].

(Oliver, 2020) surveys various techniques for aligning Wikipedia with WordNet. The author concludes that the evaluation of alignments between the two is still an open research task. In our mapping strategy we use a rule-based approach with a post-editing validation by a human.

In (Rudnicka et al., 2022) the gaps in mapping Polish and English WordNets were identified and addressed. Such gaps are observed also in our case, and although we preserve the mappings with the Open English WordNet (OEW), we also try to make the Bulgarian hierarchy more natural to the Bulgarian cultural environment and speakers with removing the artificial intermediate nodes and with adding Bulgarian hypernyms and hyponyms.

---

[4]https://www.w3.org/2019/09/lexicog/
[5]https://www.iso.org/standard/68516.html

[6]https://framenet.icsi.berkeley.edu/fndrupal/
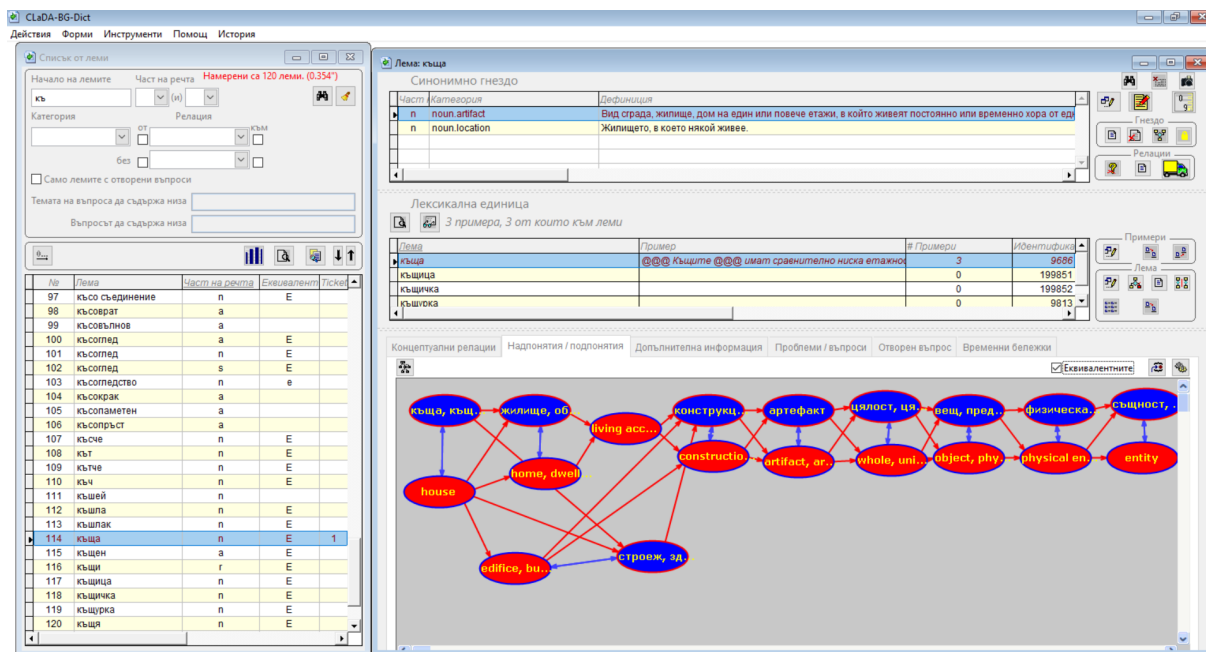[7]https://verbs.colorado.edu/verbnet/
[8]https://www.cidoc-crm.org/

Figure 1: A screenshot of the user interface of CLaDA-BG-Dict. It shows a search of lemmas against several criteria within the current BTB-WN (on the left); A synset editor (on the right) – shows all the synsets (upper part of the window) for a selected lemma; for the selected synset it shows its category, a definition and a list of synonyms. For each lemma there are assigned examples as well as mappings to the respective inflectional paradigms. At the bottom of the window a graphic representation of a noun hierarchy is given and also the mapping to the English synsets.

