

STSM Report - ENeL COST Action

Paraphrases as meaning descriptions for MultiWord Expressions in a Greek lexicographic resource

Marianna Apidianaki
LIMSI, CNRS, Université Paris-Saclay, Orsay, France
marianna.apidianaki@limsi.fr

General

Host institution: Institute for Language and Speech Processing
ILSP / “Athena” Research Center, Athens, Greece

Dates: July 15, 2016 - August 6, 2016

1 Work carried out during the STSM

The goal of this STSM has been to use paraphrases as glosses describing the meaning of MultiWord Expressions (MWEs) in a Greek lexicographic resource. The resource, called IDION, is dedicated to Greek idioms and other types of MWEs and contains manually encoded content regarding their definitions, translations and usage examples, morpho-syntactic descriptions and lexico-semantic relations.¹ The goal of the proposed STSM was twofold: (1) to enrich IDION with MWEs automatically extracted from corpora, and (2) to work on a systematic way of constructing the definition of a MWE from automatically acquired paraphrases. Apart from providing additional content for this specific resource, this work puts forward a semi-automatic methodology for resource enrichment that could be useful in other lexicographic projects.

For enriching IDION, we exploited Greek MWEs automatically acquired from parallel corpora using a data-driven approach that combines translation information with shallow syntactic knowledge (Apidianaki et al., 2017).² The MWEs were of high quality due to a number of constraints applied during extraction: a) the use of two language pivots (English and French) instead of one, which was the case in previous work (Moirón and Tiedemann, 2006), following the assumption that word sequences having non-compositional meaning would tend to be translated consistently in different languages, and b) a syntactic filter which discarded expressions that did not correspond to syntactic constituents. In Apidianaki et al. (2017), the quality of the automatically built MWE databases was evaluated extrinsically, through its integration in a Greek dependency parser (Prokopidis and Papageorgiou, 2014). During the ENeL STSM, we complemented this work with an intrinsic evaluation where four annotators manually checked the quality of the automatically extracted MWEs to be included into IDION. MWEs that were judged as good ones (i.e. having a stable structure and non compositional meaning) by four annotators were retained and the reduced lists were fed to the parser again in order to test the impact of these cleaner MWE lists on parser performance compared to the previous setting.

¹The IDION resource can be accessed here: <http://idion.ilsp.gr>

²The MWE identification method was developed during a Parseme COST STSM. The report is available at: http://typo.uni-konstanz.de/parseme/images/STSM-reports/STSMreport_Apidianaki.pdf

	Adverbial MWEs			Prepositional MWEs		
	Responses	%	Agreement	Responses	%	Agreement
0Y	121	40,88	0Y or 4Y: 54,73 %	226	45,29	0Y or 4Y: 59,72 %
4Y	41	13,85		72	14,43	
3Y - 1N	34	11,49	4Y & 3Y: 25,34%	71	14,2	4Y & 3Y: 28,66%
2Y - 2N	39	13,18		56	11,2	
1Y - 3N	61	20,61		74	14,9	
	296			499		

Table 1: Distribution of annotations for the adverbial and prepositional MWEs.

For enriching the resource with meaning descriptions (glosses), we retrieved paraphrases of the Greek MWEs from the Paraphrase Database (PPDB) (Ganitkevitch and Callison-Burch, 2014).³ PPDB is an automatically extracted database containing millions of paraphrases in 16 different languages, including Greek. The paraphrases included in this resource have been extracted from parallel corpora using the pivot paraphrasing method of Bannard and Callison-Burch (2005) and can be lexical, phrasal or syntactic. The database contains a very high number of phrases, not all of which have a linguistic status. To retrieve material that would be relevant for IDION, we solely extracted information available for the retained MWEs and discarded noisy paraphrases by applying a number of filtering mechanisms. The quality of the retained paraphrases was manually validated by two annotators.

The work was carried out in collaboration with Stella Markantonatou, the inviting person in the host institution (ILSP / “Athena” R.C.), and graduate and postgraduate students in Linguistics (University of Athens, Laboratory of Linguistics of the Department of Mediterranean Studies at the University of the Aegean) who performed the annotation. We present the results of this STSM in the next Section.

