

Špela Antloga

Researcher Assistant at the Digital Signal Processing Laboratory at the **Faculty of Electrical Engineering and Computer Science**, University of Maribor, Slovenia

<https://dsplab.feri.um.si/en/dsplab/>, <https://feri.um.si/en/>

s.antloga@um.si

Report on the ELEXIS Transnational Visit Grant at the Department of Computational Linguistics, Institute for Bulgarian Language, Bulgarian Academy of Sciences (ИНСТИТУТ ЗА БЪЛГАРСКИ ЕЗИК) (Sofia, Bulgaria, 5 July – 19 July 2022)

Introduction

I applied for a visit to the Institute for Bulgarian Language, having in mind that they have a very strong department for computational linguistics with several successfully completed and on-going projects. Their work is in many ways related to the ideas of my proposed projects, but also interesting from a broader perspective for my work. The project proposal that I submitted to the call for ELEXIS grants for research visits included the plan to get introduced to the (1) general principles of annotation, (2) the Bulgarian Sentence Aligned Corpus, (3) to get familiarized with FrameNet Frame Semantics, with special attention to Conceptual Frames in Bulframe and to the ontology of metaphorical/metonymical units, and with the (4) Bulgarian Wordnet, with special attention to the constructions (patterns) of metaphorical/metonymical units and multiword expressions.

Aims of the project KOMET (KOMET 1.0 and g-KOMET)

The aim of my projects KOMET 1.0 and g-KOMET that have been financed by CLARIN.SI Infrastructure is to build a hand-annotated corpus for metaphorical expressions in Slovene in various text types, including news texts, transcriptions of spoken language, academic texts and fiction, to (1) furthermore identify the semantic structures of metaphorical units, and to (2) compare the metaphorical expressions among different text types. The MIPVU annotation procedure lacks some formal aspects of annotation that are extremely important for the purposes of my further analysis. Therefore, I wanted to discuss possibilities of adding some of the elements of the semantic frames to my annotation of metaphorical units and/or in their analysis. I was especially interested in their approach to identifying and defining metaphorical and metonymical constructions as well as multiword expressions.

My visit at the Institute for Bulgarian Language

My visit was following a preliminary programme:

6 July

I met some of the researches at the Department for Computational Linguistics and got familiarized with the department, their research focus, their projects etc.

Topic: **General Principles of Annotation**

We discussed pros and cons of different annotation tools such as WebAnno, Inception, SketchEngine etc. and various annotation procedures depending on the annotator's aim. Together, we discussed some of the problematic points in my annotation scheme and tried to find sensible and convenient solutions.

7 July

Topic: **Bulgarian Sentence Aligned Corpus**

I was introduced to the Bulgarian-English Sentence- and Clause-Aligned Corpus. The discussion covered the reasons behind the corpus development, the structure and content of the corpus, the segmentation and alignment strategy and the tools used in the corpus processing. I was able to get hands-on experience in sentence splitting, clause splitting, correction of wrong splitting and the identification of the type of relation between pairs of connected clauses.

Topic: **FrameNet Frame Semantics**

In discussing some of the basics of lexical semantics we focused on The Berkeley FrameNet project. I was particularly interested in criteria of how words, in each of their senses, are linked to frames, which are schematic cognitive structures that represent a speaker's knowledge of the described. That was a very helpful insight into the theory which is in its core (and especially when it comes to meanings that are contextually used in metaphorical sense) extremely similar to describing metaphorical mapping between the domains. We discussed some of the possibilities of how FrameNet Frame Semantics can be used to analyse metaphorical/figurative meaning.

8 July

Topic: **Bulgarian Wordnet**

We discussed the whole process of creating a dictionary of synsets along with their explanatory definitions. I was especially interested in connecting sets of semantic relations with other words in the language since the large number of those relations shows the semantic richness of the language. We were comparing the synsets and semantic relations in Bulgarian, English and Slovene.