3 Extending and Linking BTB-WN

One important step we performed within the period of work reported here is the switching from a tool that supported only local editing (where synsets were considered within a very limited context) to a tool that supports editing of the Wordnet data within a global context. In Fig. 1 the main user interface of the system is presented.

When a lemma is selected within BTB-WN, the following information can be accessed immediately: the number of synsets related to it with the part-of-speech as well as the numbered meanings and links to the Open English WordNet. The usage of almost each synonym within a synset is illustrated with examples. Within the system the user could consult several other sources of information. The center of the system is BTB-WordNet. The user could open as many editor forms as necessary in which to observe the synsets for different words. Similarly the Open English Wordnet is available in the system. The creation of a new Bulgarian synset could start from scratch – entering all the information, including relations to other systems. But it is also possible to create such a synset with using an equivalent English synset. In this way the relations of the English synsets are automatically transferred

to BTB-WN. Also, a graphic is provided that reflects the ratio among the relations that are relevant to the synset (not visible in Fig. 1). In addition, the hypernyms and hyponyms can be observed as well.

The user can access the requested lemma in two ways: a) through writing it in the search box, or b) through finding it in the list of all lemmas. If the first option has been chosen, then the available information about the lemma is immediately presented. If the second option has been chosen, then one can see the part-of-speech and then enhance further information such as statistics of the lemma occurrence in the resource, or access the lemma information in the current representation or in a separate one. The 'Search' option provides various filters for making the inquiry more accurate. These include not only the lemma but also part-of-speech, lexicographic category, various relations. If the query is too broad for the database to return a reasonable set of examples, the user is prompted to specify it further. Another possibility to access the available lexical information is through the specific ID of the synset.

In addition to granting access to OEW, the system provides access to dictionaries that are freely available to us, among which the Bulgarian Explanatory Dictionary, our in-house Bulgarian Inflec-

Figure 2: Information about the lemma 'apple'. This is a system providing access to the BTB-WN for external users.

tional dictionary, two Bulgarian-English dictionaries. Each of these dictionaries could be consulted in isolation or simultaneously on the base of the alignments performed through lemmas. The user could also define different lists of lemmas which to be mapped to the vocabulary of BTB-WN and to the vocabularies of the included dictionaries. In this way it can be decided which new lemmas to be included within BTB-WN, or which to be used within a given application. Currently we support vocabularies co-responding to Bulgarian learners' levels like A1-A2, B1-B2, C. Also vocabularies of two student spelling lexicons and a list of the first 10 000 ranked lemmas were added against the Bulgarian Referent Corpus. The information within dictionaries is available within the editor form under the tab labeled as 'Additional information'. When exploring regular expressions, the user could observe different patterns of lemmas within the dictionaries.

Through the lexicographic classes (such as verb.social, verb.cognition, etc.) the synsets are connected also to the Bulgarian Valency Dictionary. This linking has not been implemented in our system yet since the Valency dictionary is being curated by specialists. For example, if the verb.emotion 'worry' is considered, the Bulgarian counterpart is displayed with a definition and a valency frame where the Subject has the role of Experiencer and the complement event that causes

worrying has the role of Stimulus. The link to the VerbNet frame is also given[9]. The transfer of valency frames from English to Bulgarian through an English-centered resource is not trivial. For that reason, often the initial frames are customized accordingly. As best practices for valency dictionaries we follow the Czech VALLEX (Lopatková et al., 2016) and the Polish Walenty (Przepiórkowski et al., 2014), among others.

In addition to the data access options, described above, one can search with the selected lemma in various corpora. We consider the definitions and examples already included in BTB-WN as a corpus from which to select examples for other senses. In this case we could construct sense annotated corpora similar to the (Rademaker et al., 2019). The system provides access to text corpora. For searching in the textual corpora the user has to point to a given corpus compiled from a text format where the metadata (like the source, for example) is introduced inside the text as a new line starting with a special symbol (@). The user might incrementally compile through various searches their own corpus with examples since there is an option of adding previously extracted results to the new ones.

