



FACHLITERATUR
EDITION ROMIOSINI
ΕΝΙΣΤΗΜΗ



ICGL12

12th INTERNATIONAL CONFERENCE
ON GREEK LINGUISTICS
16 – 19 SEPTEMBER 2015
FREIE UNIVERSITÄT BERLIN, CEMOG

Proceedings of the ICGL12

vol. 1

The International Conference on Greek Linguistics is a biennial meeting on the study and analysis of Greek (Ancient, Medieval and Modern), placing particular emphasis on the later stages of the language.

PROCEEDINGS OF THE ICGL12
ΠΡΑΚΤΙΚΑ ΤΟΥ ICGL12

**Thanasis Georgakopoulos, Theodossia-Soula Pavlidou, Miltos Pechlivanos,
Artemis Alexiadou, Jannis Androutsopoulos, Alexis Kalokairinos,
Stavros Skopeteas, Katerina Stathi (Eds.)**

**PROCEEDINGS OF THE 12TH INTERNATIONAL
CONFERENCE ON GREEK LINGUISTICS**

**ΠΡΑΚΤΙΚΑ ΤΟΥ 12^{ΟΥ} ΣΥΝΕΔΡΙΟΥ ΕΛΛΗΝΙΚΗΣ
ΓΛΩΣΣΟΛΟΓΙΑΣ**

VOL. 1

© 2017 Edition Romiosini/CeMoG, Freie Universität Berlin. Alle Rechte vorbehalten.
Vertrieb und Gesamtherstellung: Epubli (www.epubli.de)
Satz und Layout: Rea Papamichail / Center für Digitale Systeme, Freie Universität Berlin
Gesetzt aus Minion Pro
Umschlaggestaltung: Thanasis Georgiou, Yorgos Konstantinou
Umschlagillustration: Yorgos Konstantinou

ISBN 978-3-946142-34-8
Printed in Germany

Online-Bibliothek der Edition Romiosini:
www.edition-romiosini.de

Στη μνήμη του *Gaberell Drachman* (†10.9.2014)
και της Αγγελικής Μαλικούτη-*Drachman* (†4.5.2015)
για την τεράστια προσφορά τους στην ελληνική γλωσσολογία
και την αγάπη τους για την ελληνική γλώσσα

ΣΗΜΕΙΩΜΑ ΕΚΔΟΤΩΝ

Το 12ο Διεθνές Συνέδριο Ελληνικής Γλωσσολογίας (International Conference on Greek Linguistics/ICGL12) πραγματοποιήθηκε στο Κέντρο Νέου Ελληνισμού του Ελεύθερου Πανεπιστημίου του Βερολίνου (Centrum Modernes Griechenland, Freie Universität Berlin) στις 16-19 Σεπτεμβρίου 2015 με τη συμμετοχή περίπου τετρακοσίων συνέδρων απ' όλον τον κόσμο.

Την Επιστημονική Επιτροπή του ICGL12 στελέχωσαν οι Θανάσης Γεωργακόπουλος, Θεοδοσία-Σούλα Παυλίδου, Μίλτος Πεχλιβάνος, Άρτεμις Αλεξιάδου, Δώρα Αλεξοπούλου, Γιάννης Ανδρουτσόπουλος, Αμαλία Αρβανίτη, Σταύρος Ασημακόπουλος, Αλεξάνδρα Γεωργακοπούλου, Κλεάνθης Γκρώμαν, Σαβίνα Ιατρίδου, Mark Janse, Brian Joseph, Αλέξης Καλοκαιρινός, Ναπολέων Κάτσος, Ευαγγελία Κορδώνη, Αμαλία Μόζερ, Ελένη Μπουτουλούση, Κική Νικηφορίδου, Αγγελική Ράλλη, Άννα Ρούσου, Αθηνά Σιούπη, Σταύρος Σκοπετέας, Κατερίνα Στάθη, Μελίτα Σταύρου, Αρχόντω Τερζή, Νίνα Τοπιντζή, Ιάνθη Τσιμπλή και Σταυρούλα Τσιπλάκου.

