Greek and Serbian MWEs in Ontology-Based Lexical Resources
STSM Report
Institute for Language and Speech Processing/ Athena RIC, Athens (EL)

Host: Dr. Styliani Markantonatou
Visitor: Jelena Mitrović

October 1st – October 15th, 2015

Introduction

Figurative language has the power to affect the semantics of texts or spoken language and it has been studied in depth recently in the context of Computational Linguistics for Serbian (Mladenović and Mitrović, 2013; Mladenović et.al, 2016) – this work pertains to different rhetorical figures (many are MWES) and to the figure simile, in particular, which is always a MWE. Therefore, it was an interesting idea to try and investigate similar rhetorical constructs in Greek – a language that is geographically close to Serbian, belonging to the family of Indo-European languages, but seemingly very different. The aim of this Short-Term Scientific Mission was to find similar ways in which these two languages function, in the scope of the lexical resources based on ontologies that we have at our disposal for both Greek and Serbian.

Description of the work carried out during the STSM

The basis for the research efforts that were carried out during this STSM was the assumption that Greek MWEs have similar semantic and syntactic properties with Serbian MWEs, due to the close connections that our two nations have had for centuries. The initial idea was to try and incorporate rhetorical figure encoding somehow into the General conceptual lexicon for the Greek language, as presented in (Fotopoulou et. al, 2014) but that task proved to be too complex as it would temper with the existing structure of the database, therefore, that initial idea was abandoned. From there, we decided to change our course of action, keeping the original plan of using ontology based lexical resources as a starting point for our research into the similarities between Greek and

Serbian MWEs. For that purpose, we used the Ontology of Rhetorical Figures for Serbian to get information about syntactic and semantic properties of some rhetorical figures, the figure simile, in particular. We chose this figure as it is the focus of the recently published research and we wanted to try and draw parallels between Serbian and Greek examples and their usage. This figure of speech is used quite frequently in everyday language and it was our hope that the Greek Corpus might contain many interesting examples of this figure. The figure simile is the comparison of one concept with another, implying a certain resemblance between them. This is used either for praise or censure. Its goal is to compare something unknown or less known with something similar, but well-known. Simile is an explicit comparison, often (but not necessarily) employing "like" or "as." Leaning on the recent research on Serbian similes and their encoding (Mitrović et. al, 2015; Mladenović et. al, 2016), we performed corpus-based investigation and cross-linguistic analysis of the Greek MWEs that could serve the purpose of similes.

Results of the work carried out during the STSM

We searched the Hellenic National Corpus (http://hnc.ilsp.gr/en/) in the hopes of finding interesting MWEs relating to the simile rhetorical figure. We extracted examples of one possible form of this figure. A total of 2084 results were obtained using the following restrictions: "Find all Adjective as Noun constructs (gr. Επίθετο σαν ουσιαστικό)". The third search criterion (regarding the noun), was set to the distance 2 because of the usage of article in front of the noun in Greek – (gr. Μέγιστη απόσταση 2 λέξεις) – thus, the longest distance between the particle "as" and the noun was set to two. Although a lot of noisy data were retrieved, we obtained a lot of valuable data based on which we were able to draw conclusions, such as the following:

1 The usage of adjectives denoting colors in similes is quite close in Greek and Serbian, e.g.

Άσπρο σαν το γάλα – Beo kao mleko – "As white as milk" – obično za opisivanje lica lepe devojke Bela kao mleko, ili (Lica) belog kao mleko.

Άσπρο σαν το χιόνι – Beo kao sneg – "As white as snow".

Άσπρο σαν το πανί – Beo(bled) kao krpa – "As white (pale) as a kitchen cloth" –negative sentiment, meaning that someone is very pale, very sickly looking Bled kao krpa.

Very often, though, certain constructs apply to particular entities, e.g. *Έγινε άσπρος σαν το πανί/την κιμωλία* – "Someone's face turned white" because of some emotional reaction, but *Η δασκάλα είναι άσπρη σαν το γάλα* – "The teacher has a white complexion".

2 Simile relating to characteristics of animals are used equally in both languages, e.g.

 Πονηρός σαν αλεπού – Lukav kao lisica – "As sly as a fox"

Πιστός σαν σκύλος/σκυλί – Veran kao pas – "As faithful as a dog"

3 Some similes did not carry similar meanings in both languages as it was expected, e.g.

In Serbian, it is very common to say that something is *Prosto kao pasulj*, literally, "As easy as (cooking) beans", which is an expression similar to "piece of cake", in English. In Greek, a similar expression translates to *κάτι είναι παιχνιδάκι*, meaning that something is as simple as child's play and there is no similar food related expression in Greek.