With these functionalities, we performed a full examination of the existing version of BTB-WN (version 3.0) at the time when our working system was ready. BTB-WN contained a little more than

---

[9]https://verbs.colorado.edu/verbindex/vn/marvel31.3.php

19 000 synsets. Each synset was checked with respect to the following criteria:

- *Appropriateness of definitions.* We have checked the definitions for the different kinds of word classes and also per synset. This step was necessary, because in many cases the definition types in our resource differ from the ones in paper dictionaries. This holds especially for adjectives. In the traditional dictionaries the adjective is usually defined as qualifying a noun. In our case we go further and develop the definition of the adjective also to the specific features of the qualified noun. This holds especially for the relational adjectives like 'sofiyski' (Sofia-adjective). This adjective might relate to something: that originates in Sofia; is made in Sofia style; is placed in Sofia, etc.

- *Alignment to OEW.* In version 3.0 we supported as many relations as possible between the Bulgarian and English synsets. With the switch to the global view it became much more convenient to verify these mappings.

- *Missing senses.* The construction of the BTB-WN up to version 4 was mainly driven by specific NLP tasks, as it was mentioned in the Introduction. Thus, it reflected the needs of these tasks. Now we decided to check the coverage of the resource with respect to the most common and well-established senses.

- *Relations.* When a Bulgarian synset was created on the basis of the corresponding English one, the relations were transferred automatically. After the transfer the set of relations became eligible for modifications, if needed.

- *Appropriateness of examples.* The assigned examples were specially checked with respect to their appropriateness to the corresponding sense. The most frequent error was when the example did not provide enough context for the meaning, and thus the corresponding word form might have been interpreted ambiguously. In such cases the example was extended or deleted.

Besides the examination of the existing synsets we have extended BTB-WN with new synsets through the above mentioned vocabularies extracted from both types of sources - dictionaries

and corpora. This was performed in line with our goal to cover the senses of the most common lemmas in Bulgarian. At the moment we completed the coverage of the core vocabulary with about 6000 lemmas. Then the following information was added: derivational sets for these lemmas such as adjectives derived from nouns, aspectual variants of Bulgarian verbs that share a common basic sense, etc. In this way, more than 14 000 synsets were added. For the addition of examples we compiled and used a concise guide. For the moment it is for our internal usage only, but it will be available also in English for better accessibility by anyone who would be interested in it. The short guide explains how the examples were selected that are connected to senses, how to better search for examples in corpora and on the net, and what the recommendation criteria are for this selection. The best examples always should reflect some of the characteristics given in the definition, or add to them. For instance, if we want to give a good example to the *noun.artifact* 'pair of trousers', we might take the following one: 'The right leg of his trousers was split to allow his plastered leg to pass through'. Here the sentence reveals the following facts: that the trousers cover legs and that the trousers have two parts.

4 BTB-WN based applications

In this section we present some of the applications of BTB-WN that were developed recently or are under development.

The first application is the role of BTB-WN in the Bulgaria-centric knowledge graph. We consider the knowledge graph a core semantic repository for Bulgarian research infrastructure related to CLARIN[10] and DARIAH[11]. For that reason BTB-WN has been further enriched with terms from various Social Sciences and Humanities domains such as history and ethnography. Here two challenges appeared. The first one is related to the introduction of terminological multiword expressions while the second one refers to the register of usage such as being archaic or dialectal, etc.

For example, let us take the Bulgarian folk units of measurement. They are linked with a hyponymy relation to the concept about the official Bulgarian folk units (such as 'pedya' (span), 'prast' (finger), 'lakat' (elbow), etc.) and the concept for linear

---

[10]https://www.clarin.eu/
[11]https://www.dariah.eu/

units (such as the unit for length).

The inclusion of domain terms in the WordNet would allow the annotation of domain texts with word senses. These then might be used for training domain-specific semantic taggers and would be able to contribute to the task of natural language understanding.