Την Οργανωτική Επιτροπή του ICGL12 στελέχωσαν οι Θανάσης Γεωργακόπουλος, Αλέξης Καλοκαιρινός, Κώστας Κοσμάς, Θεοδοσία-Σούλα Παυλίδου και Μίλτος Πεχλιβάνος.

Οι δύο τόμοι των πρακτικών του συνεδρίου είναι προϊόν της εργασίας της Εκδοτικής Επιτροπής στην οποία συμμετείχαν οι Θανάσης Γεωργακόπουλος, Θεοδοσία-Σούλα Παυλίδου, Μίλτος Πεχλιβάνος, Άρτεμις Αλεξιάδου, Γιάννης Ανδρουτσόπουλος, Αλέξης Καλοκαιρινός, Σταύρος Σκοπετέας και Κατερίνα Στάθη.

Παρότι στο συνέδριο οι ανακοινώσεις είχαν ταξινομηθεί σύμφωνα με θεματικούς άξονες, τα κείμενα των ανακοινώσεων παρατίθενται σε αλφαβητική σειρά, σύμφωνα με το λατινικό αλφάβητο· εξαίρεση αποτελούν οι εναρκτήριες ομιλίες, οι οποίες βρίσκονται στην αρχή του πρώτου τόμου.

Η Οργανωτική Επιτροπή του ICGL12

ΠΕΡΙΕΧΟΜΕΝΑ

Σημείωμα εκδοτών	7
Περιεχόμενα	9
 Peter Mackridge: <i>Some literary representations of spoken Greek before nationalism(1750-1801)</i>	17
Μαρία Σηφianού: <i>Η έννοια της ευγένειας στα Ελληνικά</i>	45
Σπυριδούλα Βαρλοκώστα: <i>Syntactic comprehension in aphasia and its relationship to working memory deficits</i>	75
 Ευαγγελία Αχλάδη, Αγγελική Δούρη, Ευγενία Μαλικούτη & Χρυσάνθη Παρασχάκη-Μπαράν: <i>Γλωσσικά λάθη τουρκόφωνων μαθητών της Ελληνικής ως ξένης/δεύτερης γλώσσας: Ανάλυση και διδακτική αξιοποίηση</i>	109
Κατερίνα Αλεξανδρή: <i>Η μορφή και η σημασία της διαβάθμισης στα επίθετα που δηλώνουν χρώμα</i>	125
Eva Anastasi, Ageliki Logotheti, Stavri Panayiotou, Marilena Serafim & Charalambos Themistocleous: <i>A Study of Standard Modern Greek and Cypriot Greek Stop Consonants: Preliminary Findings</i>	141
Anna Anastassiadis-Symeonidis, Elisavet Kiourti & Maria Mitsiaki: <i>Inflectional Morphology at the service of Lexicography: ΚΟΜΟΛεξ, A Cypriot Morphological Dictionary</i>	157

Γεωργία Ανδρέου & Ματίνα Τασιούδη: <i>Η ανάπτυξη του λεξιλογίου σε παιδιά με Σύνδρομο Απνοιών στον Ύπνο</i>	175
Ανθούλα- Ελευθερία Ανδρεσάκη: <i>Ιατρικές μεταφορές στον δημοσιογραφικό λόγο της κρίσης: Η οπτική γωνία των Γερμανών</i>	187
Μαρία Ανδριά: <i>Προσεγγίζοντας θέματα Διαγλωσσικής Επίδρασης μέσα από το πλαίσιο της Γνωσιακής Γλωσσολογίας: ένα παράδειγμα από την κατάκτηση της Ελληνικής ως Γ2</i>	199
Spyros Armostis & Kakia Petinou: <i>Mastering word-initial syllable onsets by Cypriot Greek toddlers with and without early language delay</i>	215
Julia Bacsakai-Atkari: <i>Ambiguity and the Internal Structure of Comparative Complements in Greek</i>	231
Costas Canakis: <i>Talking about same-sex parenthood in contemporary Greece: Dynamic categorization and indexicality</i>	243
Michael Chiou: <i>The pragmatics of future tense in Greek</i>	257
Maria Chondrogianni: <i>The Pragmatics of the Modern Greek Segmental Markers</i>	269
Katerina Christopoulou, George J. Xydopoulos & Anastasios Tsangalidis: <i>Grammatical gender and offensiveness in Modern Greek slang vocabulary</i>	291
Aggeliki Fotopoulou, Vasiliki Foufi, Tita Kyriacopoulou & Claude Martineau: <i>Extraction of complex text segments in Modern Greek</i>	307
Αγγελική Φωτοπούλου & Βούλα Γιούλη: <i>Από την «Έκφραση» στο «Πολύτροπο»: σχεδιασμός και οργάνωση ενός εννοιολογικού λεξικού</i>	327
Marianthi Georgalidou, Sofia Lampropoulou, Maria Gasouka, Apostolos Kostas & Xanthippi Foulidi: <i>“Learn grammar”: Sexist language and ideology in a corpus of Greek Public Documents</i>	341
Maria Giagkou, Giorgos Fragakakis, Dimitris Pappas & Harris Papageorgiou: <i>Feature extraction and analysis in Greek L2 texts in view of automatic labeling for proficiency levels</i>	357