2 Results

2.1 Validation of automatically extracted MWEs

We conducted a manual evaluation of the adverbial and prepositional MWE candidates that were extracted by Apidianaki et al. (2017) from the Europarl corpus (Koehn, 2005). The candidate MWEs were identified based on one-to-many automatic alignments (i.e. Greek word sequences aligned to one foreign (English or French) word were considered as candidates) and the candidate lists were cleaned using a syntactic filter which discarded MWEs that didn’t correspond to syntactic constituents. Four annotators were asked to judge the quality of the MWEs (assign a “Yes” or “No” value to each of them) with respect to their stable structure and non-compositional meaning.

The retained adverbial MWE candidates were 296 in total. The annotators were more or less strict in their judgments: the positive answers (“Yes”, the candidate is a good MWE) ranged from 89 (30%) to 117 (40%) for different annotators. In average, 103 MWEs (34,79%) were judged to be correct by the annotators. Table 1 gives a more detailed picture of the agreement between the annotators. The first two rows contain information about MWEs that were rejected or accepted by all annotators (i.e. that were assigned 0 or 4 “Yes” values). The annotators agreed on the goodness of the MWEs in 54,73% of the cases. The other rows correspond to MWEs where there was disagreement between the annotators (for instance, MWEs that were assigned 2 “Yes” and 2 “No” values) which correspond to 45,28% of the cases. We retained the MWEs that were judged as good ones by at least three annotators for inclusion into IDION. We observed that the accepted MWEs often corresponded to constructs inherited from older or ancient variations of Greek which are not used compositionally in everyday Modern Greek.

³The PPDB resource is freely available and can be found at: <http://paraphrase.org>

	Mateparser (graph-based)			Maltparser (transition-based)		
	LAS	UAS	LACC	LAS	UAS	LACC
No conversion	82.65	88.45	89.69	79.76	85.27	88.42
MWEs ≥ 5	82.75 ($\downarrow 0.13$)	88.44 ($\downarrow 0.05$)	89.90 ($\downarrow 0.08$)	79.96 ($\downarrow 0.02$)	85.38 ($\downarrow 0.03$)	88.58 ($\downarrow 0.10$)
MWEs ≥ 2	82.48 ($\uparrow 0.17$)	88.25 ($\uparrow 0.22$)	89.63 ($\downarrow 0.01$)	80.00 ($\downarrow 0.07$)	85.37 ($\downarrow 0.02$)	88.59 ($\downarrow 0.02$)
all MWEs	82.48 ($\downarrow 0.21$)	88.40 ($\downarrow 0.37$)	89.62 ($\uparrow 0.01$)	80.20 ($\downarrow 0.06$)	85.50 ($\uparrow 0.03$)	88.70 ($\downarrow 0.06$)

Table 2: Parser performance when using the whole lists of adverbial and prepositional MWEs obtained through two language pivots vs. the cleansed lists of manually validated MWEs. In parentheses, we show whether the use of the cleansed lists resulted in an improvement or a deterioration in parser performance.

... δεν προστατεύουν την τοπική αυτοδιοίκηση όταν μας στέλνουν μια επιστολή **άνευ περιεχομένου**.
... δεν προστατεύουν την τοπική αυτοδιοίκηση όταν μας στέλνουν μια επιστολή **κενή περιεχομένου**.
... δεν προστατεύουν την τοπική αυτοδιοίκηση όταν μας στέλνουν μια επιστολή **άνευ νοήματος**.
... δεν προστατεύουν την τοπική αυτοδιοίκηση όταν μας στέλνουν μια επιστολή **ανούσια**.
... δεν προστατεύουν την τοπική αυτοδιοίκηση όταν μας στέλνουν μια επιστολή **χωρίς φορτίο**.
... δεν προστατεύουν την τοπική αυτοδιοίκηση όταν μας στέλνουν μια επιστολή **άνευ σημασίας**.
... δεν προστατεύουν την τοπική αυτοδιοίκηση όταν μας στέλνουν μια επιστολή **άδειο**.
... δεν προστατεύουν την τοπική αυτοδιοίκηση όταν μας στέλνουν μια επιστολή **χωρίς νόημα**.

