

# Semantic Classification of Prepositions in BulTreeBank WordNet

Zara Kancheva

Artificial Intelligence and Language Technologies Department,  
Institute of Information and Communication Technologies,  
Bulgarian Academy of Sciences, Bulgaria  
zara@bultreebank.org

## Abstract

The paper presents the work in progress for a PhD thesis about preposition incorporation in the Bulgarian BulTreeBank WordNet. Being one of the most polysemous parts of speech, prepositions are still relatively challenging for NLP and are usually missing in wordnets. A preposition semantic classification, a model for preposition synsets and synset relations are proposed. The planned applications of the prepositions and the directions for future processing are introduced.

## 1 Introduction

The aim of the paper is to present the work in progress for a thesis about incorporation of prepositions in the structure of BulTreeBank WordNet (BTB-WN) for Bulgarian. Prepositions are viewed as a necessary part of speech in a lexical resource like wordnet, because their integration would seriously expand its range of applications and would be beneficial for several NLP tasks (such as semantic annotation, word sense disambiguation, machine translation, parsing, knowledge extraction, word embeddings, text analysis and generation, etc.). Prepositions are governing words and present the language-specific semantics more completely, so they would contribute to better automatic translation generation and better word embeddings in Bulgarian. The processing and representation of prepositions is a challenging task mainly due to their high polysemy, so they are often missing in lexical resources.

In addition to the extension of BTB-WN the next stage of the work is going to be dedicated to neural models building and prepositions are important for this task because they have a considerable role in the semantics of the text. Algorithms will be developed for the generation of artificial texts in Bulgarian and these texts will use the semantic classification of prepositions and their relations with BTB-WN in order to produce more natural pseudo corpora. Within CLaDA-BG<sup>1</sup> the work will play manifold roles: 1) representation of preposition semantics; 2) supporting language model training for Bulgarian; 3) NLP applications such as word sense disambiguation. The following features will be considered: semantic classification of prepositions, the BTB-WN categories of verbs and nouns to which the prepositions refer and later the frames of the verbs and the semantic roles of the nouns from the Bulgarian OntoValence lexicon. A model for preposition synsets and relations in BTB-WN will be presented.

Explanatory dictionaries typically contain prepositions, but do not present them in a way that is satisfactory for NLP tasks. Some resources such as treebanks, parsers, etc. include prepositions, but for wordnets this part of speech is unusual. The processing of prepositions typically includes semantic and syntactic classification, thematic roles, categorisation by wordnet hierarchy structure or data from resources like valency lexicons, FrameNet<sup>2</sup>, VerbNet<sup>3</sup>, PennTreebank (Marcus et al., 1994) and PropBank<sup>4</sup>.

In this study a semantic classification of Bulgarian prepositions is done and the classes are used as synset categories for the prepositions, like the synsets for any other part of speech have such categories.

---

This work is licenced under a Creative Commons Attribution 4.0 International Licence. Licence details: <http://creativecommons.org/licenses/by/4.0/>

<sup>1</sup><https://clada-bg.eu/en/>

<sup>2</sup><https://framenet.icsi.berkeley.edu/fndrupal/>

<sup>3</sup><https://verbs.colorado.edu/verbnet/>

<sup>4</sup><https://propbank.github.io/>

Preposition synsets have the same structure and features as the synsets for other parts of speech in BTB-WN, however less relations are relevant for them. These synsets have definitions, examples, synonyms and antonyms if available, and the synset category shows their semantic class. This topic is going to be further explored in the future work, because some studies (Amaro, 2018; Harabagiu, 1996) show that there are more applicable relations for preposition synsets, for example *hyponymy/hyperonymy* and *causes/is caused* relations, which are typically used for nouns and verbs.

BTB-WN was created in several steps. The work on it started on the base of the Core WordNet subset<sup>5</sup> of Princeton WordNet (PWN) (Fellbaum, 1998) which contains the 5000 most frequent English senses. Later it was expanded with content words from the BulTreeBank and a Bulgarian frequency list, and also with encyclopedic data: senses from the Bulgarian versions of Wikipedia and Wiktionary. BTB-WN was successively expanded by several initiatives, such as expansion with multi-word expressions, named entities, terms from certain domains, etc. Originally BTB-WN was mapped to the PWN, but since 2020 the mapping is transferred to the Open English WordNet (EOW) (McCrae et al., 2020), because it is being currently maintained and developed, unlike PWN. Recently several new synset relations were introduced in BTB-WN in addition to the relations from PWN and OEW, which are also used with some modifications. The latest version of BTB-WN is 4.0<sup>6</sup>. It contains more than 33 000 synsets and is available for online browsing<sup>7</sup>. Soon BTB-WN 4.0 will be freely available for downloading in WordNet-LMF XML format.