Dionysis Goutsos, Georgia Fragaki, Irene Florou, Vasiliki Kakousi & Paraskevi Savvidou: <i>The Diachronic Corpus of Greek of the 20th century: Design and compilation</i>	369
Kleanthes K. Grohmann & Maria Kambanaros: <i>Bilectalism, Comparative Bilingualism, and the Gradience of Multilingualism: A View from Cyprus</i>	383
Günther S. Henrich: „Γεωγραφία νεωτερική“ στο Λίβιστρος και Ροδάμνη: μετατόπιση ονομάτων βαλτικών χωρών προς την Ανατολή;.....	397
Noriyo Hoozawa-Arkenau & Christos Karvounis: <i>Vergleichende Diglossie - Aspekte im Japanischen und Neugriechischen: Veretäten - Interferenz</i>	405
Μαρία Ιακώβου, Ηριάννα Βασιλειάδη-Λιναρδάκη, Φλώρα Βλάχου, Όλγα Δήμα, Μαρία Καββαδία, Τατιάνα Κατσίνα, Μαρίνα Κουτσουμπού, Σοφία-Νεφέλη Κύτρου, Χριστίνα Κωστάκου, Φρόσω Παππά & Σταυριαλένα Περρέα: <i>ΣΕΠΙΜΕ2: Μια καινούρια πηγή αναφοράς για την Ελληνική ως Γ2</i>	419
Μαρία Ιακώβου & Θωμαΐς Ρουσουλιώτη: <i>Βασικές αρχές σχεδιασμού και ανάπτυξης του νέου μοντέλου αναλυτικών προγραμμάτων για τη διδασκαλία της Ελληνικής ως δεύτερης/ξένης γλώσσας</i>	433
Μαρία Καμηλάκη: «Μαζί μου ασχολείσαι, πόσο μαλάκας είσαι!»: Λέξεις-ταμπού και κοινωνιογλωσσικές ταυτότητες στο σύγχρονο ελληνόφωνο τραγούδι.....	449
Μαρία Καμηλάκη, Γεωργία Κατσούδα & Μαρία Βραχιονίδου: <i>Η εννοιολογική μεταφορά σε λέξεις-ταμπού της ΝΕΚ και των νεοελληνικών διαλέκτων</i>	465
Eleni Karantzola, Georgios Mikros & Anastassios Papaioannou: <i>Lexico-grammatical variation and stylometric profile of autograph texts in Early Modern Greek</i>	479
Sviatlana Karpava, Maria Kambanaros & Kleanthes K. Grohmann: <i>Narrative Abilities: MAINing Russian–Greek Bilingual Children in Cyprus</i>	493
Χρήστος Καρβούνης: <i>Γλωσσικός εξαρχαϊσμός και «ιδεολογική» νόρμα: Ζητήματα γλωσσικής διαχείρισης στη νέα ελληνική</i>	507

