

Encoding MWEs in a conceptual lexicon of Modern Greek

Angeliki Fotopoulou, Stella Markantonatou and Voula Giouli

Institute for Language and Speech Processing, "Athena" RIC

{[afotop; marks; voula](mailto:afotop@ilsp.athena-innovation.gr)}@ilsp.athena-innovation.gr

WP1

1 Introduction

We report on work in progress aimed at the development of a conceptual lexicon for the Greek language and the encoding of MWEs in it. This lexicon builds on an existing resource, *Ekfrasis* (Markantonatou and Fotopoulou, 2007; Markantonatou et al, 2010) that is currently being revised and expanded further in order to be suitable for a number of NLP applications.

2 The conceptual lexicon EKFRASIS

Ekfrasis is a conceptually organised dictionary that capitalizes on two basic notions: (a) the notion of *lexical fields*; and (b) the Saussurian notion of *sign* and its two inseparable facets, namely, the SIGNIFIER and the SIGNIFIED as the building blocks (main classes) of the underlying ontology.

In this sense, the intended language resource is a linguistic ontology and *words* are instances in the SIGNIFIER class. At this level, morphological, syntactic and functional information about lemmas is encoded. Similarly, *word meanings* are instances in the SIGNIFIED class. These are further organised on the basis of *lexical fields*. Each instance in the SIGNIFIER class is mapped onto a concept, the latter represented as an instance in the SIGNIFIED class.

The *Instances* of the class SIGNIFIER are specified for (a) features pertaining to lexical semantic relations (i.e., synonymy, antonymy); (b) lexical relations such as word families, allomorphs, syntactic variants etc.; and (c) morphosyntactic properties (PoS, gender, declension, argument structure, word specific information etc.).

3 MWEs in Ekfrasis

Multi-word expressions are treated in *Ekfrasis* in a systematic way using consistent lexicographic criteria and practices. Our schema caters for verb, noun, adjective and adverb classes of MWEs; syntactic configurations pertaining to each class are also represented as distinct sub-classes hierarchically organized. More precisely, a typology of Greek verbal MWEs has been defined in (Fotopoulou 1993; Mini, 2009) and of nominal MWEs in (Anastasiadis, 1986; Fotopoulou et al, 2008) on the basis of the lexical and syntactic configurations involved. Fixed Subject and object complements of verbal MWEs are also taken into account. This typology has been mapped onto a hierarchy under the class SIGNIFIER and its sub-class MWEs (as opposed to the class Simple Lexical Units).

One of the crucial aspects in MWEs recognition and classification is to model the kind and degree of *fixedness* (M. Gross, 1988; G. Gross, 1996; Sag et al, 2002; Lamiroy, 2003). In our approach, the main distinction into collocations and fixed MWEs is initially made (Fig. 1). The degree and type of fixedness are then encoded as features for the Nominal MWEs : (a) *lexical invariance* (**N_lexical_variance=no** in the example given in Fig. 1 *ταξίδι*"trip" *αστραπή*"lightning"), and (b) *full_metaphor: Is_Actually*. The feature *Is_Actually* (Fig.1) encodes information about the interpretation pattern, i.e., whether the meaning of the MWE is non-compositional, partially compositional (results from the meanings of at least one of its constituents) or compositional. In Fig. 1, the MWE has the meaning of the first noun *ταξίδι* "trip" and is encoded in as *collocation*.

Moreover, the standard set of features inherited from the class SIGNIFIER is also retained (PoS, Gender, Number, Tense, synonyms, antonyms, etc.). Finally, the MWE is linked to its meaning that is an instance in the SIGNIFIED class (a concept) with the *Instantiates* feature.