Section 2 gives an overview of relevant studies about prepositions, Section 3 presents the semantic preposition classification that is used, Section 4 contains the synset model for prepositions in BTB-WN, and Section 5 introduces the plans for future work and Section 6 concludes the paper.

## 2 Related Works

The processing of prepositions is a challenging task for NLP mainly because of the high polysemy of prepositions. Yet there are many studies which attempt at sense disambiguation and semantic classification of prepositions.

Typically wordnets do not contain closed-class words including prepositions (currently, the only wordnet with prepositions is the BulNet for Bulgarian but they do not have relations and hierarchy<sup>8</sup>), but there are recent studies which show such attempts. Amaro (2018) presents an approach for incorporation of prepositions in wordnet, particularly of Portuguese prepositions for movement. The work provides tests for the establishment of several semantic relations between prepositions, following the relations for nouns, verbs, adjectives and adverbs in wordnets: *synonymy*, *antonymy*, *hyponymy*, *hypernymy*, *cause* and *is caused by*. These tests could be applied to prepositions in different languages. The approach towards prepositions in BTB-WN follows that of Amaro (2018), but with the aim to identify even more relations.

The research of Da Costa and Bond (2016) provides another proof for the benefits of enriching wordnets with different parts of speech by introducing non-referential concepts. They incorporate interjections, numeral classifiers and exclamatory pronouns in the Open Multilingual WordNet<sup>9</sup> and establish several kinds of relations between them and the other parts of speech.

A resource fully dedicated to prepositions is the PrepNet (Saint-Dizier, 2008). It introduces a categorisation based on thematic roles particularly for French prepositions, but it is applicable also for English, Spanish and German. PrepNet distinguishes prepositions in two levels – abstract notion level and language realization level. The first one is conceptual and does not differ in languages and the second one regards the realizations in different languages.

The Preposition Project is a resource “designed to provide a comprehensive characterization of preposition senses suitable for use in natural language processing” (Litkowski and Hargraves, 2005). It uses

<sup>5</sup><http://wordnetcode.princeton.edu/standoff-files/core-wordnet.txt>

<sup>6</sup><https://clada-bg.eu/bg/centers-and-services/language-technologies/btb-wordnet.html>

<sup>7</sup><https://concordance.webclark.org/>

<sup>8</sup><http://dcl.bas.bg/en/resursi/wordnet/>

<sup>9</sup><https://omwn.org/>

the presentations of prepositions from three different sources and compares them – the Oxford dictionary of English, tagged prepositions from FrameNet and the prepositions from an English grammar. This way every preposition sense in the Preposition Project is described by a semantic role name, syntactic and semantic properties of its complement and attachment point, link to its definition in the dictionary, syntactic function and meaning from the grammar, different prepositions with a similar semantic role, frames and frame elements from FrameNet, other syntactic forms in which the semantic role may be found and the position of the given preposition in a network of prepositions.

Lassen (2006) presents a novel approach towards an ontology-based preposition processing – 16 preposition senses are used in the study to mark the semantic relations between noun phrases and the result of that is aimed at determining conceptual distance for information retrieval purposes.

The resource of Schneider et al. (2015) based on preposition supersenses provides a general purpose classification of prepositions, aimed at automatic WSD, where every supersense has detailed explanations, dictionary senses, example sentences and mapping to other resources, all of that for the benefit of the annotators.

O’Hara and Wiebe (2003) perform a WSD of prepositions based on the semantic role annotations of the Penn TreeBank and FrameNet. They use the semantic roles as senses of the prepositions. Moreover, they enrich the standard approach of WSD for using collocations by taking into consideration also WordNet hypernyms. High-level synsets are serving as collocations in the form of semantic categories. The two resources that are used are providing data with different levels of granularity – the Penn Treebank semantic roles classification is more compact (7 types) and that of FrameNet is very fine-grained (140 classes) (the classifications are presented in Section 3). The results show that the accuracy is higher with the less fine-grained roles from PennTreebank, but in the processing of the both resources the wordnet hypernyms prove beneficial. In the later research of O’Hara and Wiebe (2009) the topic of preposition disambiguation is further explored with the addition of data from more semantic role resources.

For the aims of predicting semantic relations Srikumar and Roth (2013) use the annotated prepositions from the SemEval 2007 shared task of WSD for prepositions and semantic roles from FrameNet. Their approach includes identifying the semantic types of the prepositions’ governor and object and for that wordnet hypernyms are used. The method is very beneficial, because for polysemous prepositions, the sense prediction is only possible with information for the arguments.

Similarly, Bailey et al. (2015) exploit VerbNet frames, wordnet relations and selectional restrictions with the purpose of resolving prepositional phrase attachment.