Demetra Katis & Kiki Nikiforidou: <i>Spatial prepositions in early child Greek: Implications for acquisition, polysemy and historical change</i>	525
Γεωργία Κατσούδα: <i>Το επίθημα -ούνα στη ΝΕΚ και στις νεοελληνικές διαλέκτους και ιδιώματα</i>	539
George Kotzoglou: <i>Sub-extraction from subjects in Greek: Its existence, its locus and an open issue</i>	555
Veranna Kyprioti: <i>Narrative, identity and age: the case of the bilingual in Greek and Turkish Muslim community of Rhodes, Greece</i>	571
Χριστίνα Λύκου: <i>Η Ελλάδα στην Ευρώπη της κρίσης: Αναπαραστάσεις στον ελληνικό δημοσιογραφικό λόγο</i>	583
Nikos Liosis: <i>Systems in disruption: Propontis Tsakonian</i>	599
Katerina Magdou, Sam Featherston: <i>Resumptive Pronouns can be more acceptable than gaps: Experimental evidence from Greek</i>	613
Maria Margarita Makri: <i>Opos identity comparatives in Greek: an experimental investigation</i>	629
2ος Τόμος	
Περιεχόμενα	651
Vasiliki Makri: <i>Gender assignment to Romance loans in Katoitaliótika: a case study of contact morphology</i>	659
Evgenia Malikouti: <i>Usage Labels of Turkish Loanwords in three Modern Greek Dictionaries</i>	675
Persephone Mamoukari & Penelope Kambakis-Vougiouklis: <i>Frequency and Effectiveness of Strategy Use in SILL questionnaire using an Innovative Electronic Application</i>	693

Georgia Maniati, Voula Gotsoulia & Stella Markantonatou: <i>Contrasting the Conceptual Lexicon of ILSP (CL-ILSP) with major lexicographic examples</i>	709
Γεώργιος Μαρκόπουλος & Αθανάσιος Καρασίμος: <i>Πολυεπίπεδη επισημείωση του Ελληνικού Σώματος Κειμένων Αφασικού Λόγου</i>	725
Πωλίνα Μεσηνιώτη, Κατερίνα Πούλιου & Χριστόφορος Σουγανίδης: <i>Μορφοσυντακτικά λάθη μαθητών Τάξεων Υποδοχής που διδάσκονται την Ελληνική ως Γ2</i>	741
Stamatia Michalopoulou: <i>Third Language Acquisition. The Pro-Drop-Parameter in the Interlanguage of Greek students of German</i>	759
Vicky Nanousi & Arhonto Terzi: <i>Non-canonical sentences in agrammatism: the case of Greek passives</i>	773
Καλομοίρα Νικολού, Μαρία Ξεφτέρη & Νίτσα Παραχεράκη: <i>Το φαινόμενο της σύνθεσης λέξεων στην κυκλαδοκρητική διαλεκτική ομάδα</i>	789
Ελένη Παπαδάμου & Δώρης Κ. Κυριαζής: <i>Μορφές διαβαθμιστικής αναδίπλωσης στην ελληνική και στις άλλες βαλκανικές γλώσσες</i>	807
Γεράσιμος Σοφοκλής Παπαδόπουλος: <i>Το δίπολο «Εμείς και οι Άλλοι» σε σχόλια αναγνωστών της Lifo σχετικά με τη Χρυσή Αυγή</i>	823
Ελένη Παπαδοπούλου: <i>Η συνδυαστικότητα υποκοριστικών επιθημάτων με β' συνθετικό το επίθημα -άκι στον διαλεκτικό λόγο</i>	839
Στέλιος Πιπερίδης, Πένυ Λαμπροπούλου & Μαρία Γαβριηλίδου: <i>clarin:el. Υποδομή τεκμηρίωσης, διαμοιρασμού και επεξεργασίας γλωσσικών δεδομένων</i>	851
Maria Pontiki: <i>Opinion Mining and Target Extraction in Greek Review Texts</i>	871
Anna Roussou: <i>The duality of mipos</i>	885