Table 3: Paraphrases of the MWE *άνευ περιεχομένου* substituted for the original phrase in sentential context.

3 Evaluation through integration in the Greek dependency parser

To evaluate the impact of the manually validated MWEs on parsing performance, we fed the cleansed MWE lists to two dependency parsers trained for Greek: the transition-based Maltparser (Nivre et al., 2007) and the graph-based Mateparser (Bohnet, 2010). We compared the results to the ones reported by Apidianaki et al. (2017) where the entire automatically extracted MWE lists were used. We preserved the same experimental setting as in Apidianaki et al. (2017): we focused on adverbial and prepositional MWEs and used a recently extended version of the Greek Dependency Treebank (GDT) (Prokupidis et al., 2005) for training. If the components of a MWE were found as a sequence of tokens in a sentence of the train or test partitions of the GDT, we examined whether the sequence constituted a sub-tree. If yes, we joined the words of the sequence using underscores, assigned an adverb part-of-speech tag to the newly created token and attached it as an adverbial modifier of the governor of the original subtree.

The results presented in Table 2 show the performance of each parser when it had no access to MWE information (“No conversion” setting); when it exploited information on MWEs occurring more than 5, or more than 2, times in the corpus (MWEs ≥ 5 and ≥ 2); and when it used all extracted adverbial and prepositional MWEs regardless of their frequency. We report the Labelled and Unlabelled Attachment Scores (LAS and UAS) and the Label Accuracy (LACC) obtained by the parsers.⁴ As shown by the results of the previous experiments shown in the table, the use of the automatically extracted MWE lists resulted in consistent improvements of the transition-based Maltparser compared to the setting where it had no access to MWE knowledge. The numbers in parentheses show the impact of using the manually cleansed lists on parser performance. The direction of the arrows denotes whether the use of the cleansed MWEs resulted in an improvement or a deterioration of the parser. For instance, the use of the cleansed list of MWEs with frequency over 5 resulted in a deterioration of 0.13 in the LAS of the Mateparser, while the cleansed MWEs with frequency over 2 improved the parser’s LAS by 0.17. In the previous experiments, the Mateparser gave lower results than the Maltparser but it seems to be helped by the cleansed MWE lists. We intend to explore this pattern in the future in order to further improve the graph-based parser.

⁴The LAS corresponds to the percentage of tokens that are assigned a correct head and a correct dependency type. The UAS corresponds to the percentage of tokens that are assigned a correct head, and the LACC corresponds to the percentage of tokens with the correct dependency.

MEANING

5	All of the meaning of the original phrase is retained, and nothing is added
4	The meaning of the original phrase is retained, although some additional information may be added but does not transform the meaning
3	The meaning of the original phrase is retained, although some information may be deleted without too great a loss in the meaning
2	Substantial amount of the meaning is different
1	The paraphrase doesn't mean anything close to the original phrase

GRAMMAR

5	The sentence with the paraphrase inserted is perfectly grammatical, and would require no correction
4	The sentence is grammatical, but might be a little awkward sounding
3	The sentence has an agreement error (such as between its subject and verb, or between a plural noun and singular determiner)
2	The sentence has multiple errors and/or omits words that would be required to make it grammatical
1	The sentence is totally ungrammatical

Table 4: Paraphrase evaluation scales

	Meaning	Grammar		Meaning	Grammar
Adverbial MWEs	5: 0.29	5: 0.38	Prepositional MWEs	5: 0.26	5: 0.33
	4: 0.02	4: 0.04		4: 0.07	4: 0.05
	3: 0.03	3: 0.04		3: 0.04	3: 0.07
	2: 0.09	2: 0.01		2: 0.08	2: 0.05
	1: 0.57	1: 0.54		1: 0.56	1: 0.49

Table 5: Distribution of the paraphrase substitutions over scores.