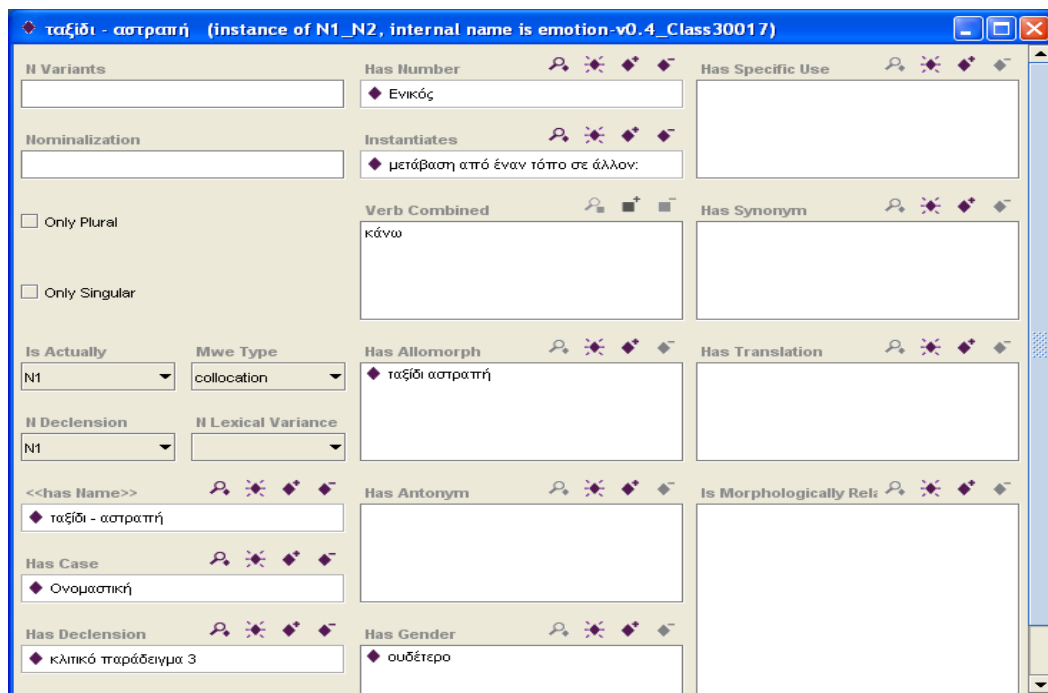


Fig. 1 Encoding MWEs in Ekfrasis: the MWE noun *ταξίδι – αστραπή* (=sudden trip)

4 Conclusions and future research

We have presented an overview of the conceptual lexicon *Ekfrasis* and the treatment of MWEs in it. We have so far treated noun and verbal MWEs. Future work involves the population of the lexicon with new MWEs also pertaining to adjectives and adverbs and the definition of a fine-grained typology for the new categories. Compatibility of the resource with diverse syntactic approaches will also be investigated. The evaluation of the final resource will be performed by integrating it in a tool that automatically recognizes MWEs in texts.

References

- Αναστασιάδη-Συμεωνίδη, Άννα. 1986. *Η Νεολογία στην Κοινή Νεοελληνική*, Θεσσαλονίκη. ΕΕΦΣ του ΑΠΘ.
- Baldwin, Timothy and Kim, Su Nam Kim. 2010. Multiword Expressions. In N. Indurkha and F. J. Damerau (Eds.), *Handbook of Natural Language Processing* (2nd ed.).
- Fotopoulou, Angeliki. 1993. *Une Classification des Phrases a Complements Figes en Grec Moderne*. Doctoral Thesis, Universite Paris VIII.
- Fotopoulou, A., Giannopoulos, G., Zourari, M., Mini, M. 2008. "Automatic recognition and extraction of multiword nominal expressions from corpora". In *Proceedings of the 29th Annual Meeting, Department of Linguistics, Aristotle University of Thessaloniki, Greece* (in Greek)
- Gross, Gaston. 1996. Les expressions figées en français. Noms composés et autres locutions. Paris/Gap: Ophrys.
- Gross, Maurice. 1988. Les limites de la phrase figée. *Langage* 90: 7-23
- Lamiroy, B. 2003. Les notions linguistiques de figement et de contrainte. *Linguisticae Investigationes* 26:1, 53-66, Amsterdam/Philadelphia: John Benjamins.
- Markantonatou, Stella and Fotopoulou, Aggeliki. 2007. The tool "Ekfrasi". In *Proceedings of the 8th International Conference on Greek Linguistics, The Lexicography Workshop*. Ioannina, Greece.

- Markantonatou, Stella, Fotopoulou, Aggeliki, Mini, Marianna. & Alexopoulou, Maria. 2010. In search of the right word. In *Proceedings of Cogalex-2: Cognitive Aspects of the Lexicon, 2nd SIGLEX endorsed Workshop*. Beijing.
- Mini, M. 2009. *Linguistic and Psycholinguistic Study of Fixed Verbal Expressions with Fixed Subject in Modern Greek: A Morphosyntactic Analysis, Lexicosemantic Gradation and Processing by Elementary School Children*. Unpublished doctoral dissertation. University of Patras.
- Sag, Ivan A., Timothy Baldwin, Francis Bond, Ann Copestake and Dan Flickinger. 2002. Multiword Expressions: A Pain in the Neck for NLP. LinGO Working Paper No. 2001-03. In Alexander Gelbukh, ed., (2002) *Proceedings of CICLING-2002*. Springer.