Stathis Selimis & Demetra Katis: <i>Reference to static space in Greek: A cross-linguistic and developmental perspective of poster descriptions</i>	897
Evi Sifaki & George Tsoulas: <i>XP-V orders in Greek</i>	911
Konstantinos Sipitanos: <i>On desiderative constructions in Naousa dialect</i>	923
Eleni Staraki: <i>Future in Greek: A Degree Expression</i>	935
Χριστίνα Τακούδα & Ευανθία Παπαευθυμίου: <i>Συγκριτικές διδακτικές πρακτικές στη διδασκαλία της ελληνικής ως Γ2: από την κριτική παρατήρηση στην αναπλαισίωση</i>	945
Alexandros Tantos, Giorgos Chatziioannidis, Katerina Lykou, Meropi Papatheohari, Antonia Samara & Kostas Vlachos: <i>Corpus C58 and the interface between intra- and inter-sentential linguistic information</i>	961
Arhonto Terzi & Vina Tsakali: <i>The contribution of Greek SE in the development of locatives</i>	977
Paraskevi Thomou: <i>Conceptual and lexical aspects influencing metaphor realization in Modern Greek</i>	993
Nina Topintzi & Stuart Davis: <i>Features and Asymmetries of Edge Geminates</i>	1007
Liana Tronci: <i>At the lexicon-syntax interface Ancient Greek constructions with ἔχειν and psychological nouns</i>	1021
Βίλλυ Τσάκωνα: <i>«Δημοκρατία είναι 4 λύκοι και 1 πρόβατο να ψηφίζουν για φαγητό»:Αναλύοντας τα ανέκδοτα για τους/τις πολιτικούς στην οικονομική κρίση</i>	1035
Ειρήνη Τσαμαδού- Jacobberger & Μαρία Ζέρβα: <i>Εκμάθηση ελληνικών στο Πανεπιστήμιο Στρασβούργου: κίνητρα και αναπαραστάσεις</i> ...	1051
Stavroula Tsiplakou & Spyros Armotistis: <i>Do dialect variants (mis)behave? Evidence from the Cypriot Greek koine</i>	1065
Αγγελική Τσόκογλου & Σύλα Κλειδή: <i>Συζητώντας τις δομές σε -οντας</i>	1077

Αλεξιάννα Τσότσου:

Η μεθοδολογική προσέγγιση της εικόνας της Γερμανίας στις ελληνικές εφημερίδες 1095

Anastasia Tzilinis:

Begründendes Handeln im neugriechischen Wissenschaftlichen Artikel: Die Situierung des eigenen Beitrags im Forschungszusammenhang..... 1109

Κυριακούλα Τζωρτζάτου, Αργύρης Αρχάκης, Άννα Ιορδανίδου & Γιώργος Ι. Ξυδόπουλος:
Στάσεις απέναντι στην ορθογραφία της Κοινής Νέας Ελληνικής: Ζητήματα ερευνητικού σχεδιασμού 1123

Nicole Vassalou, Dimitris Papazachariou & Mark Janse:

The Vowel System of Mišótika Cappadocian 1139

Marina Vassiliou, Angelos Georgaras, Prokopis Prokopidis & Haris Papageorgiou:

Co-referring or not co-referring? Answer the question!..... 1155

Jeroen Vis:

The acquisition of Ancient Greek vocabulary..... 1171

Christos Vlachos:

Mod(aliti)es of lifting wh-questions..... 1187

Ευαγγελία Βλάχου & Κατερίνα Φραντζή:

Μελέτη της χρήσης των ποσοδεικτών λίγο-λιγάκι σε κείμενα πολιτικού λόγου 1201

Madeleine Voga:

Τι μας διδάσκουν τα ρήματα της ΝΕ σχετικά με την επεξεργασία της μορφολογίας 1213

Werner Voigt:

«Σεληνάκι μου λαμπρό, φέγγε μου να περπατώ ...» oder: warum es in dem bekannten Lied nicht so, sondern eben φεγγαράκι heißt und ngr. φεγγάρι

1227

Μαρία Βραχιονίδου:

Υποκοριστικά επιρρήματα σε νεοελληνικές διαλέκτους και ιδιώματα 1241

Jeroen van de Weijer & Marina Tzakosta:

*The Status of *Complex in Greek.....* 1259

Theodoros Xioufis:

