

Identification of Modern Greek collocations with Fips

Eric Wehrli, Luka Nerima, Vasiliki Foufi



UNIVERSITÉ DE GENÈVE



LATL
LABORATOIRE D'ANALYSE ET DE TECHNOLOGIE DU LANGAGE

WG1, WG2, WG3
PARSEME 7th General Meeting
26-27 September 2016
Dubrovnik, Croatia

Objective:

Identification of Modern Greek nominal and verbal collocations

Method:

- Enhancement of the lexical database with lexemes and collocations
- Enhancement of the grammar-based processing model
- Parsing-based extraction of candidate-collocations from a corpus
- Filtering of candidates using standard association measures
- Manual validation of the best candidates
- Addition of validated collocations to the collocation database

Tools:

Fips:

Multilingual parser (French, English, German, Italian, Spanish, Greek, Portuguese, Romanian)

Grammar-based processing model (rules and procedures)

Collocation extraction tool (Seretan 2011):

Identification of collocations in the input sentence during the syntactic analysis (Wehrli et al. 2010)

Syntactic processes may impact the identification of verb-object collocations, especially in morphology-rich languages, like Modern Greek:

- Syntactic flexibility (relatively free order)
- Wh-questions
- Relative clauses
- Passivization
- Pronominalization (replacement of the noun object by a pronoun either in the same or in the next phrase)

Experiments & Results

- Measurement of:
 - the parser's performance to identify lexicalized collocations
 - the impact of the collocation knowledge on the performance of the parser (in percentage of complete analyses)
- Newspaper corpus (20,000 words) (<http://www.greek-language.gr/greekLang/index.html>, <http://www.clarin.gr/en>)
- Manual identification of 638 collocations
- Parsing results:
 - 1st analysis: 43.26% of complete analyses, 124 identified collocations
 - 2nd analysis after enrichment of the lexical database: 44.33% of complete analyses, 482 identified collocations
- 100% precision in the collocation identification task, recall 75.54%, F-measure 0.86
- 156 collocations not identified due to incomplete parsing

Importance

- Extension of FipsGreek lexical coverage:
 - 18,342 lexemes
 - 21,121 collocations
- Integration with parsing:
 - helps lexical disambiguation
 - gives priority to collocation-driven syntactic attachment
- A tool for helping collocation dictionary building
- Available online: latlapps.unige.ch

References

Seretan Violeta. 2011. *Syntax-Based Collocation Extraction*. Springer Verlag.
Wehrli Eric, Seretan Violeta, and Luka Nerima. 2010. "Sentence analysis and collocation identification". In *Proceedings of the Workshop on Multiword Expressions: from Theory to Applications (MWE 2010)*, pp. 27–35, Beijing, China.

Examples of parsing and collocation extraction

Syntactic flexibility

Έκκληση στους διοικητικούς υπαλλήλους να σταματήσουν την απεργία έκανε ο υπουργός.
'An appeal to the administrative staff to stop the strike made the minister.'

Έκκληση	NOUN-SIN-FEM	1	έκκληση		
στους	PREP	9			
διοικητικούς	ADJ-PLU-MAS	15	διοικητικός		
υπάλληλους	NOUN-PLU-MAS	28	υπάλληλος		διοικητικός υπάλληλος
να	CONJ-SUB	39	να		administrative staff
σταματήσουν	VERB-SUB-AOR-3-PLU	42	σταματώ		
την	DET-SIN-FEM	54	ο		
απεργία	NOUN-SIN-FEM	58	απεργία		
έκανε	VERB-IND-IMP-3-SIN	66	κάνω		κάνω έκκληση
ο	DET-SIN-MAS	72	ο		to make an appeal
υπουργός	NOUN-SIN-MAS	74	υπουργός		
.	PUNC-point	82			

Figure 1: Verb-object collocation where the noun object precedes the verb.

Wh-questions

Ποιο ρεκόρ θέλει να σπάσει ο Μελισσανίδης;
'Which record does he want to break Melissanidis?'

Ποιο	DET-SIN-NEU	1	ποιος		
ρεκόρ	NOUN-SIN-NEU	6	ρεκόρ		
θέλει	VERB-IND-PRE-3-SIN	12	θέλω		
να	CONJ-SUB	18	να		
σπάσει	VERB-SUB-AOR-3-SIN	21	σπάζω		σπάζω το ρεκόρ
ο	DET-SIN-MAS	28	ο		to break a record
Μελισσανίδης	NOUN	30	Μελισσανίδης		
.	PUNC-point-virgule	42			

Figure 2: Example of a wh-question.

Relative clauses

Οι ασφαλιστικές εισφορές που καταβάλλονται εξ ολοκλήρου από τους κρατικούς προϋπολογισμούς.
'The insurance contributions that are paid totally by the national budget.'

