

# A systematic study on the fixedness degree of verbal multiword expressions: application to Modern Greek and French

Matthieu Constant (1,2) and Angeliki Fotopoulou (3)

(1) Université Paris-Est, LIGM, France (2) Alpage Team, INRIA (3) ILSP R.C “Athena”, Greece  
[Mathieu.Constant@u-pem.fr](mailto:Mathieu.Constant@u-pem.fr) [afotop@ilsp.athena-innovation.gr](mailto:afotop@ilsp.athena-innovation.gr)

(WG1)

Multiword expressions are complex linguistic phenomena (Sag et al. 2002) as they have multidimensional properties and varying degree of compositionality (Nunberg et al. 1994). This poster aims at presenting a preliminary work on the identification of the fixedness degree of multiword expressions (MWE). We specifically focus on Modern Greek and French as lexical MWE database exist with a larger coverage of encoded syntactic-semantic properties of verbal MWE that are organized in the form of a lexicon-grammar (M. Gross 1986). Our study is built on the work of M. Gross (1984) for French and the work of Fotopoulou (1993) and Mini (2009) for Modern Greek. Our final objective is to build a model representing the different fixedness cases, by means of a spectrum of linguistic features.

For our study, we constructed a small set of 65 verbal multiword expressions for both languages. The extracted expressions have different syntactic structures in order to test the fixedness degree with respect to a series of tests. These tests are those which, in principle, enable to define a fixed expression like more and more studies show such as in Vincze (2011), Sailer and Wintner (2014), Stone (2015). Our criteria are the following:

1. **Lexical criteria:** Fixedness can be identified by testing whether there exists a paradigmatic break, e.g. while having *Max casse (le jouet+le verre) à Marie* with the compositional meaning *Max breaks Marie's (toy+glass)* [lit. *Max breaks the (toy+glass) to Marie*], we get *Max a cassé (les pieds +\*le jouet+\*le verre) à Marie (Max lui a cassé les pieds)* with the meaning ‘*Max gets on Marie's nerves*’ [lit. *Max broke the (feet +\*toy+\*glass) to Marie*]; These criteria enables to evaluate the exclusive co-occurrence of expression components.
2. **Morphosyntactic criteria:** non-regular restrictions apply on the determiner distribution or on the morphological variants (e.g. number), as well as on some transformations like passivation or pronominalization. For instance, *Μου κόπηκαν τα ήπατα* (me.PRO-GEN cut.V-3-PL-PASS the.ART-DEF liver.N-PL-ACC with the meaning *I was very frightened*) verbal expression do not allow any modification over the fixed constituents: *Μου κόπηκε το ήπαρ* (me.PRO-GEN cut.V-3-SG PASS the.ART-DEF liver.N-SG-ACC).
3. **Semantic criteria:** the meaning of the expression is non-compositional, i.e. it is not predictable from the meaning of its components. Similarly to Mini et al. (2011), we examined independently the verb and the nominal arguments of the expressions with tests like
  - a. the element keeps its literal meaning: e.g. *Luc nage dans le bonheur*, lit. *Luc swims in the happiness*, with the meaning *Luc is happy* (the complement *bonheur* keeps its literal meaning)
  - b. the element has a metaphorical meaning, or is an extension of the literal meaning : *μου ράγισε την καρδιά* me.PRO-GEN broke.V-3SG the.ART-DEF heart.N-ACC (meaning. *he broke my heart*)
  - c. the element has nothing to do with the literal meaning: e.g. in Modern Greek *η Μαρία τα φόρτωσε στον κόκορα*, the.ART-DEF Maria.N-NOM them.PRO-ACC loaded.V-3-SG to-the.PREP-ART-DEF cock.N-SG-ACC, with the meaning. *Maria gave up (the activity)*.

In practice, we systematically implemented all these criteria for the 65 expressions both in Modern Greek and French. We also used standard tests to identify support verb constructions. They are specific multiword expressions as the nominal element keeps its literal meaning, while the verb has a neutral value. We examined three criteria: reduction in a noun phrase with deletion of the support verb, nominalization of the expression (*prendre une décision* [lit. *take a decision*, with the meaning *make a decision*] = *prise de décision* [meaning *decision making*]), unique relation of the nominal element to the subject. We considered that encoding such tests would contribute to model the fixedness degree.

All expressions are encoded in the form of a table: a row corresponds to a lexical entry (a multiword expression), a column corresponds to a lexical, syntactic or semantic property like in M. Gross (1986). In addition of the different properties listed above, we also encoded the prepositions selected by the multiword expression (e.g. *Luc casse la gueule à Max*, litt. *Luc breaks the mouth to Max*, with the meaning *Luc smashes Max's face*). As stated in the first paragraph, most of the examined expressions belong to large-scale lexical database where complementary syntactic and lexical properties were encoded. A specific column is reserved to put the reference to the lexica in order to be analyzed later on. In total, for French and Modern Greek, we have 37 columns: 10 for lexical elements (a column for a given syntactic position), 2 for lexical criteria, 12 for morphosyntactic/syntactic properties, 6 for prepositional selection, 6 for semantic criteria, 1 for reference to existing database.

The poster will be devoted to present these tables, which encode the degree of fixedness for verbal multiword expressions in Modern Greek and French, with a spectrum of linguistic features. From the study of 65 sentiment expressions with different syntactic structures, we observed different correlations between these features. For instance, morphosyntactic restrictions and absence of transformations seem to correlate with the paradigmatic break (mainly on the verbal element) and with the semantic information on the verb (metaphoric meaning or no relationship with literal meaning). This pilot study is a preliminary step towards a better linguistic understanding of the fixedness degree of multiword expressions.

## References

- Fotopoulou A. 1993. Une classification des phrases à compléments figés en grec moderne - étude morphosyntaxique des phrases figées. Thèse de Doctorat. Paris: Université Paris VIII - St. Denis.
- Gross, M. 1984. Une classification des phrases "figées" du français, *Revue Québécoise de Linguistique*, 11.2, 151-185.
- Gross M. 1986. "Lexicon-Grammar. The Representation of Compound Words". Proceedings of the 11th conference on Computational linguistics (COLING '86)
- 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
- Mini, M., Diakogiorgi, K. & Fotopoulou, A. 2011. What can children tell us about idiomatic phrases' fixedness: the psycholinguistic relevance of a linguistic model?. *DISCOURS (Revue de linguistique, psycholinguistique et informatique)* <http://discours.revues.org/8579>.
- Nunberg, G., Sag, I., Wasow, T. 1994. Idioms. *Language* 70: 491-538.
- Sag, I. A., Baldwin, T., Bond, F., Copestake, A., and Flickinger, D. 2002. *Multiword Expressions: A Pain in the Neck for NLP*. In: *Lecture Notes in Computer Science*, Vol. 2276, pp. 1-15.
- Sailer M. and Wintner S. 2014. Multiword expressions: linguistic properties and lexical representation. *Slides from the PARSEME Prague Training School*.
- Stone M. 2015. Systematic flexibility in Verb-Object Idioms. *Proceedings of the 8th Brussels Conference on Generative Linguistics*.
- Vincze V. 2011. *Semi-Compositional Noun + Verb Constructions: Theoretical Questions and Computational Linguistic Analyses*. PhD thesis, University of Szeged, August 2011.