

Modeling MWEs in BTB-WN*

Laska Laskova, Petya Osenova, Kiril Simov, Ivajlo Radev, Zara Kancheva

IICT-BAS

Sofia, Bulgaria

{laska|petya|kivs|radev|zara}@bultreebank.org

Abstract

The paper presents the characteristics of the predominant types of MultiWord expressions (MWEs) in the BulTreeBank WordNet – BTB-WN. Their distribution in BTB-WN is discussed with respect to the overall hierarchical organization of the lexical resource. Also, a catena-based modeling is proposed for handling the issues of lexical semantics of MWEs.

1 Introduction

In this paper we present the distribution and treatment of MultiWord Expressions (MWEs) within BTB-WN — a data-driven Bulgarian WordNet.¹ Currently BTB-WN contains about 22 000 synsets covering CoreWordNet synsets, all the content words within BulTreeBank (about 8 000 lemmas) and the top part of a frequency list over 70 million running words. For the purpose of this work we use two subsets: (1) the current version of BTB-WN; and (2) a subset mapped to the Bulgarian Wikipedia in order to establish a connection between the lexical information in BTB-WN and the encyclopedic knowledge — (Simov et al., 2019). The second set is used to evaluate the impact of the encyclopedic domain on the distribution of the MWEs. From the first subset 981 examples of MWEs have been extracted, while from the second one – 506 examples.

In the past few years extensive literature has been dedicated to MWEs. In spite of that there is no single guiding principle or widely accepted classification, since MWEs are not homogeneous and can be classified at different levels that interact in various ways: morphology, lexicology, syntax, and semantics. Also, the typology becomes more

complex at a cross-language level due to the differing approaches to MWEs and differing language systems. For that reason we rely on the classification² developed within WG 4 of PARSEME COST Action.³ This classification takes into account the part-of-speech of the MWE head which in our view is suitable for the treatment of MWEs in wordnets. Thus, it categorises the MWEs into the following types: Nominal (Named Entities, NN compounds, other), Verbal (phrasal verbs, light verb constructions, VP idioms, other), Adjectival, Prepositional and Other.

We focus on modeling compositionality of MWEs as reflected in their morphosyntactic and semantic properties. With respect to semantics we follow (Bentivogli and Pianta, 2004) in requiring the representation of both types of meanings – a) related to the whole MWE and b) related to its constituent words. Such an approach is especially important for cases when the MWE allows also a fully compositional usage. For example, the classical MWE “kick the bucket” comprises an idiomatic meaning, but in an appropriate context it might have also a compositional (literal) usage. We differ from the above mentioned authors, since we do not introduce a new relation for handling compositionality *composed-of*, but directly annotate the corresponding words within the MWE with their literal meaning.⁴ As a modelling device for these MWEs we extend the catena framework of (Osenova and Simov, 2018b), since this approach can handle the morphosyntactic behaviour as well as the compositionality issues sourcing from semantics. The novelty here is that the focus is put on the incorporation of the lexical meaning

*Laska Laskova and Petya Osenova are also affiliated at Sofia University "St. Kl. Ohridski", Faculty of Slavic Studies.

¹For more information on the creation and development of BTB-WN see (Osenova and Simov, 2018a).

²http://clarino.uib.no/iness/page?page-id=MWEs_in_Parseme

³<https://typo.uni-konstanz.de/parseme/>

⁴We do not presuppose a definition for a literal meaning so that the annotators can decide themselves in each case.

coming from the WordNet into the catena model.

The structure of the paper is as follows: the next section outlines the related work; Section 3 presents a classification of the MWEs in BTB-WN. Section 4 proposes an extension of the catena model that incorporates lexical semantics. Section 5 concludes the paper.

2 Related Work

(Constant et al., 2017) elaborate on the diversity of the MWEs and the schemes for their categorization. The article lists the most commonly seen categories of MWEs: idioms; light-verb constructions; verb-particle constructions; noun and verb compounds; complex function words; multiword named entities and multiword terms. The authors note that these categories are non-exhaustive and can overlap. Recently, the work on identification of MWEs continued with a focus on Verbal MWEs. The 2018 edition of the shared task PARSEME (Ramisch et al., 2018) relied on enhanced and revised guidelines defining the following verbal MWEs typology: light-verb constructions; verbal idioms; inherently reflexive verbs; verb-particle constructions; multi-verb constructions; inherently clitic verbs; inherently adpositional verbs. Here we do not go into such a detailed typology, thus relying on the more general verbal classification from the PARSEME WG 4 presented briefly above.

