sciendo
DOI 10.2478/jazcas-2021-0035

TOWARDS CLASSIFICATION OF STATIVE VERBS IN VIEW OF CORPUS DATA

SVETLOZARA LESEVA – IVELINA STOYANOVA
– HRISTINA KUKOVA
Institute for Bulgarian Language Prof. Lyubomir Andreychin, Bulgarian Academy of Sciences, Sofia, Bulgaria

LESEVA, Svetlozara – STOYANOVA, Ivelina – KUKOVA, Hristina: Towards classification of stative verbs in view of corpus data. Journal of Linguistics, 2021, Vol. 72, No 2, pp. 383 – 393.

Abstract: The paper presents work in progress on the compilation and automatic annotation of a dataset comprising examples of stative verbs in parallel Bulgarian-Russian corpora with the goal of facilitating the elaboration of a classification of stative verbs in the two languages based on their lexical and semantic properties. We extract stative verbs from the Bulgarian and the Russian WordNets with their assigned conceptual information (frames) from FrameNet. We then assign the set of probable Bulgarian and Russian stative verbs to the verb instances in a parallel Bulgarian-Russian corpus using WordNet correspondences to filter out unlikely stative candidates. Further, manual inspection will ensure high quality of the resource and its application for the purposes of semantic analysis.
Keywords: stative verbs, parallel corpora, semantic annotation

1 MOTIVATION AND TASK OVERVIEW

Despite the advances in the creation of ever larger corpora, parallel or comparable corpora for specific pairs of languages may still be scarce, especially ones with task-specific labelling. In this paper, we describe a methodology for compiling and annotating a parallel corpus for two Slavic languages, Bulgarian and Russian, tailored to a specific linguistic task: contrastive description of stative verbs.

1.1 Predefined vs. resource-driven classification of stative verbs

Vendler's aspectual classification of verbs into activities, states, achievements and accomplishments [1] subsequently developed and elaborated by Dowty [2] and Van Valin and LaPolla [3], among many others, has provided deep insights into the aspectual nature of situations and predicates. There have been other proposals for classifications according to semantic classes that usually take into account the aspectual class: two such accounts are Paducheva's [4] and Van Valin and Lapolla's [3] classifications. A cursory look at the representation of stative verbs (which we deal with) shows that at certain points (such as predicates of emotion, cognition,

desire), the classifications show substantial similarities, while at others they show different levels of granularity or employ different classes altogether. Table 1 provides a juxtaposition between several very similarly treated classes. We use the original examples in the relevant works.

Paducheva (1996)	Van Valin and LaPolla (1997)
Intention and will: желать 'wish' жаждать 'crave', надеяться 'hope', стремиться *'strive, aspire'*	**Desire**: *want*, *wish*
Temporary emotional states: беспокоиться 'worry', веселиться 'rejoice', возмущаться 'be indignant'	**Internal experience**: experience, feel **Emotions**: *love*, *hate*
Permanent emotional states and relations: любить 'love', обожать 'adore', страдать 'suffer, hurt'	
Mental states: интересоваться 'be interested in', колебаться 'hesitate', знать 'know', помнить 'remember', считать 'consider'	**Cognition**: *know*, *believe*, *understand*
	Propositional attitude: *consider*
	Perception: *hear*, *see*

Tab. 1. Emotion, perception, desire and cognition verbs as presented in [3] and [4]

The classes exemplified represent universally acknowledged semantic distinctions, which nonetheless yield different accounts. A worthwhile effort, which constitutes part of our work, would be to compare relevant classifications with the goal of accommodating meaningful distinctions and enriching the description.

1.2 Stative verbs in language resources

A number of lexical semantic resources, such as FrameNet, WordNet, VerbNet, etc., have employed semantic groupings of different granularity to identify semantically coherent verbs. In WordNet such groupings are defined by semantic primitives that divide the verbal domain into 15 classes [5], while FrameNet provides a much more fine-grained approach based on the definition of conceptual frames [6], cf. Section 2. As a result, distinct semantic classes emerge; as both resources have a netlike structure implemented through a number of relations between the basic units of the resource (synonym sets in WordNet; frames in FrameNet), the so-induced classifications have a partially hierarchical organisation.

Such lexical resources provide a schema for annotating verbs in corpora. At the current stage, we adopt an approach to providing the corpus with as much classificatory information as possible, combining information from different

resources. In order to annotate stative verbs in particular, we need to identify them in the resources. As we employ an alignment between WordNet synsets and FrameNet frames, it is sufficient to identify the relevant verb synsets in WordNet as they are explicitly marked or deduced from the WordNet structure.