### 3 Semantic Classification of Prepositions

The semantics of Bulgarian prepositions are very well explored and many classifications are available. Several of them (Stoyanov, 1983; Konstantinova, 1982; Boyadzhiev et al., 1998) were consulted for the goal to adapt a more compact general-purpose classification of prepositions for the integration in the BTB-WN. The first categorisation (Stoyanov, 1983) is the most thorough – some prepositions there have around 30 senses, because the classification separates all subsenses and provides examples for rare, archaic, dialectal, etc. usages. The classification of Boyadzhiev et al. (1998) is on the contrary very generalized – it has 13 categories and does not include examples from different speech registers. The overview of the history of classifications of Bulgarian prepositions made by Konstantinova (1982) is also taken into consideration.

The resulting adapted classification contains 15 categories of prepositions: location, time, transition, manner and instrument of action, possession, quantity, degree and exceeding of limit, purpose, origin and part of a whole, opposition, comparison, cause and object class: exchange, exclusion, opinion and thought. Some modifications differing from the above-mentioned classifications are done in order to make the groups of prepositions in BTB-WN more compact.

For example, the closely related classes `manner of action` (вЪРВЯТ в редица *vървят v reditsa* “they walk in a line”) and `instrument of action` (ЯМ с вилица *jam s vilitsa* “I eat with a fork”) here are united in one category.

The same approach is applied to the *origin* (изследовател от България *izsledovatel ot Bălgarija* “researcher from Bulgaria”) and *part of a whole* (филм в две части *film v dve časti* “movie in two parts”) classes.

The approximation of time (около 9 часа *okolo 9 časa* “around 9 o’clock”) and approximation of quantity (към 3 километра *kăm 3 kilometra* “about 3 kilometers”) classes from Stoyanov (1983) here are generalised and included respectively in the *time* and *quantity* classes.

The exceeding of limit sense (това е свръх силите ни *tova e svrăh silite ni* “this is beyond our powers”) is considered a part of the *quantity* category.

An *object superclass* is created to unite the expression of relations for *exchange* (ще го направя вместо теб *šte go napravja namesto teb* “I will do it instead of you”), *exclusion* (яке без копчета *jake bez kopčeta* “jacket without buttons”), *thought* (разсъждавам върху проблема *razsăždavam vărhu problema* “I reflect on the problem”) and *opinion* (за мен това е най-доброто решение *za men tova e naj-dobroto rešenje* “for me this is the best decision”). The *prep.obj.thought* class includes expression of object of thought, speech and writing.

Additionally, a decision is made for the metaphorical usages of a given class to be considered part of it, not a separate group. For instance, usages like *тя го чу сред всички гласове* *tja go ču sred vsichki glasove* “she heard him among all the voices” are considered as examples of the *location* class.

Table 1 shows the semantic classes and synset categories of the prepositions and the preposition distribution in them.

Semantic Class	Synset category	Prepositions
locative	prep.location	в (във), връз, всред, въз, върху, до, за, зад, из, извън, иззад, измежду, изпод, край, към, между, на, над, накрай, насред, низ, о, около, от, отвъд, откъм, отсам, оттам, отгатак, по, под, подир, подире, покрай, помежду, посред, пред, през, при, против, пряко, след, спроти, сред, срещу, у
temporal	prep.time	в (във), всред, до, за, край, към, между, на, накрай, насред, около, от, по, подир, подире, покрай, помежду, посред, пред, преди, през, при, с (със) след, спроти, сред, срещу, у
manner and instrument of action	prep.manner	без, в (във), като, на, по, под, посредством, с (със), според, чрез
cause	prep.cause	за, заради, от, оттам, по, поради, пред
purpose	prep.purpose	до, за, заради, към, поради
possession	prep.possession	на, от, с (със), у
origin and part of a whole	prep.origin	в (във), на, от
quantitative, degree and exceeding of a limit	prep.quantity	до, за, към, между, на, над, около, от, по, под, с (със), около, свръх
exchange	prep.obj.exchange	вместо, за, заради, наместо, срещу, спроти
exclusion	prep.obj.exclusion	без, извън, освен
opinion	prep.obj.opinion	за, според
thought	prep.obj.thought	върху, връз, въз, до, за, заради, към, над, около, по, спрямо
transition	prep.transition	в (във), на, от
comparison	prep.comparison	като
opposition	prep.opposition	въпреки, против, пряко, спроти, срещу

Table 1: Semantic classes and synset categories of prepositions.