The contemporary terminological lexicons very often comprise detailed encyclopedic knowledge. We do not incorporate such detailed entries in BTB-WN, but just concise definitions and references to the respective terminological lexicon. This step is similar to the operation of mapping from BTB-WN to Wikipedia.

The generalization of this approach grew into the creation of a hub for a bigger net of dictionaries and resources, called in our case 'All about words'. In this application we reused the integration of dictionaries within the system for further creation of BTB-WN in order to provide as much as information as possible about the Bulgarian words. The system includes a concordancer, a Wordnet viewer, a word form analyser, a viewer for the Bulgarian inflectional dictionary, viewers for other dictionaries. Thus, the user can run the concordancer with the query expression of interest. From the returned concordance lines the users could select arbitrary word forms and require information about them. The system applies the word form analyser which returns all possible lemmas for the word form with the appropriate part of speech. For example, if we type the word 'belya', it will return three part-of-speech types: peel (verb), peel oneself (verb), white (adjective) and mischief (noun). Then the system switches to a different browser tab where the user could consult different resources via these lemmas. At the heart of these interrelated resources come BTB-WN and the Inflectional lexicon. The user could observe the paradigm of the selected lemma, its meanings in BTB-WN, brief information from other dictionaries in which the lemma is presented, and a list of examples extracted from the sense annotated treebank of Bulgarian, etc. From this tab the user could switch to other tabs in order to consult the corresponding resource in more detail. Also the user could switch back to the concordancer for searching examples about other word forms.

For example, if we choose the noun (mischief) from the above list of ambiguous lemmas, then the noun paradigm will be made visible. If one of the

verbs is chosen, then the verb paradigm of present tense in all persons and numbers is made visible. All other verb forms are planned to be made available as well, irrespective of whether they are synthetic or analytic. If the user clicks on a specific wordform of the paradigm information, they can see the respective description of the grammatical characteristics like the following: for the lemma 'belya' the description is: verb, personal, imperfective, transitive, indicative, present tense, 1 person, singular; for the word form 'belyat' the description will change in the indicated places which are: 3 person, plural. When the user selects the BTB-WN visualization page, all synsets of the lemma are listed, a graphical presentation of the relational graph around the synset is visualized. Additionally, users can type another lemma and see all the synsets in which this lemma participates. When a synset is selected, also a graphical view with the available relations is shown. An example for the lemma 'apple' is given in Fig. 2.

In Fig. 2 the following information is displayed. On the left side the lemma 'apple' is given as number 1. As number 2 one can see an idiom starting with this lemma, namely 'apple of discord'. However, the first one has been selected. In the middle column both meanings of the lemma have been listed - as an apple tree and as the apple fruit - together with the mappings to Open English WordNet. The third column presents a visualization of some of the relations in which the meaning for 'apple tree' participates. These are as follows: the immediate hypernym is a 'fruit tree' and the next level hypernym is a 'tree'. There is an immediate hyponym which is a 'wild apple tree'. Through the equality relation on the right, the Bulgarian lemma is related to the English one. Last, but not least, on the left, a derivation relation is established to the adjective 'apple'.