1.3 Motivation

Our interest in stative verbs is motivated by a joint project undertaken by Bulgarian and Russian researchers that aims at an ontological description of stative verbs in the two languages. Stative verbs are a natural place to start as in a number of theoretical accounts ([1], [3]) among others, they form one of the building blocks (together with activities) employed in the construal of more complex situations. The features of stative verbs are, nonetheless, far from being exhaustively and definitively determined and the membership of verbs to this ontological class is still subject to debate. The corpus aims at providing a test setting for linguistic observations on stative verbs but may readily be extended to include other ontological classes and to perform other tasks.

2 STATIVE VERBS: DATA SELECTION AND ANNOTATION

2.1 Lexical-semantic resources: WordNet

We combine information from several previously developed resources for Bulgarian and Russian, as well as for English for those resources where the two Slavic languages are linked through it (WordNet).

The Princeton WordNet, PWN [7] is a large-scale lexical database that encompasses the lexis of English organised as a network of synonym sets (synsets) comprising conceptual synonyms (individual members of a synset are called literals) linked to each other by means of conceptual, lexical, and other relations. The semantic description pertaining to a synonym includes a semantic label assigned to each verb or noun synset that denotes the semantic primitive of the respective verb or noun synset [5]. In addition to PWN, we use available wordnets for two other languages, Bulgarian [8] and Russian [9], each of which is mapped onto PWN 3.0 through unique synset identifiers. The Bulgarian WordNet contains 14,103 verb synsets, while the Russian WordNet is considerably smaller with 7,634 verb synsets. These wordnets provide the verb inventories used in the study; corresponding synsets are paired at the synset level through their mapping onto PWN.

Example 1.

ID: eng-30-02756359-v
PWN Synset: {belong:5}
Semantic primitive: verb.stative

Gloss: be a member, adherent, inhabitant, etc. (of a group, organization, or place)
Example: They belong to the same political party.
Bulgarian Synset: {принадлежа:7}
Russian Synset: {принадлежать 02365119}

According to their semantic primitive (or atomic predicate) [5], the verb synsets in WordNet are organised in 15 classes such as verbs of change, verbs of motion, verbs of cognition, verbs of communication, verbs of emotion, among others, and are accordingly labelled at the synset level. The class of stative verbs (marked as verb.stative) includes not semantically coherent verbs, but rather verbs that aspectually belong to the category of states. In addition, stative verbs are also found across other classes, although not necessarily characterised as such (e.g., cognitive, emotion verbs, verbs denoting bodily states, verbs of possession, etc.).

To obtain the verbs that denote states, we assume that the verbs labelled as stative qualify as viable candidates. In the first step, we take the set of stative verb synsets in the Bulgarian WordNet, thus obtaining a collection of 559 synsets. We then expand this number by adding verbs that are hyponyms of stative synsets (659 synsets). Further, we add a selection of verbs labelled with the primitives verb. emotion, verb.cognition and verb.perception, increasing the overall number of synsets to 1,786.

We then match these synsets to their Russian WordNet equivalents whenever they exist. Table 2 shows the number of verb synsets under analysis (and the number of literals they contain) in the Bulgarian and Russian WordNets and their corresponding semantic primitive.

	Bulgarian WordNet		Russian WordNet	
Semantic prime	# synsets	# literals	# synsets	# literals
verb.stative	559	1725	392	641
verb.cognition	503	1776	389	605
verb.perception	342	1173	261	432
verb.emotion	264	1035	210	385
verb.change	27	29	6	10
verb.body	19	86	14	21
Others	72	252	56	96

Tab. 2. Distribution of stative verbs across primes in the Bulgarian and the Russian WordNet

As mentioned above, we also collect verbs whose hypernym is a stative verb but they themselves are assigned a different semantic class. Such synsets can be considered as stative verbs with additional semantic characteristics expressed by the

semantic class assigned: for instance, Example 2 shows two hyponyms of a verb. stative synset where one is classified as verb.consumption (denoting the meaning of a state reached through consumption), and the other, an emotional state, is defined as verb.emotion.

Example 2.