### 3.1 Parallel with Classifications for Other Languages

A comparison of the semantic classification of Bulgarian prepositions could be made with the semantic roles used in the Penn Treebank for the prepositional phrases. They determine the semantic relation which the prepositions express. In Penn Treebank there are seven types of semantic roles for prepositional phrases – beneficiary, direction, spatial extent, manner, location,

purpose/reason and temporal. Most of them correspond to the classes of the Bulgarian classification, but with several differences: there the purpose and reason classes are united in contrast with the Bulgarian; also the location relation here is divided in more fine-grained subtypes – spatial extent and location; and the direction relation here is a separate class, but in the Bulgarian categorisation it could be either in the purpose or location classes, depending on the given instance.

As mentioned above, the thematic roles-based classification of PrepNet (Saint-Dizier, 2008) is applicable for several languages, and additionally it is observed that it has many similarities with the Bulgarian classification.

It contains the following senses: localization (with subsenses source, destination, via/passage, fixed position, quantity (numerical or referential quantity, frequency and iterativity, proportion or ratio), manner (manners and attitudes, means (instrument or abstract), imitation or analogy), accompaniment (adjunction, simultaneity of events, inclusion, exclusion), choice and exchange (exchange, choice or alternative, substitution), causality (cause, goal or sequence, intention), opposition, ordering (priority, subordination, hierarchy, ranking, degree of importance) and minor groups like about, in spite of and comparison.

The generalized classifications with small number of classes, such as that of Penn Treebank, surely have benefits, as it is proved by the experiments of O'Hara and Wiebe (2003) for WSD of prepositions on the base of the semantic roles from Penn Treebank and FrameNet. The classification of FrameNet is much more fine-grained with more than 140 types of semantic roles and the results of the research show that the accuracy of the WSD is higher with the data from Penn Treebank. However, the semantic roles of FrameNet are frequently used for preposition disambiguation and processing (see Section 2).

In O'Hara and Wiebe (2003) the top 25 roles from FrameNet are sorted: speaker, message, self-mover, theme, agent, goal, path, cognizer, manner, source, content, experiencer, evaluatee, judge, topic, undefined, cause, addressee, perceptual source, phenomenon, reason, area, degree, body part, protagonist. Even though in FrameNet the roles distinguish very finely the different subsenses, similar types with the Bulgarian classification could be observed: in both of them there are manner and source (in the Bulgarian it is formulated as origin class) types; the cause class in FrameNet is divided in two – cause and reason; goal is corresponding to the purpose class; degree is included in the quantity class in the Bulgarian categorisation; area in FrameNet is a subtype of the more general location class, etc.

The classification used by Lassen (2006) also has similarities with the Bulgarian one. Some of the similar classes (concerning the same senses but rather different in formulation) that they share are temporal aspects, location, position, purpose, function, cum (for accompaniment, etc.), causes, caused by, by means of, instrument, via, comprising, has part and part of. There are also several classes regarding acts and processes (agent/patient/source/result/destination of act or process), which are not available in the Bulgarian classification.

The work of Srikumar and Roth (2013) presents an inventory of 32 preposition relations, many of which match with Bulgarian classes, such as cause, location, manner, purpose, etc. However, the relations are generally more fine-grained than in our classification, for example destination and direction classes are separate, they are not considered as a part of a broader location category. Based on the preposition relations of Srikumar and Roth (2013) and the thematic roles from VerbNet, Schneider et al. (2015) introduce a classification of 73 fine-grained preposition supersenses, which form a hierarchical taxonomy.

#### **4 Preposition Synset Model**

The main intention towards the structure and relations of the preposition synsets is that they follow the model of all the other synsets in BTB-WN (for nouns, verbs, adjectives and adverbs) as much as possible,

given the differences between them.

Preposition synsets have synset category (which shows their semantic category), detailed definition, examples, synonyms if available and relations. In BTB-WN preposition synsets have the part of speech value *p*, following the format of the Global WordNet Association<sup>10</sup> where this value is planned for adpositions.

Currently six type of relations between preposition synsets and between a preposition synset and other parts of speech are established – *synonymy*, *antonymy*, *hyponymy*, *hyponymy*, *similarity* and *semantic derivation*. The relations are mainly semantic, but also two derivational are used – *sem-derived-from-p* which links the preposition with the noun or adverbs that it is derived from and the opposite – *sem-derives-to-p*<sup>11</sup>. Additionally, it is planned to introduce relations between verbs and prepositions and nouns and prepositions that link combinations of these two parts of speech to express a given meaning.

Figure 1 presents an example for the model with the synset of върху *vărhu* “over, on” (*prep.location*) which has several synonyms, the first two being archaic (врѣз, въз *vřáz, vřáz*); one *antonymy* relation with под *pod* “under”; it has *derivational* relations with the noun връх *vřáh* “top” and the adverb сврѣх *svřáh* “above, beyond”; it has a *similar* relation with the preposition над *nad* “over, above” that also expresses a spatial position higher than something.