As already mentioned, our approach is similar to the one proposed in (Bentivogli and Pianta, 2004). They consider the addition of syntagmatic information to WordNet by providing co-occurrences of meanings within a MWE. In order to do this they related each noun, verb, adjective or adverb in a given MWE with the appropriate synset via the new relation `composed-of`. In addition to MWEs the authors proposed to include Recurrent Free Phrases in WordNet that are completely compositional, but have some additional features that distinguish them from the arbitrary compositional phrases. These features might source from additional knowledge carried by the phrase, or from statistically idiosyncratic patterns. The grouping of the phrases by their meaning has been called a *phrasets*. The *phrasets* are useful not only by providing co-occurrences of meanings for their constituent words, but also in multilingual settings where they might fill lexical gaps. Also they are useful in NLP tasks such as Word Sense

Disambiguation, Machine Translation, etc. A similar approach to MWEs has been undertaken also in the creation of the Basque WordNet — (Agirre et al., 2006). As already mentioned above, we do not introduce a new relation, but directly annotate the words in MWEs with the appropriate literal meanings. Furthermore, we do not restrict the annotation only to the compositional parts of MWEs. Whole MWEs are annotated as well. For the moment no *phrasets* have been added to BTB-WN, but we consider such a step as a good development in future.

In series of papers (Simov and Osenova, 2014), (Simov and Osenova, 2015), and (Osenova and Simov, 2018b) we presented the modeling of MWEs in terms of catena. These papers demonstrate how the (partial) variability and compositionality can be represented. The last paper reflects the multilingual application of the model. In our work here we extend this model to represent also the literal meanings of the distinct components in MWEs.

3 MWE types in BTB-WN

The classification we present preserves the general grouping of synsets into the four syntactic types which can be found in Princeton WordNet (PWN) and other wordnets alike: nominal, verbal, adjectival and adverbial ones. All prepositional MWEs are classified either as adjectival or as adverbial MWE. It is worth noting that both phrases – PP and, less often, AdvP – can be modifiers or adjuncts depending on the context. For example, `от първа ръка` (“first-hand”) can modify the verb `знам` (“know”) (i.e. I know something from first-hand) and the noun `информация` (“information”) (i.e. I have information from first-hand) which denotes one of the components involved in the situation described by the verb.

We examine each of the four subsets for recurring syntactic patterns and evaluate them in terms of semantic compositionality, grammatical deviation (archaic, morphologically frozen forms included), and flexibility — the last one understood as a complex feature that takes into account morphological variation, word order permutation, and the possibility to modify the sub-units of a MWE.

It is assumed that MWEs exhibiting the degree of compositionality and flexibility typical for phrases generated *ad hoc* in discourse, should still be included in the lexicon if they are as-

sociated with a particular type of genre, speech act or otherwise conventionalized (Calzolari et al., 2002). One such example is the terminological unit промяна на климата (change of climate, ‘climate change’), which corresponds to the two MWE forms in the PWN synset {*climate change*, *global climate change*}. A small number of the two or three word sequences extracted from BTB-WN appeared to be marginal for the MWE spectrum. They were born in the process of the bidirectional mapping of BTB-WN and PWN synsets (Simov et al., 2015) as instances of periphrastic translation; whenever there is no word or MWE in the target language to express the concept, dictionaries offer descriptive phrases whose length and syntactic level of complexity may vary. Consider these two examples: the VP гледам гневно of the structural type V + Adv, ‘look disapprovingly’ which is mapped to the PWN synset {*glower*, *glare*} or the four-word sequence казвам|изговарям буква по буква, lit. ‘say/pronounce letter per letter’ which is used to translate the English verb “spell”. While the meaning of both Bulgarian expressions can be derived from the meaning of their sub-units, the latter is a collocation, and the former is not.

Type	Number	%
Total: 981		
named entities	117	11.93
A + N	565	57.59
N + PP	79	8.05
N + N	25	2.54
A + N + PP	2	0.20
N + Conj + N	5	0.51
V	124	12.64
PP	27	2.75
Adj	12	1.22
other	25	2.54

Table 1: MWE types distribution in BTB-WN.