Hypernym ID: eng-30-02604760-v
PWN Synset: {be:4}
Semantic primitive: verb.stative
Gloss: have the quality of being; (copula, used with an adjective or a predicate noun)

Hyponym 1 ID: eng-30-01188342-v
PWN Synset: {be full:1}
Semantic primitive: verb.consumption
Gloss: be sated, have enough to eat

Hyponym 2 ID: eng-30-02604760-v
PWN Synset: {seethe:3; boil:4}
Semantic primitive: verb.emotion
Gloss: be in an agitated emotional state

In the next stages of the research, we intend to focus on ways of increasing (through creating, translating, etc.) the Slavic data with more synsets denoting stative meanings that are frequently found in Bulgarian and Russian parallel, comparable or monolingual corpora, including prefixed verbs that are typical for Slavic languages, but are not included in the Princeton WordNet.

2.2 Lexical-semantic resources: FrameNet

FrameNet [6] is a network of conceptual frames, where each frame represents a script-like description of the conceptual structure of situations, objects or events by means of their participants and props, called Frame Elements [10]. The frames are instantiated by word-meaning pairings called Lexical Units. In addition, frames are linked to each other by means of several hierarchical (Inheritance, Using, Subframe, Perspective) and non-hierarchical relations (Causation, Inchoation, Precedence).

FrameNet frames are assigned to synsets in WordNet using one of the proposed automatic mappings between the resources where lexical units in FrameNet and synonyms in WordNet synsets are aligned; where such alignment is impossible, the synsets are assigned a frame from their parent synset or another suitable frame is assigned using a number of additional automatic procedures [11]. More than 5,000 frame-to-synsets assignments have been validated manually. As conceptual information

is to a large extent language-independent, the semantic information is transferrable across languages. Fig. 1 shows an excerpt of corresponding Bulgarian and Russian synsets labelled as stative either on the basis of their primitive (verb.stative) or as hyponyms of a verb.stative synset. The FrameNet frame assigned to the synsets and the pertaining Frame Elements are also exemplified.

SYNSET	SEM. CLASS	BG SYNSET	RU SYNSET	FRAME	FRAME Elements
eng-30-00033599-v	verb.body	изглеждам:1; имам вид:1; излизам:15	[смотреть, глядеть]	Give_impression	Phenomenon:; Characterization:; Appraisal:; Inference:;
eng-30-00047610-v	verb.body	нося:33	[быть_одетым]	Have_associated	Entity:; Topical_entity:;
eng-30-00047745-v	verb.body	имам:25; нося:34	[носить]	Wearing	Wearer:Sentient; Clothing:Artifact; Body_part:Body_part;
eng-30-00065370-v	verb.body	боледувам:1; страдам:2; болен съм:1; имам:11	[обладать, иметь]	Medical_conditions	Patient:Living_thing; Ailment:;
eng-30-00077698-v	verb.body	задавям се:3; задавя се:3; давя се:3; задушавам се:1	[задыхаться]	Perception_body	Experiencer:Sentient; Body_part:Body_part;
eng-30-00623006-v	verb.cognition	поставям:4; поставя:4; задавам:1; задам:1	[говорить_загадками]	Stimulate_emotion	Experiencer:Sentient; Stimulus:;
eng-30-01188144-v	verb.consumption	умирам от глад:2; умра от глад:2; гладувам:3	[голодать, быть_голодным]	Perception_body	Experiencer:Sentient; Body_part:Body_part;
eng-30-01763101-v	verb.emotion	преливам:3; прелея:3; преизпълнен съм:1; кипя:5	[переливаться_через _край]	Emotion_heat	Experiencer:Sentient; Emotion:; Seat_of_emotion:;
eng-30-01763303-v	verb.emotion	вря:4; бушувам:6	[вызывать_брожение]	Emotion_heat	Experiencer:Sentient; Emotion:; Seat_of_emotion:;
eng-30-02134927-v	verb.perception	звуча:1; прозвучавам:1; прозвуча:1; озвучавам:1	[издавать_звук, звучать]	Give_impression	Phenomenon:; Characterization:; Appraisal:; Inference:;
eng-30-02603699-v	verb.stative	съществувам:3; съм:13; има:4; битувам:1	[быть, существовать]	Existence	Entity:;
eng-30-02614181-v	verb.stative	живея:2; жив съм:1; съм:18	[существовать]	Dead_or_alive	Protagonist:Sentient; Figure:;
eng-30-02614387-v	verb.stative	живея:8; поминувам:1; поминавам:1; помина:1	[жить]	Manner_of_life	Experiencer:; Lifestyle:; Manner:;
eng-30-02198234-v	verb.perception	струва ми се:2; сторва ми се:2; стори ми се:2	[казаться, представляться]	Unattributed_information	Unattributed_information 1 Reported_fact:;

Fig. 1. A sample of the synset-to-frame alignment for several stative verbs

In order to expand the number of stative verbs, we use both the manually checked WordNet-to-FrameNet alignment and the netlike frame organisation, in particular part of the FrameNet frame-to-frame relations. The FrameNet 'tree' stemming from the frame State represents stative situations according to the relation of Inheritance.