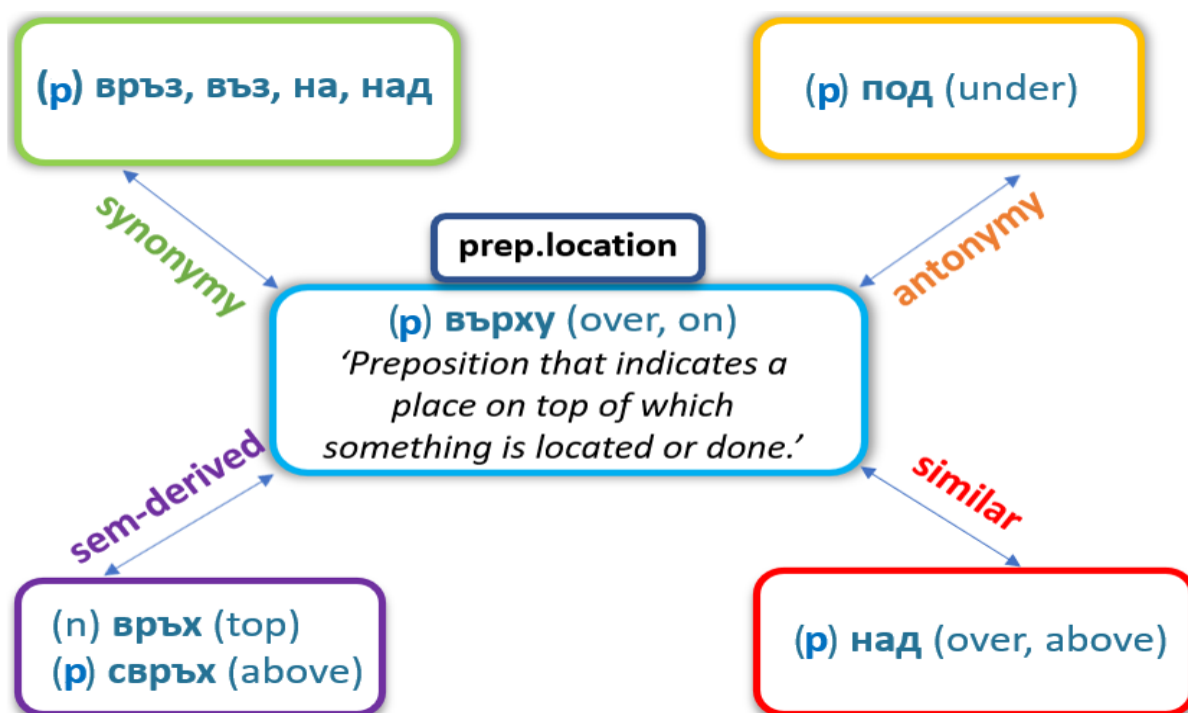


Figure 1: Example of the preposition synset model in BTB-WN

*Synonymy* is observed, for example, in the prepositions преди *predi* and пред *pred* “before” that are synonyms in their temporal sense and are united in one synset. The synset with the most synonyms – 7 – is from the manner class: в, на, по, под, с (*със*), посредством, чрез *v, na, po, pod, s (săs), posredstvom, črez* (“in”, “on”, “under”, “with”, “through”) meaning *expression of a way, form of doing, course of something*. (see Figure 2).

The *antonymy* relation links the synsets of без *bez* “without” (in the sense of *expression of lack from a part of a whole, prep.obj.exclusion*) and с *s* “with” (*expression of a belonging part or quality of something, someone, prep.possession*). An example for the *hyponymy* and *hyponymy* relations could be made with the synset for в, на, у *v, na, u* “in, at” (*expression of position in time, a moment when*

<sup>10</sup><https://globalwordnet.github.io/schemas/>

<sup>11</sup>More information about the newly introduced relations could be found in the BTB-WN Guidelines – <https://clada-bg.eu/en/centers-and-services/language-technologies/btb-wordnet.html>



Example	Lemma	ID	Order	Part of speech
Вървят @@@ в @@@ стройни редици.	в	217662	1	p
	на	217672	2	p
	по	217673	3	p
	под	217702	4	p
	с	217709	5	p
	посредством	217710	6	p
	чрез	217711	7	p
	със	217712	8	p

Figure 2: The preposition synset with the most members in BTB-WN.

*something happens*, prep.time), which is the hypernym of several narrower temporal senses such as от of “from” (*expression of initial limit in time, beginning of something*, prep.time), по, към, около по, към, около “by, around” (*expression of approximate time*, prep.time) and a few others, that are shown in Figure 3.

Example	Lemma	ID	Order	Part of speech
@@@ в @@@ 10 часа.	в	217662	1	p
	на	217672	2	p
	у	217663	3	p

Relations: **Hyponym chain** | Additional information | Tickets | Open ticket | Temporary notes

Изравяване на положение във времето, момент, отрязък или период от време, когато нещо се случва, извършва.

