

## REPORT ON THE RESEARCH CONDUCTED DURING THE STSM

### 1. THE PROBLEM

While developing IDION [www.idion.ilsp.gr](http://www.idion.ilsp.gr), an environment for encoding a wide range of MWE properties addressed to both humans and machines, several questions presented themselves concerning, among others, the sources of MWEs and the validity of the information that is eventually encoded.

It is clearly the case that Modern Greek sources of MWEs are few. The reliable sources are the established lexica and a small number of dedicated collections but their coverage is less than satisfactory. Furthermore, MWEs that function as adverbials or pps (the distinction is not that clear in MG where composite prepositions are very frequent and are used to introduce constituents with an adverbial function) have not really attracted the attention of researchers so far. On the other hand, they are underrepresented in the established lexica.

The meaning of MWEs is a more serious issue. It clearly varies depending on the context and lexicographical definitions, when they exist, do not really capture the variation.

### 2. THE RESULTS OF THE STSM

The work conducted with this STSM provided a good number (more than 200) of MWEs that find adverbial usages. These are MWEs of high validity because they have been retrieved from corpora and have been validated by four native speakers, all of them with linguistic training. Furthermore, the paraphrases will provide more MWE material, after it has been evaluated by native speakers. Therefore, source-wise this STSM was an important opportunity to develop a method for retrieving fixed MWEs and the results are satisfactory.

The paraphrase part of the work, that was the main topic of the STSM, was most important. We have been able to classify paraphrases of MWEs according to syntactic environment. We used three general syntactic patterns, namely *\_within an NP\_*, *\_S OR VP\_* and *\_complement of a copula\_*. Not all MWEs avail themselves to all the three contexts indicated above and not all of them can be paraphrased in all contexts. A MWE that occurs in all the contexts along with the validated paraphrases per context is given in Table 1.

	<b>within an NP</b>	<b>S OR VP</b>	<b>complement of a copula</b>
εις βάθος	στόχος αυτής της μελέτης είναι να παρουσιαστεί μια ~ κριτική επανεξέταση της πολιτικής	η επιθυμία μας ήταν να δούμε την μόδα " ~ " και όχι επιφανειακά	αυτό αποδεικνύει ότι οι έλεγχοι είναι ~ και συστηματοποιημένοι
	βαθύτητα	βαθιά.ADV	σε βάθος

	<b>εκ βάθρων</b>	<b>εκ βάθρων</b>	εμπειριστατωμένος
	εμπειριστατωμένος	<b>σε βάθος</b>	ενδεδεχής
	ενδεδεχής		διεξοδικός
	διεξοδικός		
	βαθιά.ADV		
	<b>σε βάθος</b>		

Table 1. Validated paraphrases for the MWE εις βάθος ‘in depth’.

The retrieved paraphrases for the MWE εις βάθος (‘in depth’) are illustrated in Table 1 for each context. The results contain two candidate MWEs (boldfaced). We note that σε βάθος is the Modern Greek counterpart of the (archaic in form) MWE εις βάθος—this type of pair is frequently observed in our data.

Another observation concerning the paraphrases of these fixed MWEs is about the position of the MWE in the sentence, namely whether the MWE introduces the sentence or it is placed within it. Table 2 shows a MWE whose paraphrases are sensitive to its position in the sentence. In the first example, the MWE για παράδειγμα (‘for example’) introduces the sentence, in the second it is placed within it. The paraphrases for the in-the-sentence position form a subset of the paraphrases allowed for the position introducing the sentence.

για παράδειγμα	~ , το 16 % των εσόδων της εθνικής τράπεζας προέρχεται από τις δραστηριότητές της στο εξωτερικό
	ως παράδειγμα
	όπως λόγου χάρη
	παράδειγμα
	ένα παράδειγμα
	όπως για παράδειγμα
	παραδείγματος χάρη
	παραδείγματος χάριν
	λόγου χάρη
	οι τραυματισμοί , ~ , συχνά οδηγούν σε προσωρινή ακινησία
	ως παράδειγμα
	παράδειγμα
	παραδείγματος χάρη
	παραδείγματος χάριν
	λόγου χάρη

Table 2. Meanings of the MWE για παράδειγμα ‘for example’ according to sentence position.

Therefore, IDION is now called to encode these facts:

1. Meaning of a MWE according to the syntactic context
2. Meaning of a MWE according to position in the sentence

IDION has been designed to encode verb MWEs that function as predicates supporting independent sentences. The MWEs studied with this STSM do not form independent sentences although some of them do select nominal complements. How exactly these facts are going to be incorporated in IDION is a matter of research that will be conducted in the immediate future.

Last but not least, the STSM indicated that lexicographic MWE resources are of value for parsing since the improvement in the performance of the parser was more or less retained despite the fact that the cleaned MWEs were considerably less in number than the ones

contained in the set of the automatically retrieved MWEs (before their validation by native speakers). A richer list of verified MWEs may improve parser performance further. On the other hand, one should consider the semantic accuracy that verified MWEs offer to NLP.

### **3. IMMEDIATE FUTURE WORK**

We plan to:

1. Retrieve more paraphrases
2. Evaluate the retrieved paraphrases as regards their MWE status
3. Establish the syntactic properties of the MWEs
4. Encode all this information in IDION
5. Compare our data with the data available in established lexica and collections
6. Publish this work

## **Confirmation by the host institution of the successful execution of the STSM**

Dr. Markantonatou reports:

The work conducted with this STSM provided a good number (more than 200) of adverbial/prepositional Greek MWEs that were retrieved from corpora and validated by four native speakers. In addition, the retrieved paraphrases have provided approximately another 200 MWEs. In order to select valid paraphrases, we used three syntactic patterns, namely *\_within an NP\_*, *\_S OR VP\_* and *\_complement of a copula\_*. Four native speakers validated the obtained paraphrases in a substitution test and provided interesting results as regards the conditions that affect the meaning of adverbial/prepositional MWEs of Greek. The automatically acquired and manually validated MWEs and their paraphrases will be integrated into IDION, a web based lexicographic tool hosted by ILSP/“Athena” RC, initially designed to encode verb MWEs. Encoding adverbial/prepositional MWEs may result in the modification of the IDION encoding schema; this research will be conducted in the immediate future.

In the overall, this STSM produced a satisfactory method for retrieving fixed MWEs with 2 or 3 words and their paraphrases according to syntactic position, thus providing validated ‘definitions’ of MWEs. Dr Apidianaki worked on ILSP’s premises and the overall cooperation was smooth and intensive.