While the above described applications serve mostly as a guide to the specifics of Bulgarian words, the next one that we discuss here is more educationally oriented. It is a newly developed application called 'Game of Meanings'. The user receives a task where they have to select the correct definition per lemma in a sentence from the examples associated with one of the synsets for this lemma. The definitions in the multiple choice task as well as the contexts in which a certain lemma was used come from BTB-WN. An example from the beta version is shown on Fig. 3. Each game
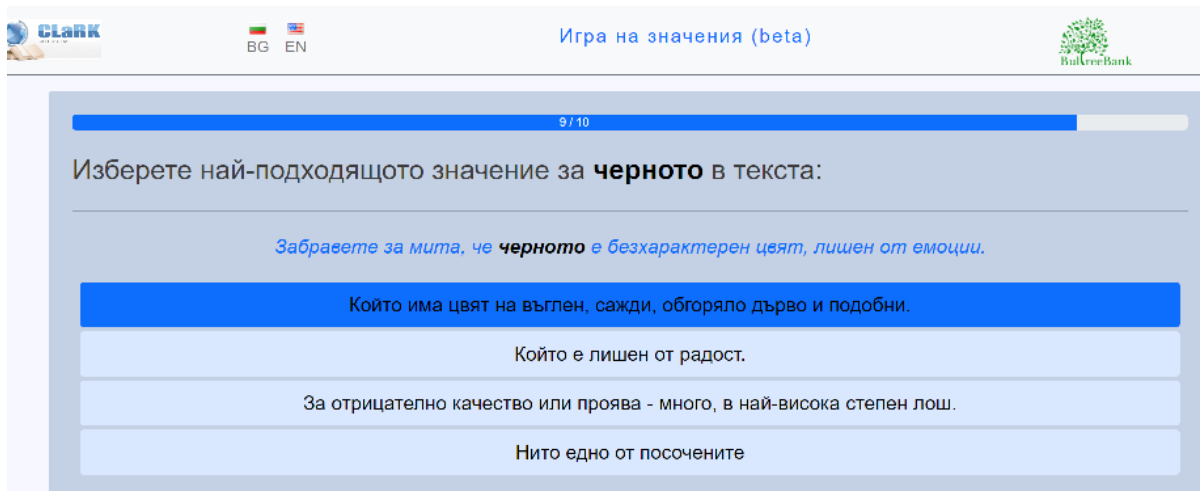
Figure 3: The multiple choice question on the color of black.

consists of a set of 10 tasks. Each task includes an example for a selected lemma in a selected synset with four alternative answers. Each answer is a definition or a message for a missing definition - the algorithm for generation of tasks is given below.

In Fig. 3 the following task is given: Select the most appropriate meaning for the word 'the black' in the text that says: *Forget the myth that 'black' is a featureless color that lacks emotion.* Four possible definitions are given to the player to select from. They are: 1. Which has the color of charcoal, soot, burnt wood and the like; 2. Which lacks joy; 3. For a negative quality or manifestation - very bad; 4. None of these.

Here the correct answer is supposed to be the first one (1). It should be noted that the more similar definitions to select from, the more difficult the task is, and vice versa.

The algorithm for generating the tasks includes these steps: a) a lemma is selected; b) a synset is chosen for this lemma; c) from this synset an example is selected; d) the available definition is given as an option to choose from. e) the other alternatives are selected over the synsets with the lexical chains navigation algorithm (Hirst and St-Onge, 1996) over the BTB-WN graph including the other synsets of the lemma, if available. In cases when there are no enough options, a string-based similarity search is performed with respect to the initial lemma.

We imagine that such a type of game would increase the ability of students but also of the whole interested community to improve their reading with understanding. It should be noted that with respect to the task of 'reading with understanding' Bulgar-

ian students perform poorly in comparison to their peers in Europe.

5 Conclusions

In this paper we present an environment where the BTB-WN plays a central role in displaying all the available information about a lemma in Bulgarian - synsets, associated definitions and examples, grammatical information in the form of paradigms and descriptions, possibilities to search in corpora of all definitions or in external ones. Thus, our approach is lemma-based but at the same time it starts from the lexical semantics and through various linking strategies incorporates also the grammar and pieces of world knowledge.

Our future plans are to add more information of all kinds and more relations as well as relation directions among the resources. Needless to say, approaches for automation of resources enrichment and linking are also envisaged.

In addition to the presented tasks, we have been working also on generation of exercises for mastering Bulgarian grammar. Since the exercises use our dictionaries and patterns to produce as many drills as possible, very often their semantics is questionable. This fact causes a serious problem to the freedom of the underlying generating algorithms since the users should be prevented from seeing and memorizing nonsense or pedagogically and ethically flawed messages. Thus, even in automatized exercises such as drills we should be very careful about what suggestions we provide to trainees. Following this line, we plan not to stop the generating power per se but to use BTB-WN (integrated with

the Bulgarian Valency Dictionary) as a semantic filter in the exercise production module.