3.1 MWE paraphrase extraction

Paraphrases were extracted from the different PPDB packages for the validated set of MWEs. The PPDB resource comes in different sizes (S, M, L, XL, XXL, XXXL); smaller packages contain high precision paraphrases while bigger ones contain all paraphrase pairs, for highest recall. The number of paraphrases doubles with each increase in size, and larger sizes subsume smaller ones. We chose to use the XXL package of Greek phrasal paraphrases which, on the one hand, has a good coverage and, on the other hand, contains a reasonable amount of noise that could be automatically reduced. We focused on the adverbial and prepositional MWEs that were used in the parsing experiments. Not all the MWEs were covered in the PPDB. We retrieved paraphrases for 17 adverbial and 90 prepositional MWEs. We filtered out paraphrases ending by a conjunction, a personal pronoun, and tense and subjunctive particles, and beginning with indefinite and definite articles and relative pronouns. To discard morphological variations of the same phrase, we considered using an edit distance filter that identified paraphrases differing by one or two characters. We finally decided to keep the paraphrase variants as these could be useful for automatic processing.

The quality of the retained paraphrases was evaluated using a substitution mechanism, following Bannard and Callison-Burch (2005), and standard criteria used in paraphrase evaluation.⁵ We gathered sentences containing each MWE from the Hellenic National Corpus (HNC)⁶ or the web and substituted it with the retained paraphrases in each sentential context. Table 3 shows some of the substitutions performed for the MWE *άνευ περιεχομένου* in a sentence using its paraphrases (*κενή περιεχομένου*, *άνευ νοήματος*, *χωρίς νόημα*, etc). The annotators were asked to judge the generated sentences according to two dimensions: a) whether the paraphrase retains the same meanings as the original phrase, and b) whether the sentence with the paraphrase substituted into it remains grammatical. The sentences were assigned two values using the 5-point scales described in Table 4. We consider an item to have the same meaning as the original MWE if it was assigned a score of 3 or greater, and to be grammatical if it was

⁵The criteria used are detailed here: <http://www.cs.jhu.edu/~ccb/howto-extract-paraphrases-page-2.html>

⁶The HNC is available here: <http://hnc.ilsp.gr>

assigned a score of 4 or 5 (Callison-Burch, 2008).

The number of paraphrases available for each MWE varied a lot. We retained a maximum of 20 paraphrases for each MWE and substituted them in 3 sentences extracted from the HNC. All items were scored by two annotators regarding the meaning and grammaticality aspects. Due to time constraints, we performed this evaluation for the 17 adverbial MWEs and for 47 out of the 90 prepositional MWEs. For the adverbial MWEs, there were a total of 98 sentences that paraphrases were substituted into, with a total of 1160 judgments collected (580 per aspect). For prepositional MWEs, there were 272 contexts (5.79 per phrase in average) with 3846 judgments in total. We measured inter-annotator agreement (IAA) using the Kappa statistic (Carletta, 1996). IAA was very high, reaching $\kappa = 0.862$ for adverbial MWEs and $\kappa = 0.588$ for prepositional MWEs. If we measure agreement in terms of how often the annotators both judged an item to be above or below the meaning and grammaticality thresholds, then agreement is $\kappa = 0.889$ for adverbial and $\kappa = 0.698$ for prepositional MWEs. This shows that estimating the goodness of the paraphrases is an easy task for the human annotators. As shown in Table 5, the majority of the paraphrases were assigned a 1 or 5 score while intermediate values were rarely selected, which makes it straightforward to sort out the good and bad paraphrases of the MWEs.

4 Future collaboration with the host institution

This STSM has paved the way for future collaboration with members of the host institution on several topics. In a lexicographic perspective, the identified paraphrases will serve as meaning descriptions of the automatically extracted and manually validated MultiWord Expressions that will be included in the IDION resource. Apart from being good quality glosses that illustrate the meaning of the MWEs, the paraphrases will also constitute individual entries by themselves increasing the coverage of the resource. The integration will be carried out at a later stage, once that the XML specifications of the IDION resource is adapted to include this additional knowledge.

