STSM Scientific Report
FipsCo (a collocation extraction system based on the Fips parser) for Modern Greek

Vasiliki Foufi

Translation and Natural Language Processing Laboratory

Aristotle University of Thessaloniki

foufi@auth.gr

General

- Host institution: Laboratoire d'Analyse et de Technologie du Langage (LATL), Centre Universitaire d'Informatique (CUI), University of Geneva, prof. Eric Wehrli
- STSM participant: Dr. Vasiliki Foufi, Translation and Natural Language Processing Laboratory, Aristotle University of Thessaloniki
- Dates: January 4 – Mars 31, 2016

1 Purpose of the STSM

The goal of this mission was to develop a version of FipsCo (a collocation extraction system based on the Fips parser) for Modern Greek. While we dealt with a wide range of MWEs, including compounds, particular attention was given to verb-object collocations and its idiosyncratic characteristics in Modern Greek, like free word order, passivization, relativization.

2 Work carried out

The first part of the work consisted in the enhancement of the Fips grammar of Modern Greek and the associated lexical database. We enriched the existing monolingual lexicon with 1,692 lexemes and 1,791 collocations of various types (adjective + noun, simple noun + simple noun in genitive, simple noun + compound noun in genitive, verb + object collocations) that constitute both units of the general vocabulary and terms. We also enriched the bilingual lexical database (FR-GR). When needed, we added new collocations for the French language. Additionally, we extracted and validated multiword expressions from the EuroParl corpus by means of the Fips collocation extraction tool (based on the syntactic parser mentioned above).

3 Main results obtained

The main achievements carried out during the STSM was the improvement of the results issued by the syntactic analysis via the Fips parser and the extraction of the most frequent collocations and other multiword expressions from the EuroParl corpus. As of now, the Greek version of both Fips and FipsCo (the parser and the collocation extraction system) are available on-line on our webservice site (http://latlapps.unige.ch).

4 Future collaboration with host institution

The collaboration between the STSM grant holder and the host institution continues with the aim to enrich the system with data and improve as much as possible both the syntactic analysis and the translation results for the language pair French to Greek.

5 Foreseen publications to result from the STSM

The participants in the STSM have already submitted for a common publication in the collaborative book on MWE representation and parsing entitled "Multiword Expressions: Representation and Parsing" organised by the Parseme-WG2 working group. We also intend to submit a proposal for a communication at the next PARSEME meeting in Dubrovnik (Croatia), describing our work and results and at the 12th Workshop on Multiword Expressions (MWE 2016) in Berlin (Germany).

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

I'm pleased to report that Vasiliki's STSM has been a big success. All the research objectives, as well as the dissemination objectives, described in the plan have been accomplished. As mentioned above, both the parser for Greek and the collocation extraction system are now available as web services. Vasiliki and the LATL team will continue this successful collaboration during the next few months.

Eric Wehrli, LATL-CUI, University of Geneva