Towards Incorporating Prepositions in BTB-WordNet: A Case Study

Zara Kancheva

Institute of Information and Communication Technologies, Bulgarian Academy of Sciences

zara@bultreebank.org

Introduction

- case study on preposition incorporation in BulTreeBank WordNet
- substantial role in many NLP tasks, but their polysemy constitutes one of the greatest challenges for this research area
- seriously benefit the performance of wordnets text analysis and generation, word-sense disambiguation, automatic translation, etc.
- focus on the preposition '***на***' (na) in ***verb+preposition+noun*** constructions
- '***на***' (na) is the most frequent Bulgarian preposition, because it is loaded with the most numerous and most abstract senses
- the prepositions are categorized by semantic classification and the verbs by the categories from Princeton WordNet

Classification of prepositions

- the adapted version of the classification of prepositions (Stoyanov, S. Ed.: *Grammar of contemporary Bulgarian standard language*, 1983) contains **7** classes: ***locative***, ***temporal*** (example (1)), ***transition***, ***manner***, ***property and possession***, ***quantitative*** and ***purpose***

(1) Koledната vakancija šte započne na 21 dekemvri i šte svărši na 14 januari.
'Christmas vacation will start on 21 December and will end on 14 January.'

Classification of verbs

- the categories from PWN which are inherited in OEW and BTB-WN are used for the verb analysis
- **15** classes of verbs***:*** **verb.body**, **verb.change**, **verb.cognition**, **verb.communication**, **verb.competition**, **verb.consumption**, **verb.contact**, **verb.creation**, **verb.motion**, **verb.perception**, **verb.possession**, **verb.stative**, **verb.social**, **verb.emotion** and **verb.weather**

Preposition category	Verb categories
locative	**verb.contact** (*hang, lean, stay, lie, put, bang, rest*) **verb.motion** (*go, stay, sit, return, go out, land, leave*) **verb.stative** (*live, attend, be*) **verb.perception** (*look, look around*) **verb.creation** (*write*) **verb.communication** (*express*)
purpose	**verb.stative** (*be, attend*) **verb.motion** (*go, send*) **verb.change** (*show up*) **verb.contact** (*put*)
quantitative	**verb.stative** (*be*), **verb.change** (*happen, set*) **verb.social** (*split up, divide*) **verb.cognition** (*separate*) **verb.change** (*disintegrate*)
property and possession	**verb.stative** (*be*)
manner	**verb.cognition** (*leave*) **verb.emotion** (*feel*) **verb.creation** (*perform*) **verb.stative** (*exist*)
transition	**verb.communication** (*pretend*)
temporal	**verb.stative** (*be, end*) **verb.change** (*start*)