Our working assumption is that frames inheriting from State must also be stative, as well as the verbs they describe (Fig. 2). We thus consider a set of 178 frames regarded as describing stative verbs and situations. Some frames can cover both stative and active verbs, e.g., Assessing, which is assigned to verbs such as *value* (stative) and *grade* (active).

Through their alignment with English, the synsets in the Bulgarian and the Russian WordNets are also assigned a frame, confer Table 3 for the most frequent frames. The last column signifies whether the frame is a confirmed stative frame, marked with an X (either coming from the FrameNet tree rooted in State, or manually confirmed as stative).

- State, ID# 150. Definition. (Total related: 134)
  - 132 frames related to State via Inheritance (20 immediate Inheritance, 54 total Inheritance).
    - Locative_relation, ID# 199. Definition. (contains 16 Inheritance relations)
    - Process_initial_state, ID# 152. Definition. (contains 2 Inheritance relations)
    - Change_of_state_initial_state, ID# 185. Definition
    - Change_of_state_endstate, ID# 186. Definition
    - Dead_or_alive, ID# 310. Definition. (contains 1 Inheritance relation)
    - Bearing_arms, ID# 470. Definition.
    - Being_located, ID# 960. Definition. (contains 3 Inheritance relations)
    - State_of_entity, ID# 1600. Definition. (contains 1 Inheritance relation)
    - Being_in_operation, ID# 880. Definition
    - Attention, ID# 791. Definition.
    - Existence, ID# 660. Definition. (contains 1 Inheritance relation)
    - Posture, ID# 18. Definition
    - Being_attached, ID# 307. Definition.
    - Emotions, ID# 1712. Definition. (contains 8 Inheritance relations)
    - Process_completed_state, ID# 234. Definition. (contains 1 Inheritance relation)
    - Process_uncompleted_state, ID# 1764. Definition. (contains 1 Inheritance relation)
    - Thriving, ID# 1771. Definition
    - Dying, ID# 2055. Definition.
    - Transportation_status, ID# 2645. Definition
    - Chaos, ID# 2705. Definition

Fig. 2. The shallow hierarchy beginning with State according to the Inheritance relation

FrameNet frame assigned	# synsets BG	# synsets RU	Stative frame
NO FRAME ASSIGNED	154	92	
Stimulate_emotion	132	111	X
Locative_relation	37	27	X
Categorization	38	32	
Assessing	24	11	
Purpose	22	16	
Perception_body	20	13	
Compatibility	16	11	X
Similarity	15	11	X
Have_associated	15	8	X
Give_impression	11	5	X
Posture	11	9	X
Residence	11	9	X
Existence	10	6	X
Expertise	10	6	X

Tab. 3. The most frequent frames (10+ examples) assigned to stative verbs in the two wordnets

The semantic information from WordNet and FrameNet, including the gloss, semantic prime, examples, etc. from the WordNet synsets and the frame definition, lexical units (other verbs) assigned to the frame, etc. provide valuable semantic information that will be used in the analysis of the stative verbs. The synset semantic primes and the FrameNet frames (as distinct entities from other frames) are especially helpful as they suggest meaningful classificatory categories. For instance, Being_ located (*sit, lie, stand 'be located at'*, etc.) and Spatial_contact (*meet*, *contact*, *touch, adjoin 'be in physical contact with'*) verbs may be defined as subcategories of a more general classification category Location; Residence verbs (*live*, *occupy*, *dwell*, *camp*, *bivouac*, *room*, *stay*, *squat*, *lodge*) and Existence verbs (*live*, *exist*, *be 'have existence'*; *consist in*, *lie in*, *dwell* 'originate in') may be defined as distinct categories, etc. In addition, the analysis of the Frame Elements (the last column of Fig. 1) is very helpful in identifying the semantic and selectional properties of the verbs' arguments.