The pattern of the metaphor within metonymy in the figurative language of romantic love in modern Greek..... 1275

EXTRACTION OF COMPLEX TEXT SEGMENTS IN MODERN GREEK

Aggeliki Fotopoulou, Vasiliki Foufi, Tita Kyriacopoulou & Claude Martineau
ILSP/RC “Athena”, AUTH, UPEM

afotop@ilsp.athena-innovation.gr, foufi@auth.gr, {tita, martinea}@univ-mlv.fr

Περίληψη

Η παρούσα μελέτη εστιάζει στην αναγνώριση και εξαγωγή σύμπλοκων δομών της νέας ελληνικής. Αρχικά, θα ορίσουμε τον όρο «σύμπλοκη δομή» και θα παρουσιάσουμε τα ιδιοσυγκρασιακά χαρακτηριστικά αυτών των δομών. Στη συνέχεια, θα παρουσιάσουμε τη μέθοδο που εφαρμόζουμε για την αναγνώριση τέτοιων δομών που δεν είναι δυνατό να καταγραφούν σε λεξικά λόγω του απεριόριστου αριθμού και της ετερογενούς φύσης τους. Η μέθοδός μας, η οποία αναπτύχθηκε στο περιβάλλον του αναλυτή κειμένων Unitex, μας επιτρέπει να εντοπίζουμε τις σύμπλοκες δομές στα κείμενα και να εμπλουτίζουμε τα ήδη υπάρχοντα λεξικά με νέες καταχωρίσεις μέσα από τοπικές γραμματικές.

Keywords: Information extraction, complex segments, compound nouns, fixed expressions, named entities, local grammars, electronic dictionaries

1. Introduction

The development of computational linguistic resources (electronic dictionaries and local grammars) for the automatic extraction, identification, and further fine-grained annotation of complex text segments, is the core of our work. We use and extend the notion of multi-word expressions (MWE) by allowing a large description of linguistic objects: compound nouns, named entities (henceforth NE), verbal forms (compound

tenses and negate forms, inference of lexical elements between the constituents, etc.) and frozen expressions (i.e. idioms). Complex segments are very frequent in texts and are characterized by their flexible and heterogeneous nature. As a consequence, they constitute a major challenge for natural language processing (NLP) systems and applications based on corpora analysis and information retrieval.

The information extraction task consists of three different steps: identification of one or more patterns, delimitation and tagging of the identified structures and annotation. In order to proceed to these tasks, we propose a symbolic method based on dictionary graphs that enable users on one hand to create dynamically resources useful for corpus analysis and on the other hand to enrich the existing resources. We construct linguistic resources such as electronic dictionaries, local grammars and dictionary graphs. Our ultimate goal is to elaborate a unique method in order to identify segments that are difficult to enumerate and list (they could also be simple structures e.g. simple proper nouns), to create new resources and to enrich the existing resources in a semi-automatic way.

First, we will define the notion of a ‘complex segment’ and we will describe its special characteristics. Then, we will present our method which enables users to: a) identify complex structures that cannot be lemmatized in dictionaries via local grammars called dictionary graphs and b) create the dictionary of a given text. By applying this method that was developed in the Unitex environment¹ (Paumier 2003), users could also develop “expert” grammars easily reusable.

2. Related work

MWE in Greek, like in several other languages, have been a major focus of interest for many researchers (see, among others, Anastasiadis-Simeonidis 1986, Ralli 1991 and 2008, Panayotopoulou 1992, Fotopoulou 1993, 2014, Daltas 1995, Sklavounou 1993, Ralli and Stavrou 1995 and 1998, Tsopanakis 1998, Holton et al. 1999, Bakakou-Orfanou 2005, Kyriacopoulou 2005, Thomou 2006, Xydopoulos 2008, Klairis and Babinotis 2009, Kyriacopoulou, 2011, Foufi 2014 and others). Many studies have focused on the automatic identification of collocations. The parser developed at the LATL²