The distribution of the various structural types within BTB-WN resource shows a slightly bigger share of the patterns which do not have an N for a head and the third most numerous group is that of the verbal MWEs. (see Table 1). Only 2 of the 25 compound noun phrases have not been matched to a Wikipedia article. In contrast to English, the NN pattern in Bulgarian is not only rare, but it is reserved for terms in which at least one of the sub-units has reduced its semantic transparency, as in

елен лопатар, “fallow deer”, or is foreign web страница, “web page”.

Type	Number	%
Total: 506		
named entities	110	21.74
A⁵ + N	333	65.81
N + PP	26	5.14
N + N	23	4.55
N + Conj + N	3	0.59
V	4	0.79
PP	5	1.00
AdvP	2	0.36

Table 2: MWEs type distribution in BTB-WN aligned to the Bulgarian Wikipedia.

In comparison, Table 2 presents the percentage of the different structural types of MWEs within the synsets mapped to Wikipedia articles in an initial attempt to enrich BTB-WN with encyclopedic knowledge. Not surprisingly, the first three most numerous groups are all nominal, which stems from the fact that Wikipedia articles mainly cover general concepts and named entities (McCrae, 2018).

In the following subsections a more detailed description for each MWE type is provided.

3.1 Multiword Adverbials

With a few exceptions, all of the examined prepositional phrases are adverbial adjuncts corresponding to a Prep(ositional) head followed by a post-modifier N(oun) or Adv(erb); in some cases the second element is modified by another PP or Adj(ective) — see Table 3.

The opposite, however, is not true — some of the adverbial adjuncts follow different syntactic patterns which often but not always have phonetic, rhythmic and/or lexical repetition as their common denominator. This feature is related to the iconicity that reflects the meaning of the MWE (e.g. examples 9 and 10, and especially 8 where the two sub-units are nonsensical if not concatenated, which is to say that they do not have a lemma status on their own).

Example 9 – Сегиз-тогиз (lit. “now-then”) – and example 10 – напред-назад (lit. “forth-back”) – represent the result of a type of syntactic contraction where the conjunction is omitted. Example 13 represents another typical syntactic transformation that accompanies the process

No	Expressions	Gloss	Translation	Pattern
1	за жалост	'for pity'	unfortunately	Prep + N
2	на ръка	'on hand'	manually	Prep + N
3	у дома	'in home-LOC'	at home	Prep + N
4	де факто	'de facto'	de facto	Prep + N (foreign)
5	за малко	'for little-SG.N'	for a while	Prep + Adj
6	по човешки	'on human-SG.M'	humanly	Prep + Adj
7	от време на време	'from time to time'	—”—	PP + PP
8	чат-пат	xx-xx	from time to time	opaque
9	сегиз-тогиз	'now-then'	from time to time	Adv Adv
10	напред-назад	'back-forth'	back and forth	Adv Adv
11	очи в очи	'eye-PL in eye-PL'	face to face	N + Prep + N
12	живот и здраве	'life and health'	hopefully	N + Conj + N
13	известно време	'known time'	for a while	Adj N

Table 3: Multiword adverbials and their syntactic patterns.

of lexicalization, i.e. the omission of a preposition (известно време (certain time) vs. за известно време (for certain time). Examples 5 and 6 in Table 3 illustrate two productive derivational models and consequently — a predictable multiword time and manner adverbial constructions, where a fixed preposition (за, “for” or по, “on”) is followed by an adjective which in turn has to be semantically and grammatically compatible with the elliptical head noun (време, “time” and начин, “manner”, respectively). Neither the MWEs with a prepositional head, nor any of the adverbials show any degree of morphosyntactic variation. All of them have a fixed word order.

3.2 Multiword Adjectives

There are only three MWEs of the PP modifier type in BTB-WN, на високо равнище, на високо ниво, “top-level”, and от първа ръка, “first-hand” which belong to two different synsets. The rest of the modifiers have as their head a syntactic Adjective (see Table 4), and it is the only subunit subject to morphological modification. Interesting cases are the following ones: example 14 рохко сварен, “soft-boiled” and example 15 добре дошъл, “welcome”. The former represents an interesting example of a MWE that has a limited selective power, since it typically collocates with the neutral noun яйце, “egg”. This respectively narrows down the possible morphological realizations to two forms, рохко сварен-о ‘soft-boiled-SG.N’ or рохко сварен-и ‘soft-boiled-PL’. The latter is usually predicatively used and referring to some person. Thus its form depends on the

gender of the referred person and on the singularity/plurality of these objects. Again, the order of the adjectival MWE elements is fixed.