2.3 Corpus data and preliminary annotatiton

For the purposes of the current work, we employed the Polish-Bulgarian-Russian Corpus ([12], [13]), a parallel corpus for the three languages incorporated in the CLARIN framework. The Corpus consists of 55 parallel texts, comprising 2.23 mln. words for Bulgarian and 2.04 mln. words for Russian from several text genres such as fiction, instruction manuals and technical documentation, legal texts, etc. The parallel texts are automatically aligned at sentence-level and the annotations have been post-edited manually. For the two languages under study we thus obtain 89,562 parallel sentences.

The Bulgarian corpus was POS-tagged using the Bulgarian Language Processing Chain [14]. The Russian part of the corpus was POS-tagged with an available UDPipe language model for Russian [15]. The tagging is necessary in order to identify the relevant verb lemmas to the end of matching them to possible WordNet senses.

3 DATASET OF ANNOTATED EXAMPLES OF STATIVE VERBS

The task is to annotate the stative verbs in the parallel Bulgarian-Russian Corpus obtained from the Polish-Bulgarian-Russian Corpus. The annotation involves the assignment of a relevant WordNet synset that best describes the sense using the Bulgarian and the Russian WordNets. As sense annotation is very sensitive and prone to mistakes, the decision making will be delegated to human experts who will choose the most relevant sense (synset) out of a number of automatically assigned synsets. To facilitate the process, we have adopted a procedure for filtering out non-relevant synsets, which we describe below.

Step 1. We first assign all the possible senses to the lemmatised verbs in the Bulgarian part of the parallel corpus that have at least one stative sense in WordNet. These verbs (or rather their graphic form) have counterparts in the collection of possible relevant synsets and are thus potentially stative.

Step 2. For each sense assigned to a potential stative verb in the Bulgarian part of the corpus, we collect the corresponding synsets from the Russian WordNet, where available.

Step 3. We identify the verbs in the tagged Russian part of the corpus that may potentially belong to the same synset as the corresponding Bulgarian verb in the parallel Bulgarian sentence. The task boils down to finding the intersection of the set of Bulgarian and Russian synsets which are assigned to a Bulgarian verb and a Russian verb, respectively, found in a pair of equivalent sentences: candidates from corresponding synsets appearing in a pair of parallel sentences are very likely translational equivalents.

Step 4. If no pair of verbs from the corresponding Bulgarian and Russian synsets are identified, for each Bulgarian verb, we extract all Russian stative verb translations (in the corresponding sentence) and include them in the list of possible candidates. The assumption is that a state is more likely to be expressed by stative verbs in both languages, even if not from the same synset.
At this stage, a number of heuristics based on semantic relations between synsets can be employed in order to improve filtering of invalid suggestions and reduce further manual validation.

Step 5. After the list of possible senses is reduced through the filtering procedures, we assign the FrameNet frames mapped to the relevant synsets.
As a result, the potentially stative verbs in the Bulgarian-Russian parallel corpus are assigned a number of (filtered-out) senses. Each verb is thus supplied with semantic information derived from the respective WordNet synsets and the assigned FrameNet frames: the semantic prime and the description of the conceptual frame as well as the semantic relations with other synsets or frames. The corpus is then ready to be further disambiguated by human experts.
Initially, we extracted over 30,000 pairs of parallel sentences from the corpus, which were then filtered down to 7,568 examples representing possible stative verbs in Bulgarian and their parallel equivalents in Russian (Example 3).

Example 3.

BG verb: съвпадам
BG sentence: – Вашият разказ е изключително интересен, професоре, въпреки че далеч не ***съвпада*** *с евангелските.*

EN translation: 'Your story is extremely interesting, Professor, though it does not ***coincide*** *at all with the Gospel stories.'*

Potential synsets*:*

eng-30-02658734-v verb.stative {съвпадам:4; съвпадна:4} {coincide:2} 'be the same'

Frame: Compatibility; FEs: Item_1; Item_2; Items; Parameter

eng-30-00345312-v verb.change {съвпадам:1; съвпадна:1} {concur:1; coincide:1} 'happen simultaneously'

No frame assigned

eng-30-02660442-v verb.stative {съвпадам:3; съвпадна:3} {coincide:3; co-occur:1; cooccur:1} 'go with, fall together'

Frame: Existence; FEs: Entity

RU verb: совпадать

RU sentence: – Ваш рассказ чрезвычайно интересен, профессор, хотя он и совершенно не совпадает с евангельскими рассказами.