**hyponym**

Lemma	Definition	Part of speech	EquiID
сред,сред,насред,поср	Изравяване на момент в средата на определено време.	p	155916
по,към,около	Изравяване на приблизителност по време, приблизителен момент, в който става нещо.	p	155921
накрай	Изравяване на край на някакъв момент, период.	p	155922
от	Изравяване на начален предел във времето, започване на нещо.	p	155923
през	Изравяване на период от време, в границите на който се случва, извършва нещо.	p	155927
при	Изравяване на положение във времето спрямо други придружаващи събития, обстоятелства, на и	p	155929
срещу,спроти	Изравяване на извършване, осъществяване на нещо в навечерието на нещо друго, близост по вре	p	155931

Figure 3: Example of *hyponymy* and *hyponymy* relations in the CLaDA-BG Dict – the editing system for BTB-WN

Figure 4 shows a locative sense of в в “in” (*expression of location, place where something is, happens or is done*, prep.location). This is the preposition synset with most numerous hyponyms so far – 10.

The *similarity*<sup>12</sup> relation is established for example between до, пред, при do, pred, pri “to, in front of,

<sup>12</sup>Similarity relation is used in PWN to link closely related senses, originally only adjectives – <https://globalwordnet.github.io/gwadoc/#similar>

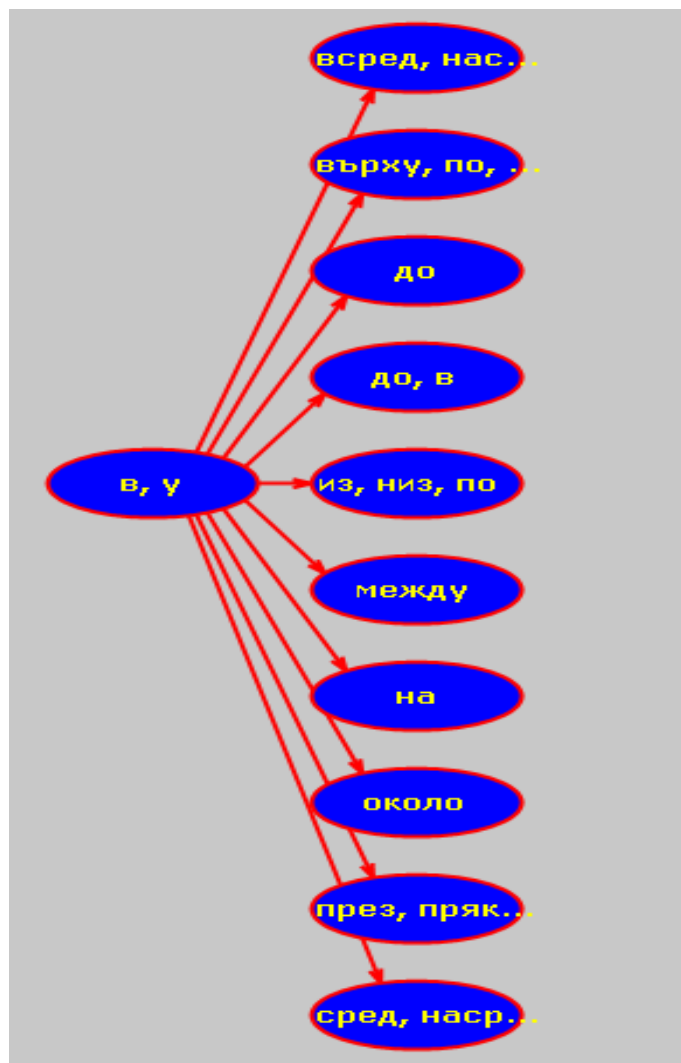


Figure 4: The preposition synset with the most hyponyms in BTB-WN – в “in”

at” (*expression of close proximity with something or someone in, during movement, prep.locative*) and зад, след *zad, sled* “behind, after” (*expression of close proximity with something or someone, prep.locative*).

The *sem-derived-from-p* and *sem-derives-to-p* are used for cases such as след *sled* “after, behind” (*prep.locative* and *prep.time*) that are derived from the noun следа *sleda* “trace” (*a visible mark left by the passage of person, animal or vehicle*) (see Figure 5).

In terms of origin Bulgarian prepositions are traditionally classified in two groups: simple and compound. The compound prepositions can be derived from nouns (typically they are related with a certain case form of the noun) or from adverbs. Derivational relations which link adjectives and nouns, adverbs and verbs and, adverbs and adjectives (*sem-derived-to/from*) were recently introduced in BTB-WN 4.0 and are now extended also for linking prepositions with related nouns and adverbs.