3.3 Multiword Verbs

The majority of verbal MWEs contain at least one reflexive *se-* or *si-*verb (отморявам си, отдъхна си, relax), or dative/accusative clitics. 93.83 % of all verbal MWEs in BTB-WN are of this kind. Although they are often mapped to English phrasal verbs in translation (Kordoni and Simova, 2014), we do not consider reflexive verbs and verbs that include accusative or dative pronominal particles, such as унася ме, “doze off” or хрумва ми, “come to mind” as MWEs (for a different approach see (Ramisch et al., 2018)). Thus, there are only 124 multiword verbs *per se*.

Among these 124 verbal phrases we distinguished several syntactic patterns as illustrated in Table 5. Examples 18 and 19 illustrate the light verb construction, with правя, “make” and водя, “lead” as phrasal heads respectively. Another frequent light verb in BTB-WN is давам, “give”. Typically, light verb MWEs are found in synsets with verbs that are derived from the nominal sub-unit, e.g. (правя) гаргар-а → гаргар-я се (make a gargle, to gargle) or vice versa, e.g. кореспондирам → (водя) кореспонденция (to correspond, correspondence). In these cases the two synsets are {правя гаргара, гаргар-я се} and {водя кореспонденция, кореспондирам}. Examples 20 and 24 belong to different structural types but they have one thing in common, a sub-unit that refers to a body part,

No	Expressions	Gloss	Translation	Pattern
14	рохко сварен	'soft boiled'	soft-boiled	Adv + Adj
15	добре дошъл	'well come'	welcome	Adv + Adj
16	труден за разбиране	'difficult for understanding'	baffling	Adj + PP
17	загубил надежда	'lost hope'	desperate	Adj + N

Table 4: Multiword adjectives and their syntactic patterns.

No	Expressions	Gloss	Translation	Pattern
18	правя гаргара	'make gargle'	gargle	V + N
19	вода кореспонденция	'lead correspondence'	communicate	V + N
20	затварям си устата	'close PTCL.REFL.POSS the-mouth'	shut up	V + N
21	завършвам наравно	'finish equally'	draw a game	V + Adv
23	карам с превишена скорост	'drive with exceeded speed'	speed	V + PP
24	говоря под носа си	'speak under nose self'	mumble	V + PP

Table 5: Multiword verbs and their syntactic patterns.

accompanied by the reflexive possessive marker *si*. In example 20 the body part is 'mouth' while in example 24 it is 'nose'. Even when they are used in a sentence with a plural subject, the number of the noun element typically remains singular, e.g. Затваря-йте си уст-а.та!, 'Shut-IMP.2PL PTCL.REFL.POSS mouth-SG.DET'. The verbal MWEs might allow for adjectival modification of their noun elements.

3.4 Multiword Nouns

This type reflects predominantly named entities and specialized terminological units or everyday idiomatic phrases. Thus, they can be highly recursive in structure.

In Table 6 the named entity types are presented. Examples 25–28 list patterns of person names. It can be seen that the Bulgarian⁶ pattern of a proper noun plus adjectival middle and/or family name is listed in examples 27–28, while in examples 25–26 the foreign tradition is illustrated of noun-noun phrases. The names of people are not very frequent in BTB-WN. They are included on the basis of mappings to English WordNet. In future, we are planning to extend the coverage of named entities through Wikipedia and other similar sources.

Examples 29–34 demonstrate patterns for geographical names. Here the patterns are more diverse structurally. The pattern 'adjective(s) plus noun' seems to be regular (examples 29 and 30). Also, the pattern 'noun plus (adjective) noun' (ex-

amples 31 and 32) and the pattern 'noun plus prepositional phrase' can be distinguished (example 33). Not surprisingly, there are some names that are opaque to the Bulgarian morphosyntax (example 34). From the point of view of the annotation with literal meanings the non-opaque cases require special attention because of the usage of common words in them. Components like "strip" (example 29), "dead" (example 30), and "new" (example 32) need to be annotated with the appropriate meanings. If we consider "New South Wales" and "New York", the adjective "new" needs to be annotated with two different meanings in the two cases — recently discovered and recently created.

Examples 35–37 illustrate the organization names. The observed patterns are: 'noun plus prepositional phrase' (example 35) and 'adjective(s) plus noun plus (prepositional phrase)' (examples 36 and 37). These names are included in BTB-WN because of the mapping to the PWN. Since the organization names could be quite complex, a special (chunk) grammar will be required to deal with them. The grammar would include rules for annotating the literal meanings of the MWE components.