Potential synsets*:*

eng-30-02658734-v verb.stative {совпадать 02278040}

Frame: Compatibility Item_1:; Item_2:; Items:; Parameter:;

eng-30-00345312-v verb.stative {совпадать 00297090}

No frame assigned

eng-30-02660442-v verb.stative {совпадать 02279659}

Frame: Existence; FEs: Entity

4 CONCLUSIONS

The research presented in this paper suggests several lines of improvement: (i) expanding the inventory of verbs, the FrameNet-to-WordNet alignment, and the size of the parallel corpus; (ii) perfecting the automatic sense assignment and filtering procedures; (iii) outlining major classification categories on the basis of analysis informed both from theoretical work on verb classification and the semantic knowledge encoded in lexical-semantic resources. Further, the classification scheme can be applied to (semi)automatic classification of corpus examples and can be used as a starting point towards automatic semantic role labelling and word sense disambiguation.

ACKNOWLEDGEMENTS

This research is carried out as part of the project *An Ontology of Stative Situations in the Models of Language: a Contrastive Analysis of Bulgarian and Russian* funded by the Bulgarian National Science Fund under the Programme for Bilateral Cooperation, Bulgaria – Russia 2019 – 2020, Grant Agreement No. КП-06-ПРУСИЯ-78 from 2020.

References

[1] Vendler, Z. (1957). Verbs and times. The Philosophical Review, 66(2), pages 143–160.

[2] Dowty, D. (1979). Word meaning and Montague grammar. Reidel: Dordrecht.

[3] Van Valin, R. D., and LaPolla, R. J. (1997). Syntax: structure, meaning, and function. Cambridge: Cambridge University Press.

[4] Paducheva, E. V. (1996). Semanticheskie issledovaniya. Semantika vremeni i vida v russkom yazyke. Semantika narrativa, 464 p.

[5] Miller, G. A., Beckwith, R., Fellbaum, C., Gross, D., and Miller, K. (1991). Introduction to WordNet: an on-line lexical database. International Journal of Lexicography 3(4), pages 235–244. Revised: 1993. Accessible at: https://wordnetcode.princeton.edu/5papers.pdf.

[6] Baker, C. F., and Ruppenhofer, J. (2002). FrameNet's frames vs. Levin's verb classes. In Proceedings of the 28th annual meeting of the Berkeley.

[7] Fellbaum, C. (Ed.) (1998). WordNet: an electronic lexical database. Cambridge, MA: MIT Press.

[8] Koeva, S. (2010). Bulgarian WordNet – current state, applications and prospects. In Bulgarian-American Dialogues, pages 120–132, Sofia: Prof. M. Drinov Academic Publishing House.

[9] Gelfenbeyn I., Goncharuk A., Lehelt V., Lipatov A., and Shilo V. (2003). Automatic translation of WordNet semantic network to Russian language. In Proceedings of International Conference on Computational Linguistics and Intellectual Technologies Dialog-2003. Accessible at: http://wordnet.ru.

[10] Ruppenhofer, J., Ellsworth, M., Petruck, M. R. L., Johnson, C. R., Baker, C. F., and Scheffczyk, J. (2016). FrameNet II: extended theory and practice. Accessible at: https://framenet2.icsi.berkeley.edu/docs/r1.7/book.pdf.

[11] Leseva, S., and Stoyanova, I. (2020). Beyond lexical and semantic resources: linking WordNet with FrameNet and enhancing synsets with conceptual frames. In Koeva, S. (Ed.) Towards a semantic network enriched with a variety of semantic relations. Sofia: Prof. Marin Drinov Academic Publishing House of Bulgarian Academy of Sciences.

[12] Kisiel, A., Koseska-Toszewa, V., and Kotsyba, N. (2016). Polish-Bulgarian-Russian parallel corpus, CLARIN-PL digital repository. Accessible at: http://hdl.handle.net/11321/308.

[13] Sosnowski, W. (2016). The parallel Polish-Bulgarian-Russian corpus: problems and solution. In Pastor, G. C. (Ed.), Computerised and corpus-based approaches to phraseology: monolingual and multilingual perspectives, pages 339–349, Geneva: Tradulex. Accessible at: https://www.researchgate.net/publication/319243154_The_parallel_Polish-Bulgarian-Russian_corpus_problems_and_solution.

[14] Koeva, S., and Genov, A. (2011). Bulgarian language processing chain. In Proceeding of the Workshop Integration of multilingual resources and tools in Web applications, Hamburg.

[15] Straka, M., and Straková, J. (2019). Universal dependencies 2.5 models for UDPipe (2019-12-06), LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL), Faculty of Mathematics and Physics, Charles University. Accessible at: http://hdl.handle.net/11234/1-3131.