οι	DET-PLU-FEM	1030	ο		
ασφαλιστικές	ADJ-PLU-FEM	1034	ασφαλιστικός		
εισφορές	NOUN-PLU-FEM	1047	εισφορά		ασφαλιστική εισφορά
που	PRO-REL-PLU-FEM	1056	που	DO	insurance contribution
καταβάλλονται	VERB-IND-SUB-PRE-PASSIVE-3-PLU	1060	καταβάλλω	SU:proDO:εισφορές	καταβάλλω εισφορά
εξ ολοκλήρου	ADV-INT	1074	εξ ολοκλήρου		to pay a contribution
από	PREP	1087	από	ADJUNCT	
τους	DET-PLU-MAS	1091	ο		
κρατικούς	ADJ-PLU-MAS	1096	κρατικός		
προϋπολογισμούς	NOUN-PLU-MAS	1106	προϋπολογισμός		κρατικός προϋπολογισμός
.	PUNC-point	1121			national budget

Figure 3: Example of a phrase containing two nominal collocations and a verb-object collocation where the noun object has been replaced by a relative pronoun.

Passivization

Οι λανθασμένες πολιτικές που ακολουθήθηκαν από όλες τις κυβερνήσεις είναι η πρώτη αιτία.
'The wrong policies that were followed by all the governments is the main cause.'

Οι	DET-PLU-FEM	418	ο	SU	
λανθασμένες	ADJ-PLU-FEM	421	λανθασμένος		
πολιτικές	NOUN-PLU-FEM	433	πολιτική		
που	PRO-REL-PLU-FEM	443	που	DO	
ακολουθήθηκαν	VERB-IND-AOR-PASSIVE-3-PLU	447	ακολουθώ	SU:proDO:πολιτικές	ακολουθώ πολιτική
από	PREP	461	από	ADJUNCT	to follow a policy
όλες	DET-PLU-FEM	465	όλος		
τις	DET-PLU-FEM	470	ο		
κυβερνήσεις	NOUN-PLU-FEM	474	κυβέρνηση		
είναι	VERB-IND-PRE-3-INN	486	είμαι	SU:πολιτικές FO:απία	
η	DET-SIN-FEM	492	ο	FO	
πρώτη	ADJ-SIN-FEM	494	πρώτος		
αιτία	NOUN-SIN-FEM	500	απία		πρώτη αιτία
.	PUNC-point	505			main cause

Figure 4: Example of a verb-object collocation whose verb is in passive voice and the noun has been pronominalized.

Pronominalization

Είχαν υποβάλει ερωτήσεις στον επίτροπο, αλλά κανένας δεν βρισκόταν εδώ για να τις θέσει.
'They had posed questions to the commissioner, but nobody was here to make them.'

Είχαν	AUX-VERB-IND-IMP-3-PLU	1	έχω		
υποβάλει	VERB-SUB-AOR-3-SIN	7	υποβάλλω	SU:proDO:ερωτήσεις PO:επίτροπο	υποβάλλω ερώτηση
ερωτήσεις	NOUN-PLU-FEM	16	ερώτηση		questions
στον	PREP	26		PO	
επίτροπο	NOUN-SIN-MAS	31	επίτροπος		
.	PUNC-virgule	39			
αλλά	CONJ-COO	41	αλλά		
κανένας	PRO-IND-SIN-MAS	46	κανείς	SU	
δεν	adverbe-NEG	54	δεν		
βρισκόταν	VERB-IND-IMP-PASSIVE-3-SIN	58	βρίσκω	SU:κανένας	
εδώ	ADV-LOC	68	εδώ		
για να	CONJ-SUB	72	για να		
τις	PRO-CLI-PLU-FEM	1	του		
θέσει	VERB-SUB-AOR-3-SIN	83	θέτω	DO:τις	θέτω ερώτηση
.	PUNC-point	88			to make a question

Figure 5: Verb-object collocation where the noun has been replaced by a personal pronoun.

Ας αναλογιστούν τις ευθύνες τους. Να τις αναλάβουν.
'Let them consider their responsibilities. Should they take them on.'

Ας	PREP	1111	ας		
αναλογιστούν	VERB-SUB-AOR-PASSIVE-3-PLU	1114	αναλογίζομαι	SU:proDO:ευθύνες	responsibilities
τις	DET-PLU-FEM	1127	ο	DO	
ευθύνες	NOUN-PLU-FEM	1131	ευθύνη		
τους	PRO-CLI-PLU-ING	1131	του		
.	PUNC-point	1143			
Να	CONJ-SUB	1145	να		
τις	PRO-CLI-PLU-FEM	1145	του		
αναλάβουν	VERB-SUB-AOR-3-PLU	1152	αναλαμβάνω	SU:proDO:τις	αναλαμβάνω την ευθύνη
.	PUNC-point	1161			to take on responsibility

Figure 6: Verb-object collocation where the noun in the form pronoun refers to the object of the previous sentence.

Acknowledgements: This work was partly funded by the PARSEME European COST Action (IC1207)