

The Greek FrameNet project: populating and interlinking a lexical resource

Voula Giouli¹, Vera Pilitsidou² & Hephæstion Christopoulos³

¹ National and Kapodistrian University of Athens, Institute for Language & Speech Processing
ATHENA RC, voula@athenarc.gr, ² National and Kapodistrian University of Athens,
verapilitsidou@gmail.com, ³ National and Kapodistrian University of Athens, hchristo@turkmas.uoa.gr

Keywords: FrameNet, Greek, semantics

Over the last decades, NLP researchers have long sought to develop resources tailored to build meaning representations beyond the levels of lexis and syntax. Even today, with the advent of large language models, the need for semantic knowledge that might be useful for various end-user applications is more than ever imperative. One step beyond, artificial agents are still lacking and thus, call for cognitive-based language representations. In this respect, framenets, that is, semantic lexical resources that have been developed for many languages after the renowned Berkeley FrameNet (BFN) for the English language (Fillmore and Baker, 2010; Fontenelle, 2003) are more than welcome. Based on the principles of Frame Semantics (Fillmore 1976; 1982), a cognitive approach to semantics, and organized around frames, rather than words, BFN is a cognitive-based resource that builds on two interfaces: the syntax-semantics interface, and the corpus-lexicon one. More recently, the Global FrameNet initiative (Torrent et al. 2018) seeks to address questions following the development of framenets for languages other than English, and to provide a multi-lingual workbench including parallel corpora and tools for the alignment of framenets in various languages.

The proposed paper describes work still in progress aimed at (a) developing a Greek FrameNet, (b) aligning it to BFN 1.7 at the frame level and (c) linking it to an existing Greek semantic lexicon. To do so, we opted for creating our resource from scratch and then aligning/linking it with external resources manually; moreover, our approach to lexicon building was a modular one, in that we considered specific domains (i.e., finance) and semantic fields (i.e., cognition, communication) as our starting point. Following the specifications already set by BFN, lexicon development is based on corpus evidence. In this regard, Greek verbs and nouns (both single- and multi-word ones) in the selected domains/semantic fields were manually extracted from corpora and then grouped together based on their meaning. As expected, word sense disambiguation was performed in the case of polysemous predicates resulting, thus, in the determination of the Lexical Units (LUs) of the respective frames. Greek frames were, thus, created from scratch; consequently, these frames were further populated with more LUs where appropriate. Syntactic/semantic considerations were also detailed in view of corpus evidence. To this end, a corpus specifically created for accompanying the final lexical resource was created; semantic role labelling was, then, performed manually by trained linguists.

In the paper, we will present the lexical resource in terms of the Database developed and its contents, the methodology employed towards lexicon creation and annotation, and the results obtained thereof. We will then discuss cross-lingual considerations that arise from our attempt to align the Greek FrameNet to BFN. The focus will be on the steps taken towards linking the Greek FrameNet with the semantic dictionary of Modern Greek.

References (indicative)

- Fillmore, C. J. and Baker, C. F. 2010. *A Frames Approach to Semantic Analysis*. In Bernd Heine et al., editors, *Oxford Handbook of Linguistic Analysis*, pages 313–341.
- Fillmore, C. J. 1976. *Scenes-and-frames Semantics*. In *Linguistic Structures Processing. Fundamental Studies in Computer Science*, vol. 59. North Holland Publishing: 55-81.
- Fillmore, C. J. 1982. *Frame Semantics*. *Linguistics in the Morning Calm*, ed. The Linguistic Society of Korea. Selected papers from SICOL-1981: 111-137.
- Fontenelle, T. (editor) 2003. *International Journal of Lexicography*, Special Issue on FrameNet, volume 16. Oxford University Press.
- Torrent, T., Elsworth, M., Baker, C., da Silva Matos, E. E. 2018. *The Multilingual FrameNet Shared Annotation Task: a Preliminary Report*. In *The International FrameNet Workshop 2018: Multilingual FrameNets and Constructions*