11 July

Topic: **Constructions (patterns) of Metaphorical/ Metonymical Units**

Dealing with a similar topic this has been extremely helpful for me in sense of understanding the representation of the structure of idioms (focusing on verbal idioms). This is extremely important in NLP applications such as machine translation, paraphrasing etc. when defining a border between compositional and non-compositional word meaning. We discussed the importance of formal description in defining the elements in metaphorical (idiomatic) constructions since the majority of my annotation so far has grounded on more or less interpretational basis.

12 July

Topic: **Sense annotated Corpus**

We covered topics of sense disambiguation (lexical items defined in the context of occurrence) in BulSemCor. I was interested in specific challenges when manually assigning the unique semantic (or grammatical) meaning in the particular context to each lexical unit since that is touching the

problem of also assigning the potential metaphorical (and metonymical) meaning to each lexical unit. We continued the discussion on challenges and solutions when defining metaphorical MWEs.

13 July

Topic: **Annotation in Bulgarian National Corpus**

We talked about the key concepts related to the development of corpora, especially in light of developments in NLP.

Topic: **WordNet and Metaphorical/ Metonymical Units**

This was a very interesting discussion for me in terms of how to firstly extract the candidates - potentially metaphorical units (on the case of verbs of perceiving) and later on how to determine their metaphorical meaning. We had a very fruitful discussion about challenges when defining whether a specific sense is metaphorical or not, e. g., new metaphorical meaning of an existing word, metaphorical meaning in neologisms, registered metaphorical meanings in dictionaries, etc.

14 July

Topic: **Lecture of Špela Antloga “Metaphorical expressions in Slovene: building a corpus”**

The Institute for Bulgarian Language organized an on-line lecture where I was able to present my work. I outlined the problems with the identification of metaphorical expressions in texts (focusing on the specifics of Slovene language) comparing various methods for identifying/extracting metaphorical expressions. I focused on some of the challenges and solutions when annotating metaphorical expressions in Slovene and possibly other Slavic languages. We discussed some of the examples from the corpus KOMET 1.0 (corpus of metaphorical expressions in Slovene) that is hand-annotated for metaphorical expressions for the following relations to metaphor: indirect metaphor, direct metaphor, borderline cases and metaphor signals. I added a new frame tag, which gives an information about a concept, to which a metaphorical expression refers. I quickly shared some of my analysis based on the already annotated data.

15 July

Topic: **CPA semantics**

I was introduced to a method I did not know about and could be very interesting and applicable for me in the next stages of my project, especially when analysing the annotated data or possibly comparing the data to other languages. The Corpus Pattern Analysis tries to show the striking patterns of words use in natural language. What was extremely interesting for me was the possibility of those kind of explorations including metaphors and other “non-normal uses”, e.g., metonymy, based on the distinction from literal uses. We took some examples from my corpus and tried to use the proposed method for the analysis. That again showed the similarity of different approaches but on the other hand the diversity and possibilities of approaching a specific language phenomenon from several angles.

18 July

Topic: **Conceptual Frames and Bulframe**

We talked about the expansion the Princeton WordNet with conceptual frames. The approach was developed at visiting Institute. We discussed the syntagmatic relations of verb synsets and the semantic classes of nouns. We also discussed the possibilities of incorporating some of the

elements of semantic/conceptual frames into my project in terms of mimicking the defining of semantic and syntactic compatibility between verb predicates and noun arguments.

Topic: **Ontology of Metaphorical/Metonymical Units**

I was introduced to a more fine-grained ontology by mapping WordNet noun hierarchies with the CPA ontology and to challenges in defining semantic classes. Since the representation or organization of concepts in my annotation scheme when defining the source of metaphorical mapping is based on predefined categories and when needed on my intuition, I had to deal with the shallowness of annotated frames. This talk gave me a greater insight into more systematic organization and categorization of different semantic types.

Concluding remarks and acknowledgements

The visit has been very informative and definitely extremely useful for my future work on building a corpus of metaphorical expressions in written and spoken Slovene language. I am very grateful for receiving ELEXIS Transnational Research Visit Grant to get an opportunity to receive guidance, advice and consultations from the experts in the field of computational linguistics and other fields from experts from The Institute for Bulgarian Language I met during my stay. I would like to thank The Institute for Bulgarian Language and to each and every person who made sure that my stay succeeded beyond my expectations.