

# Database for free subject verb MWEs

Stella Markantonatou<sup>1</sup>, Eri Koletti<sup>2</sup>, Elpi Margariti<sup>2</sup>, Panagiotis Minos<sup>1</sup>, Aimilia Stripeli<sup>2</sup>,  
Giorgos Zakis<sup>2</sup>, Niki Samaridi

<sup>1</sup>Institute for Language and Speech Processing/“Athena” RIC, [marks@ilsp.gr](mailto:marks@ilsp.gr),  
[pminos@gmail.com](mailto:pminos@gmail.com)

<sup>2</sup>National and Kapodistrian University of Athens, [erk7@yahoo.gr](mailto:erk7@yahoo.gr), [elpimargariti@gmail.com](mailto:elpimargariti@gmail.com),  
[astripeli@gmail.com](mailto:astripeli@gmail.com), [georgizak@gmail.com](mailto:georgizak@gmail.com), [nsamaridi@gmail.com](mailto:nsamaridi@gmail.com)

## WP1

### 1 Introduction

We have developed a database (DB) that accommodates a wide range of linguistic information about free subject verb MWEs drawing on a classification of free subject verb MWEs of Modern Greek (Samaridi & Markantonatou 2014) and of English (Sailer 2014). The DB has been designed to provide facilities for:

1. The description of those grammatical features and parts of a string that single it out as a MWE, also the provision of information about possible specialized modifiers; this information is meant to support, among others, parsing in different set ups
2. Encoding lexicographic information: MWE meaning, relations among different MWEs
3. The development of corpora of correct and incorrect strings containing the MWEs.

### 2 Description of the grammatical features of MWEs

Only strings with no compositional meaning are admitted in the DB. They are thoroughly examined to obtain a full description of their identifying features: the form and number of the necessary words in the string and the form and number of optional lemmata (such as modifiers) that may occur with the MWE. Furthermore, this investigation reveals

-whether any WWSs exist

-the form, number and syntactic relations among the phrasal constituents occurring in the string

-the existence of deeper syntactic dependencies such as control/ binding relations

-whether internal or external modification of the MWE is possible and how

We encode the syntactic properties of the MWEs in usage and not the “underlying syntax” that was applied to create the MWE. These two sets of syntactic properties may be identical or intersect or not intersect. We have not classified MWEs in terms of their flexibility because the distinction among flexible, semi-flexible and non-flexible MWEs seems to bear little importance for parsing once the notion of WWS is taken into account by the grammar.

#### 2.1 Discovering constituents, WWSs and deep syntactic relations

With “free subject” verb MWEs different subjects trigger agreement on the verb as corpus derived information shows. The borders and the necessary parts of the MWE are defined. Corpus data provide grammatical structures and introspection ungrammatical structures. Word order

permutations indicate phrasal constituents of the MWE, some of which may be WWSs. Corpora hardly offer any word order information, so we resort to introspection data. The intervening XPs of any type and function offer (i) information corroborating constituency information from word order permutations, and (ii) information about WWSs. Control and binding dependencies may form a crucial feature of the MWE identity and are therefore examined and listed.

## **2.2 Detailed examination of the WWSs**

As a WWS we consider a set of lemmata in fixed order with no or limited declination freedom. WWSs undergo some special scrutiny: whether they decline and whether they accept specialized modifiers. Grammatical structures are obtained from corpora and ungrammatical ones from introspection. If the WWS is an NP we examine its ability to accept a variety of determiners, whether it can be replaced by a clitic (weak personal pronoun), whether it allows for clitic doubling, whether it can be replaced by a definite/indefinite pronoun, whether it can form the target of an anaphoric clause and, whether it can be assigned any reference at all.

## **3 Lexicographic information**

### **3.1 Meaning of the MWE**

A MWE may be assigned a meaning that is related through some modifier to another MWE that has been assigned a “basic meaning”. All MWEs are assigned a non-compositional meaning.

### **3.2 Variations**

A MWE may have more than one form that are listed as variations of the MWE, eg a small set of different verb heads may be allowed, all other parameters, including meaning, being equal.

### **3.3 Semantic relations among MWEs**

The following relations among MWEs are encoded in the DB: synonymy, “opposite”, semantic pairs, verb alternations. An explanation of some terms used above follows: “Opposite”: when the meaning of a MWE is the negation of another. “Semantic pairs”: the members of certain verb pairs stand in a causative-non causative relation but they do not share the same root. Often MWEs having these verbs as heads stand in the same type of relation. “Verb alternations”: MWEs more frequently participate to the passive and the causative-inchoative alternation. Often, the WWS of the causative MWE turns up as a fixed subject of the inchoative MWE.

## **4 The corpora**

The DB currently contains ~300 Modern Greek MWEs drawn from a list of ~1120 MWEs (Samaridi 2014, <http://users.sch.gr/samaridi/attachments/article/3/Lexical%20Resources.pdf>). The text corpora accommodated in the DB contain both grammatical and ungrammatical strings featuring MWEs. The grammatical strings are drawn from the (HNC) and from Google while the ungrammatical ones have been evaluated by native speakers (introspection). We also plan to accommodate English MWEs, probably in some meaning correspondence with the Greek ones.

## **5 References**

Manfred Sailer. 2014. MWE Template: English

[http://wiki.studiumdigitale.unifrankfurt.de/FB10\\_Parseme/index.php/MWE\\_Template:\\_English](http://wiki.studiumdigitale.unifrankfurt.de/FB10_Parseme/index.php/MWE_Template:_English).

Niki Samaridi and Stella Markantonatou. 2014. Classification of Modern Greek Verb MWEs. Frankfurt Workshop on Multi-word Expressions, co-hosting the [PARSEME 3rd general meeting](#)