6 Acknowledgements

References

Sina Ahmadi, Sanni Nimb, Thomas Troelsgård, John P. McCrae, and Nicolai H. Sørensen. 2020. Towards Automatic Linking of Lexicographic Data: the case of a historical and a modern Danish dictionary. Zenodo.

Luisa Bentivogli, Andrea Bocco, and Emanuele Pianta. 2004. ArchiWordNet: Integrating WordNet with Domain-Specific Knowledge. In *Proceedings of the 2nd International Global Wordnet Conference*, pages 39–47.

Darja Fišer and Benoît Sagot. 2015. Constructing a poor man's wordnet in a resource-rich world. In *Language Resources and Evaluation*, pages 601–635.

Aldo Gangemi, Roberto Navigli, and Paola Velardi. 2003. The ontowordnet project: Extension and axiomatization of conceptual relations in wordnet. In *On The Move to Meaningful Internet Systems 2003: CoopIS, DOA, and ODBASE*, pages 820–838, Berlin, Heidelberg. Springer Berlin Heidelberg.

Graeme Hirst and David St-Onge. 1996. *Lexical chains as representations of context for the detection and correction of malapropisms*. The MIT Press.

Egoitz Laparra, German Rigau, and Montse Cuadros. 2009. Exploring the Integration of WordNet and FrameNet. In *Proceedings of the 5th Global WordNet Conference*.

Markéta Lopatková, Václava Kettnerová, Eduard Bejček, Anna Vernerová, and Zdeněk Žabokrtský. 2016. *Valenční slovník českých sloves VALLEX*. Karolinum, Praha.

John P. McCrae and David Cillessen. 2021. Towards a linking between wordnet and wikidata. Zenodo.

Antoni Oliver. 2020. Aligning Wikipedia with Word-Net:a review and evaluation of different techniques. In *Proceedings of the Twelfth Language Resources and Evaluation Conference*, pages 4851–4858, Marseille, France. European Language Resources Association.

Petya Osenova, Kiril Simov, Laska Laskova, and Stanislava Kancheva. 2012. A treebank-driven creation of an OntoValence verb lexicon for Bulgarian. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, pages 2636–2640, Istanbul, Turkey. European Language Resources Association (ELRA).

Petya Osenova, Kiril Simov, Iva Marinova, and Melania Berbatova. 2022. The Bulgarian event corpus: Overview and initial NER experiments. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*, pages 3491–3499, Marseille, France. European Language Resources Association.

Adam Przepiórkowski, Elżbieta Hajnicz, Agnieszka Patejuk, Marcin Woliński, Filip Skwarski, and Marek Świdziński. 2014. Walenty: Towards a comprehensive valence dictionary of Polish. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, pages 2785–2792, Reykjavik, Iceland. European Language Resources Association (ELRA).

Alexandre Rademaker, Bruno Cuconato, Alessandra Cid, Alexandre Tessarollo, and Henrique Andrade. 2019. Completing the Princeton annotated gloss corpus project. In *Proceedings of the 10th Global Wordnet Conference*, pages 378–386, Wroclaw, Poland. Global Wordnet Association.

Ewa Rudnicka, Łukasz Grabowski, Maciej Piasecki, and Tomasz Naskręt. 2022. In Search of Gaps between Languages and Wordnets: the Case of Polish-English WordNet. *International Journal of Lexicography*. Ecac005.

Melanie Siegel and Francis Bond. 2021. OdeNet: Compiling a GermanWordNet from other resources. In *Proceedings of the 11th Global Wordnet Conference*, pages 192–198, University of South Africa (UNISA). Global Wordnet Association.

Kiril Simov and Petya Osenova. 2010. Constructing of an ontology-based lexicon for Bulgarian. In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC'10)*, Valletta, Malta. European Language Resources Association (ELRA).

Carole Tiberius, Simon Krek, Katrien Depuyd, Polona Gantar, Jelena Kallas, Iztok Kosem, and Rundell Michael. 2021. Towards the ELEXIS data model: defining a common vocabulary for lexicographic resources. Zenodo.