In Table 7 the terms and the everyday idiomatic phrases are listed. Here the most frequent structural types are: 'noun plus noun' (examples 38–39); 'adjective plus noun' (examples 40–45) and 'noun plus prepositional phrase' (examples 46–47). Most of the examples are compositional. Example 39 demonstrates a very productive compo-

⁶Also in some other Slavonic languages.

No	Expressions	Translation	Pattern
25	Франклин Делано Рузвелт	Franklin Delano Roosevelt	N + N + N
26	Франклин Рузвелт	Franklin Roosevelt	N + N
27	Никита Сергеевич Хрушчов	Nikita Sergeevich Khrushchev	N + A + A
28	Никита Хрушчов	Nikita Khrushchev	N + A
29	Ивицата Газа	Gaza Strip	N + N
30	Мъртво море	Dead Sea	A + N
31	Република Южна Африка	South African Republic	N + (A + N)
32	Нов Южен Уелс	New South Wales	A + (A + N)
33	Стратфорд на Ейвън	Stratford-upon-Avon	N + PP
34	Буенос Айрес	Buenos Aires	opaque
35	Общество на народите	League of Nations	N + PP
36	Европейска централна банка	European Central Bank	A + N
37	Държавен департамент на САЩ	U.S. Department of State	A + N + PP

Table 6: Multiword named entities and their syntactic patterns.

No	Expressions	Translation	Pattern
38	сокол скитник	peregrine falcon	N + N
39	вагон-ресторант	dining car	N + N
40	детско креватче	baby bed	A + N
41	рожден ден	birthday	A + N
42	врабчови чревца	chickweed	A + N
43	златна среда	golden mean	A + N
44	добър вечер	good evening	A + N
45	пирова победа	Pyrrhic victory	A + N
46	връх на стрела	arrowhead	N + PP
47	черешката на тортата	icing on the cake	N + PP

Table 7: Other nominal MWE and their syntactic patterns.

sitional model of noun-noun phrases in Bulgarian. Examples of this structural type are included in BTB-WN because they have specific features as mentioned in the related work — additional world knowledge associated with them or statistical idiosyncratic usage. Similar cases are also examples 40, 41 and 46. They are fully compositional, but have been included because otherwise there would exist a lexical gap with respect to the Princeton WordNet. The examples illustrate *phrasets* in BTB-WN. Examples 38 and 42 are respectively names of a bird and a plant. The meaning of their components becomes clearer only when additional knowledge about the bird and the plant are considered.

The figurative (non-compositional) meaning is displayed in examples 43, 44, 45 and 47. Example 44 is a diachronic one, but very actively used in contemporary Bulgarian as a polite greeting. The other three are idiomatic expressions. In examples 43 and 45 the head nouns determine the whole

meaning of the phrases — “mean” and “victory” — but the meaning of the whole MWEs is not compositional because of the missing appropriate meanings for the adjectives. We do not want to include such meanings as separate synsets because of their limited distribution and thus the risk of introducing unnecessary ambiguity.

The presented examples in this section demonstrate a great diversity with respect to their morphosyntactic, syntactic and semantic characteristics. In the majority of the cases it seems that the literal meanings of the constituent words of the MWEs are transparent. This allows for an easy interpretation of the literal meanings within the appropriate context. Even when the MWEs are highly idiomatic, there might exist a context in which the speaker would refer to the literal meaning of the constituent words.

4 Formal Treatment of MultiWord Expressions in BTB-WN

In order to address the possible variations of the MWE elements, their potential for morphosyntactic variation and modification, and the lexical meaning of the MWE being compositional or not, we rely on the notion of *catena*.