4 We used the structure *Adjective as Noun* for our search, but in Greek, a variety of syntactic structures is used to represent similes. This is something that would have to be analyzed in depth in future work. Some examples we found so far are:

Έγινε κόκκινη σαν παπαρούνα – She blushed (her face turned red like a poppy).
Έγινε άσπρος σαν το πανί/την κιμωλία – Someone's face turned white as a kitchen cloth.

5 Some similes carry positive conotations, while others carry negative connotations, and they are used similarly in Greek and Serbian, e.g.

Δουλεύει σαν μυρμήγκι (positive) – *Radi vredno kao mrav* – "As hardworking as an ant".

Δουλεύει σαν σκλάβος//ζώο/σκυλί (negative) – *Radi kao rob/životinja/pas* – "Works like a slave/an animal/ a dog".

6 Some other examples of similar usage in both languages:

Βρώμικος σαν γουρούνι – Prljav kao svinja "As dirty as a pig"

Λάμπει σαν τον ήλιο – Sjajan kao sunce – "As bright as the sun"

Χοντρός σαν βαρέλι – Debeo kao bure – "As fat as a barrel"

Future collaboration with the host institution

Cooperation between the two sites will continue regarding the enrichment of the Ontology of Rhetorical Figures for Serbian and its enrichment with Greek examples of the figure Simile, for the beginning, and hopefully with examples of other rhetorical figures, later. As most of the observations and results were acquired through brainstorming and with some help of Dr. Markantonatou's students, it would be very interesting if we could carry out a crowdsourcing experiment similar to the one presented in (Mitrović et. al, 2015) and compare the results. Otherwise, more substantial results might be collected through guided web search, which would require more involvement from students and other researchers.

We plan on continuing our joint efforts, hopefully resulting in publishing a comparative study of Greek and Serbian MWEs related to rhetorical figures that are also MWEs and which are used frequently in both Greek and Serbian everyday language. A part of this research will most likely be published in an upcoming go-to resource on the theory and computational aspects of Greek MWEs – a volume with the provisional title "MWEs in Greek and other languages: from theory to implementation" that will be published under the auspices of the Academy of Athens (Research Center for Neologisms and Scientific Terms).

Confirmation by the host institution of the successful execution of the STSM

Ms Jelena Mitrovic introduced a new dimension to the ongoing work on MWEs that takes place in ILSP/R.C. "Athena".

In particular, my group has been developing a Data Base for (verb at the moment) MWEs; for each MWE the DB encodes full morphosyntactic

information, corpus/Web examples, fully glossed usage examples, a description of the behavior of the MWE under structural diagnostics, the lexical relations of the MWE (synonyms, antonyms, causative-inchoative/non-causative pairs, dative genitive alternation), English translation and, of course semantic information. The DB will be made available on the WEB soon. The DB is planned to be linked to the Conceptual Lexicon of ILSP. In this way it is possible to achieve both an efficient description of MWEs and represent their functional position within the semantic universe of Modern Greek language.

Rhetorical figures, apart from the fact that they provide the structural framework for a wealth of MWEs, definitely have a strong cross-linguistic and cross-cultural dimension. Jelena and my students had the opportunity to discuss the aspect of rhetorical figures in MWEs and the Ontology of Rhetorical Figures for Serbian.

Certain similarities and differences across Serbian and Greek were expected and were observed on the corpus data that Jelena Mitrovic collected. It was a nice surprise to see that Greek provides a collection of structurally complex (including verb) MWEs that are clearly similes. Of course, such constructs exist in Serbian. This is an exciting new field of work. The plan is that by April 2016 a paper will be submitted to a volume that will be published by the Academy of Athens that will investigate these MWEs in both Greek and Serbian as regards the features of MWEs that are encoded in the DB. The impact of this cross-linguistic study on the Ontology of Rhetorical Figures will be discussed.

References

M. Mladenović and J. Mitrović, "Ontology of Rhetorical Figures for Serbian", Lecture Notes in Computer Science and Artificial Intelligence 8082/Springer-Verlag Berlin Heilderberg, 2013, pp. 386-393.

A. Fotopoulou, S. Markantonatou and V. Giouli, "Encoding MWEs in a conceptual lexicon", Proceedings of the 10th EACL Workshop on Multiword Expressions, 2014.

J. Mitrović, M. Mladenović, C. Krstev, "Adding MWEs to Serbian Lexical Resources Using Crowdsourcing" Accepted for poster presentation at the 5th General PARSEME meeting, September 23rd, 2015.

M. Mladenović, J. Mitrović and C. Krstev, A Language-independent Model for Adding New Semantic Relations to a WordNet, Global WordNet Conference, 2016 (accepted for publication).