Moreover, the paraphrases of the automatically acquired MWEs will be exploited for improving syntactic and semantic processing in several applications (parsing, summarization, translation). They will serve to achieve a generalisation at the sense level and handle out-of-vocabulary words (i.e. not seen in the training data), improving the robustness of NLP applications. Other future extensions will involve the exploitation of the extracted MWEs and their paraphrases for cross-lingual knowledge transfer and Greek Semantic Role Labelling (van der Plas et al., 2014).

5 Foreseen publications

The MWE identification methodology, the parsing experiments and the obtained results are described in a book chapter that has been accepted for publication in the “Special issue on MWEs in Greek and other languages: from theory to implementation” of the Bulletin of Scientific Terminology and Neologisms of the Academy of Athens, to appear in 2017. The MWE identification and paraphrase glossing methodologies provide means for including high quality automatically acquired content in lexicographic resources. We intend to prepare a publication presenting the automatic stages of this procedure as well as the human intervention steps that are needed for the developed methodology to be useful in other lexicographic projects.

References

Apidianaki, M., Prokopidis, P., and Papageorgiou, H. (2017). Combining Cross-lingual and Syntactic Evidence for Greek MultiWord Expression Identification. In *Special issue on MWEs in Greek and other languages: from theory to implementation, Bulletin of Scientific Terminology and Neologisms of the Academy of Athens*, Athens, Greece.

- Bannard, C. and Callison-Burch, C. (2005). Paraphrasing with Bilingual Parallel Corpora. In *Proceedings of ACL*, pages 597–604, Ann Arbor, Michigan.
- Bohnet, B. (2010). Top Accuracy and Fast Dependency Parsing is not a Contradiction. In *Proceedings of the 23rd International Conference on Computational Linguistics (COLING 2010)*, pages 89–97, Beijing, China.
- Callison-Burch, C. (2008). Syntactic constraints on paraphrases extracted from parallel corpora. In *Proceedings of the 2008 Conference on Empirical Methods in Natural Language Processing*, pages 196–205, Honolulu, Hawaii. Association for Computational Linguistics.
- Carletta, J. (1996). Assessing agreement on classification tasks: The kappa statistic. *Computational Linguistics*, 22(2):249–254.
- Ganitkevitch, J. and Callison-Burch, C. (2014). The multilingual paraphrase database. In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)*, Reykjavik, Iceland.
- Koehn, P. (2005). Europarl: A Parallel Corpus for Statistical Machine Translation. In *Proceedings of MT Summit*, pages 79–86, Phuket, Thailand.
- Moirón, B. V. and Tiedemann, J. (2006). Identifying idiomatic expressions using automatic word alignment. In *Proceedings of the EACL 2006 Workshop on Multiword Expressions in a multilingual context*, pages 33–40, Trento, Italy.
- Nivre, J., Hall, J., Nilsson, J., Chanev, A., Eryigit, G., Kübler, S., Marinov, S., and Marsi, E. (2007). Maltparser: A language-independent system for data-driven dependency parsing. *Natural Language Engineering*, 13(2):95–135.
- Prokopidis, P., Desyprí, E., Koutsombogera, M., Papageorgiou, H., and Piperidis, S. (2005). Theoretical and Practical Issues in the Construction of a Greek Dependency Treebank. In *Proceedings of The Fourth Workshop on Treebanks and Linguistic Theories (TLT 2005)*, pages 149–160, Barcelona, Spain.
- Prokopidis, P. and Papageorgiou, H. (2014). Experiments for Dependency Parsing of Greek. In *Proceedings of the SPMRL-SANCL Workshop*, pages 90–96, Dublin, Ireland.
- van der Plas, L., Apidianaki, M., and Chen, C. (2014). Global Methods for Cross-lingual Semantic Role and Predicate Labelling. In *Proceedings of COLING 2014*, pages 1279–1290, Dublin, Ireland.