62 preposition lemmas were integrated in BTB-WN, forming a total of 105 synsets. The highest polysemy is found in the prepositions на *na* (most frequently could be translated as “on”, “of”, “in”, etc.) with 12 synsets, followed by по *po* (“over”, “in”, “on”, etc.) with 11 synsets. Other prepositions with multiple senses are за *za* (“for”, “to”, “about”, etc.) and от *ot* (“from”) that are part of nine synsets each; с *s* (“with”) is in eight synsets and до *do* (“to”, “until”, etc.) and в *v* (“in”, “at”) are found in seven.



Lemma

10 examples, 7 of which to lemmas

Example	Lemma	ID	Order	Part of speech
	следа	12115	1	n
Потерите се придружаваха от хрътки, при	диря	12114	2	n
	отпечатък	12116	3	n

relations    Hypernym chain    Additional information    Tickets    Open ticket    Temporary notes

Знак, оставен от преминаването на човек, животно, превозно средство и други.

**sem-derives-to-p**

Definition	Lemma	Part of speech
Изразяване на непосредствена близост при движение.	зад,след	p
Изразяване на посока на действие или движение, която следва посоката на нещо друго.	по,след,подир,подире	p
Изразяване на по-късна позиция във времето, случване, извършване на нещо по-късно от друго.	след,подир,подире	p

Figure 5: Example of *sem-derived-from-p* and *sem-derives-to-p* relations in the CLaDA-BG Dict – the editing system for BTB-WN

The category distribution of prepositions is shown in Table 1 and there it could be observed that the *locative* class has the most prepositions – 46 (which is not surprising, since this sense of the prepositions is considered to be their oldest and primary function), followed by *time* with 28 and *quantity* with 13 prepositions. The *comparison* class proves to be smallest – this sense is expressed only by one preposition: *като* *kato* “like”.

## 5 Future Work

The methods and experiments from related works on prepositions for NLP provide several beneficial directions for future work. Bulgarian prepositions are planned to be analysed on the base of the phrase types that they participate in. Data from a Bulgarian valency lexicon and the BulTreeBank will be derived. The research of Harabagiu (1996) could be considered for this task, because it provides approach for prepositional disambiguation with the use of information from wordnet (semantic relations, noun and verb categories, glosses) as well as the work of Anand Kumar et al. (2015) aimed at preposition disambiguation for machine translation, where hypernyms and lexicographer files (that contain information for POS, category, etc.) from the PWN are used.

The prepositional classification could be additionally validated with corpus analysis following the bottom-up approach for manual annotation of Lassen (2006) but instead of ontological types for the nouns in the phrases wordnet categories could be used, as in the approach of Srikumar and Roth (2013). This information for the different parts of speech in the prepositional phrases would be used for establishing more relations with the preposition synsets in BTB-WN.

Regarding the semantic relations in wordnet, more resources are planned to be consulted. To a greater extent than the grammars, explanatory dictionaries provide information for the dialectal, archaic and colloquial variants of prepositions. They also include information about synonymy and antonymy of prepositions, which could be used for establishing the corresponding relations in BTB-WN. This data is planned to be included in BTB-WN and thus to provide a more exhaustive presentation of prepositions. The different speech register variants of all the other parts of speech in BTB-WN are present, so the approach towards the prepositions would be correlating. Etymological dictionaries would be used for establishing more *sem-derived-from/to* relations with the words that prepositions are derived from.

Additionally, the more complex in structure types of prepositions are still not taken into consideration in this research, so a next task could be the processing of polyprepositional constructions and also of prepositions with grammatical functions could be explored.

## 6 Conclusion

The paper presents the initial stages of the attempt at integrating closed-class words, namely prepositions, in the BTB-WN. Wordnets usually do not include prepositions in their structure, however relevant studies provide evidence that this task is possible and beneficial.

So far for BTB-WN are developed a semantic classification of prepositions and a synset model with several semantic relations. The paper outlines the directions for analysis and further processing of prepositions. The goal of this attempt is to improve the application of BTB-WN for semantic annotation and to use it as a resource for creating Bulgarian language models. Since prepositions express relational information, they have a key role for the semantic interpretation.

## Acknowledgements

This work was supported by *the Bulgarian National Interdisciplinary Research e-Infrastructure for Resources and Technologies in favor of the Bulgarian Language and Cultural Heritage, part of the EU infrastructures CLARIN and DARIAH – CLaDA-BG*, Grant number DO01-377/18.12.2020. and the Grant No. BG05M2OP001-1.001-0003, financed by *the Science and Education for Smart Growth Operational Program (2014-2020)* and co-financed by *the European Union through the European structural and Investment funds*.