Table 1. Verb classes distribution in the preposition categories

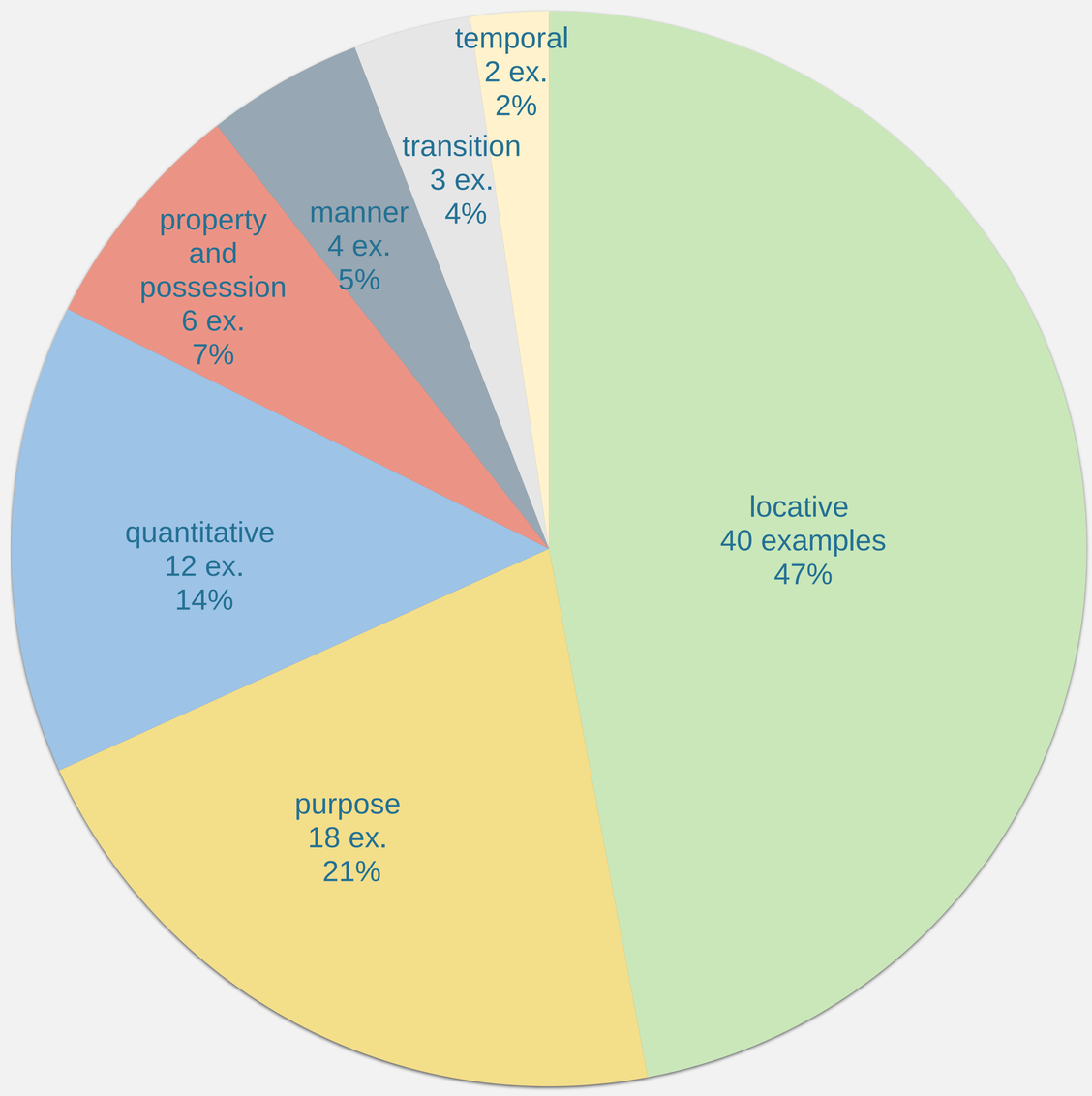
Model for preposition synset in BTB-WN

- detailed definition, synonyms if available, and examples
- generalized classification for all Bulgarian prepositions > categories of the synsets
- **relations between prepositions and other parts of speech** (for example, the verbs '***pretend***' and '***turn in***' combined with ***на*** express transition in new, different state)
- **relations between preposition synsets** (for example, the synset of ***на*** ('on') with definition "*Location relation in which something is located on some surface*" has a synonym – ***върху***, 'on', and also has an antonym relation with the synset for ***под***, 'under'

Analysis and results

- **210** phrases of the type ***verb+на+noun*** were extracted from BulTreeBank and manually sorted following a semantic classification
- in most of the examples '***на***' expresses indirect object – one of its main functions
- the most frequent semantic relation is ***locative***

Preposition categories and number of phrase occurrences



Conclusion and future work

- initial step towards the large-scale integration of prepositions in BTB-WN
- combination of semantic preposition classification, verb categorization from wordnet and later also relations and noun categories from wordnet, and features from a valency lexicon
- particular groups of verbs tend to combine with the preposition '***на***' (na) for specific senses
- the analysis will be elaborated with hierarchy inheritance of the verbs and nouns in wordnet and with the valency lexicon

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.