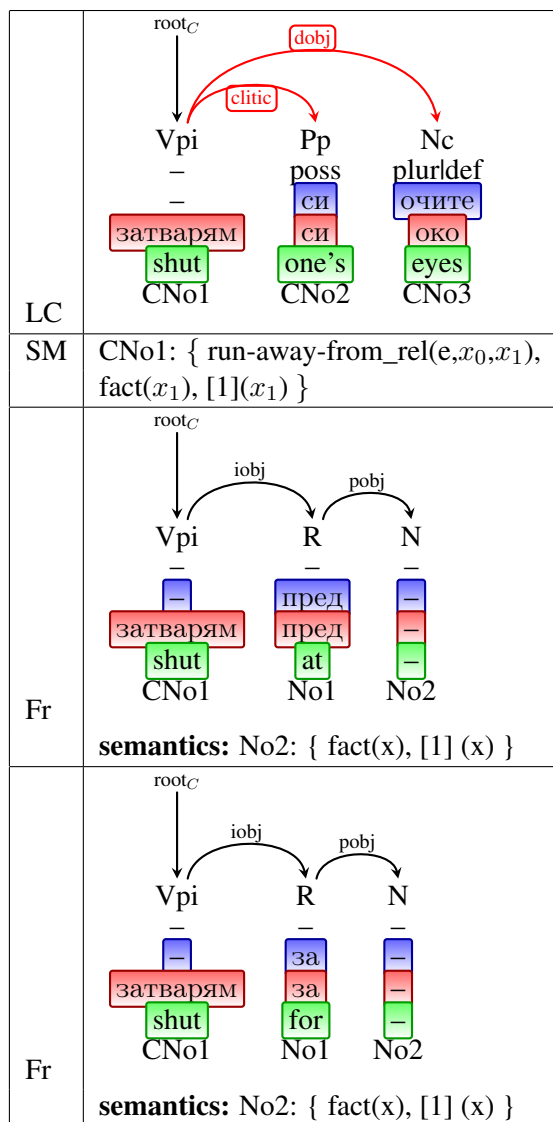


Figure 1: Lexical entry for затварям си очите, “zatvaryam si ochite”, ‘I close my eyes’.

Let us recall that the notion of catena (chain) was initially introduced in (O’Grady, 1998) as a mechanism for representing the syntactic structure of idioms. He showed that for this task a definition of syntactic patterns was needed that does not coincide with constituents. He defined 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 our work here we convert MWEs into a representation previously defined in (Simov and Osenova, 2014) and (Simov and Osenova, 2015) in which the catena is depicted as a dependency tree fragment with appropriate grammatical and semantic information. The variations of the MWEs are represented through underspecifying the corresponding features, including valency frames, non-canonical basic form.

The lexical entry uses the following format: a **lexicon-catena** (LC), **semantics** (SM) and **valency** (Frame). The lexicon-catena for the MWEs is stored in its basic 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. 1.⁷

The valency frame contains two alternative elements for indirect object introduced by two 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 catenae for the frame and for the whole lexical entry are unified on the basis of nodes with the same names.

In order to record the meaning of the whole MWE and the literal meanings of its constituent words, we extend the above lexical entry in the following way: The meaning of the whole MWE is recorded within the field SM as an additional item. In the case when the predicate semantics (as in the example) is available, it includes more than one predicate — one for the meaning of the MWE and one or more for the “assumed” arguments.

For the literal meanings of the constituent words

⁷The grammatical features are: ‘poss’ for possessive pronoun, ‘plur’ for plural number and ‘def’ for definite noun.

we include a new field called *constituent word literal meanings* (LM). In Fig. 2 one example is provided of the new lexical entry for the MWE затварям си устата (close one’s mouth-the) “shut up”. For the mapping to the synsets we use the corresponding meaning from the Princeton WordNet 3.1: **shut_up%2:30:00::** — “cause to be quiet or not talk”; **shut%2:35:00::** — “move so that an opening or passage is obstructed; make shut”; and **mouth%1:08:01::** — “the opening through which food is taken in and vocalizations emerge”. Thus, through the WordNet mappings both – figurative and compositional – meanings are provided.

LC	
SM	CNo1: { shut-up_rel(e,x ₀) } Synset: shut_up%2:30:00::
Fr	
LM	затварям : shut%2:35:00:: уста : mouth%1:08:01::

Figure 2: Lexical entry for затварям си устата (close one’s the-mouth) “shut up”.

5 Conclusions

In this paper we presented the typology and the characteristics of the MWEs in BTB-WN. Near 400 Bulgarian MWEs were encoded as lexical entries based on the catena model. This fact shows that the model is feasible not only for modeling Bulgarian MWEs but also for describing MWEs in other languages.

The approach that was taken in this work reflects the intuition of the human annotators to assign literal meanings to the constituent elements in MWEs even when they are highly idiomatic. In our work up to here no examples were found where one or more of the elements lack a literal meaning.