## References

- Amaro, R. 2018. Integrating Prepositions in Wordnets: Relations, Glosses and Visual Description. *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*. European Language Resources Association (ELRA), Miyazaki, Japan.
- Anand Kumar, M., Rajendran, S. and Soman, K. P. 2015. Cross-Lingual Preposition Disambiguation for Machine Translation. *Procedia Computer Science*, 54:291-300.
- Bailey, D., Lierler, Y. and Susman, B.. 2015. Prepositional Phrase Attachment Problem Revisited: how Verbnet can Help. *Proceedings of the 11th International Conference on Computational Semantics*, ACL, London, UK.
- Baldwin, T., Kordoni, V. and Villavicencio, A. 2009. Prepositions in Applications: A Survey and Introduction to the Special Issue. *Computational Linguistics*, 35(2):119-149. MIT Press, Cambridge, MA.
- Bond, F. and Da Costa, L. M. 2016. Wow! What a Useful Extension! Introducing Non-Referential Concepts to Wordnet *Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16)*. European Language Resources Association (ELRA), Portorož, Slovenia.
- Boyadzhiev, T., Kutsarov, I., and Penchev, Y. 1998. Contemporary Bulgarian language. Phonetics, lexicology, word formation, morphology, syntax. Petar Beron, Sofia, Bulgaria.
- Fellbaum, C. 1998. WordNet: An Electronic Lexical Database. MIT Press, Cambridge, MA.
- Fillmore, J. C., Wooters, C. and Baker, C. F. 2001. Building a Large Lexical Databank Which Provides Deep Semantics. *Proceedings of the 15th Pacific Asia Conference on Language, Information and Computation*, City University of Hong Kong, Hong Kong, China.
- Harabagiu, S. M. 1996. An application of WordNet to prepositional attachment. *Proceedings of the 34th annual meeting on Association for Computational Linguistics (ACL 1996)*. Association for Computational Linguistics, Santa Cruz, USA.
- Konstantinova, V. 1982. Prepositions in Bulgarian grammar literature. Publishing house of BAS, Sofia, Bulgaria.
- Lassen, T. 2006. An Ontology-based View on Prepositional Senses. *Proceedings of the Third ACL-SIGSEM Workshop on Prepositions*. ACL, Trento, Italy.

- Litkowski, K. and Hargraves, O. 2005. The Preposition Project. *ACL-SIGSEM Workshop on "The Linguistic Dimensions of Prepositions and Their Use in Computational Linguistic Formalisms and Applications"*. Colchester, England.
- Marcus, M., Kim, G., Marcinkiewicz, M., MacIntyre, R., Bies, A., Ferguson, M., Katz, K. and Schasberger, B. 1994. The Penn Treebank: Annotating predicate argument structure. *Proceedings of the ARPA Human Language Technology Workshop*, Plainsboro, NJ.
- McCrae, J. P., Rudnicka, E. and Bond, F. 2020. English WordNet: A new open-source WordNet for English. *K Lexical News*, 28:37-44.
- O'Hara, T. and Wiebe, J. 2009. Exploiting Semantic Role Resources for Preposition Disambiguation. *Computational Linguistics*, 35:151–184. MIT Press, Cambridge, MA.
- O'Hara, T. and Wiebe, J. 2003. Preposition Semantic Classification via Penn Treebank and FrameNet. *Proceedings of the Seventh Conference on Natural Language Learning at HLT-NAACL 2003 - Volume 4*, ACL, Edmonton, Canada.
- Osenova, P. and Simov, K. 2018. The data-driven Bulgarian WordNet: BTBWN *Cognitive Studies – Études cognitives*, volume 2018(18). Institute of Slavic Studies, Polish Academy of Sciences, Warsaw, Poland.
- Prasad, R., Webber, B., Lee, A. and Joshi, A. 2019. The Penn Discourse Treebank 3.0 Annotation Manual. *Linguistic Data Consortium*, Philadelphia, USA.
- Saint-Dizier, P. 2008. Syntactic and Semantic Frames in PrepNet. *Proceedings of the Third International Joint Conference on Natural Language Processing*, volume II. Asian Federation of Natural Language Processing.
- Schneider, N., Srikumar, V., Hwang, J.D., and Palmer, M. 2015. A Hierarchy with, of, and for Preposition Supersenses. *Proceedings of the 9th Linguistic Annotation Workshop*, ACL, Denver, Colorado, USA.
- Srikumar, V. and Roth, D. 2013. Modeling Semantic Relations Expressed by Prepositions. *Transactions of the Association for Computational Linguistics*, 1:231–242, MIT Press, Cambridge, MA.
- Stoaynov, S. 1983. Grammar of contemporary Bulgarian standard language. Morphology. Publishing house of BAS, Sofia, Bulgaria.