A more balanced and incremental view on the compositionality has been introduced since language is highly generative and might provide also contexts in which some of the literal meanings

were triggered. An open question is the handling of ambiguity when the respective element has more than one literal meaning.

Acknowledgments

This research was funded by the Bulgarian National Science Fund grant number 02/12/2016 — *Deep Models of Semantic Knowledge (DemoSem)*. The contribution of Ivajlo Radev and Zara Kancheva has been partially supported by the Bulgarian Ministry of Education and Science under the National Research Programme “Young scientists and postdoctoral students” approved by DCM # 577 / 17.08.2018. We are grateful to the anonymous reviewers for their remarks, comments, and suggestions. All errors remain our own responsibility.

References

- Eneko Agirre, Izaskun Aldezabal, and Eli Pociello. 2006. Lexicalization and Multiword Expressions in the Basque WordNet. In *Proceedings of Third International WordNet Conference*, pages 131–138.
- Luisa Bentivogli and Emanuele Pianta. 2004. Extending Wordnet with Syntagmatic Information. In *Proceedings of the 2nd Global WordNet Conference*, pages 47–53.
- Nicoletta Calzolari, Charles J. Fillmore, Ralph Grishman, Nancy Ide, Alessandro Lenci, Catherine MacLeod, and Antonio Zampolli. 2002. *Towards Best Practice for Multiword Expressions in Computational Lexicons*. In *Proceedings of the Third International Conference on Language Resources and Evaluation (LREC’02)*, Las Palmas, Canary Islands - Spain. European Language Resources Association (ELRA).
- Mathieu Constant, Gülşen Eryiğit, Johanna Monti, Lonneke van der Plas, Carlos Ramisch, Michael Rosner, and Amalia Todirascu. 2017. *Multiword expression processing: A survey*. *Computational Linguistics*, 43(4):837–892.
- Valia Kordoni and Iliana Simova. 2014. Multiword Expressions in Machine Translation. In *LREC*, pages 1208–1211.
- John P. McCrae. 2018. Mapping WordNet Instances to Wikipedia. In *Proceedings of the 9th Global WordNet Conference.*, pages 62–69.
- William O’Grady. 1998. The Syntax of Idioms. *Natural Language and Linguistic Theory*, 16:279–312.
- Petya Osenova and Kiril Simov. 2018a. *The datadriven Bulgarian WordNet: Btbwn*. *Cognitive Studies / Études cognitives*, 18.

- Petya Osenova and Kiril Simov. 2018b. [Modelling Multiword Expressions in a Parallel Bulgarian-English Newsmedia Corpus](#). *Multiword expressions: Insights from a multi-lingual perspective*, pages 247—269.
- Carlos Ramisch, Silvio Ricardo Cordeiro, Agata Savary, Veronika Vincze, Verginica Barbu Mititelu, Archana Bhatia, Maja Buljan, Marie Candito, Polona Gantar, Voula Giouli, Tunga Gungor, Abdelati Hawwari, Uxoa Inurrieta, Jolanta Kovalevskaite, Simon Krek, Timm Lichte, Chaya Liebeskind, Johanna Monti, Carla Parra Escartin, Behrang QasemiZadeh, Renata Ramisch, Nathan Schneider, Ivelina Stoyanova, Ashwini Vaidya, and Abigail Walsh. 2018. Edition 1.1 of the Parseme SharedTask on Automatic Identification of Verbal Multiword Expressions. In *Proceedings of the Joint Workshop on Linguistic Annotation, Multiword Expressions and Construction*, pages 222—240.
- Kiril Simov and Petya Osenova. 2014. 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)*, pages 198–207.
- Kiril Simov and Petya Osenova. 2015. [Catena Operations for Unified Dependency Analysis](#). In *Proceedings of the Third International Conference on Dependency Linguistics (Depling 2015)*, pages 320–329, Uppsala, Sweden. Uppsala University, Uppsala, Sweden.
- Kiril Simov, Petya Osenova, Laska Laskova, Ivajlo Radev, and Zara Kancheva. 2019. Aligning the Bulgarian BTB WordNet with the Bulgarian Wikipedia. In *Proceedings of the 10th Global WordNet Conference*.
- Kiril Simov, Alexander Popov, and Petya Osenova. 2015. Improving Word Sense Disambiguation with Linguistic Knowledge from a Sense Annotated Treebank. In *Proceedings of the International Conference Recent Advances in Natural Language Processing*, pages 596–603.