1 <http://igm.univ-mlv.fr/~unitex/>

2 Laboratoire d'Analyse et de Traitement du Langage, University of Geneva (Switzerland), <http://www.latl.unige.ch/>

(Wehrli et al. 2000, Goldman et al. 2001) identifies nominal and verbal collocations with a special focus on syntactic flexibility. Frantzi (2003) applied the C-value method on a specialized corpus in the aim to extract collocations and to enrich a specialized dictionary of Modern Greek. Kalampokis and Kyriacopoulou (2006) based on the assumption that literary forms are constituents of MWE, proposed a hybrid method for the extraction of Greek MWE. In the framework of the project Iatrolexi (Iordandou et al. 2007), 20 rules based on definite-clause grammars were developed for the identification and extraction of multiword biomedical terms. A hybrid method for the automatic extraction of nominal MWEs was constructed in the framework of the project 'Ekfrasis' (Fotopoulou et al. 2008). According to this method, grammar rules and statistics measures are combined not only to obtain as many MWE as possible but also to assure the consistency of the results. Linardaki et al. (2010) have created a dictionary of MWE based on automatic extraction and manual validation of candidate MWE. Finally, Samaridi and Markantonatou (2014) have integrated fixed verbal MWE in an LFG grammar of Modern Greek.

3. Complex segments

Complex segments are defined as MWE of great complexity. This notion comprises:

a) compound nouns, e.g.

- (1) πιστωτική κάρτα
credit card
'credit card'

b) fixed expressions, e.g.

- (2) κάνω τα στραβά μάτια
do the blind eyes
'turn a blind eye'

c) named-entities, e.g.

- (3) Βρετανικό Μουσείο,
British Museum
'British Museum'

- (4) Ολυμπιακοί Αγώνες
Olympic Games
'Olympic Games'

d) complex verbal forms (compound tenses, negative forms) like:

- (5) έχω κάνει
have done
'I have done'

- (6) δεν έχω κάνει
not have done
'I haven't done'

- (7) δεν θα έχω κάνει
not will have done
'I won't have done'

- (8) δεν θα έχω ποτέ κάνει
not will have never done
'I will have never done'

3.1. *Compound nouns*

Compound nouns are very frequent in natural languages and according to many studies 70% of a native speaker's vocabulary consists of more compound than simple nouns. Compound nouns play a significant role in NLP and Information Extraction tasks and they have been largely studied both by linguists and computer scientists not only because they are numerous but also because they present specific particularities like morphosyntactic variants, elimination of constituents, creation of abbreviations, and others.

Like simple nouns, compound nouns can be lemmatized in an electronic dictionary. In Modern Greek, the most frequent categories of compound nouns are the adjective-noun:

- (9) ψυχρός πόλεμος
cold war
'cold war'

and noun-noun in genitive:

- (10) φακός επαφής
lens contact
'contact lens'

As already mentioned, compound nouns are subject to a variety of morphosyntactic varieties. In particular, adjective-noun compounds may be abbreviated in acronyms. For instance, the abbreviated form AE of the compound noun ανώνυμη εταιρία (11) is more frequent in texts than its extended form.

- (11) ανώνυμη εταιρία
anonymous company
'incorporated company'

3.2. *Named entities*

In 1995, the Sixth Conference on Message Understanding (MUC-6) was held to evaluate the performance of information extraction systems. During the design of this evaluation, the phrase “named entity” was coined to describe textual references to real-world entities, and tasks were developed in the aim to automatically recognize mentions of named entities (NE) (now known as “named entity recognition” or NER).

NE could be defined as a segment of a document expressing a direct reference to a unique identifier. These unique identifiers of entities could describe organizations, persons, locations, times (dates, times), and quantities (monetary values, percentages) (Chinchor 1998). For instance, the following segments could be considered as NE:

- (12) 14 Δεκεμβρίου 2013
14 of December 2013
'14th December 2013'

(13) 25 ημέρες

25 days

‘25 days’

(14) 25°C βαθμούς Κελσίου

25°C degrees of Celsius

‘25°C Celsius degrees’

(15) υπηρεσιακή πρωθυπουργός Βασιλική Θάνου

interim Prime Minister Vasiliki Thanou

‘interim Prime Minister Vasiliki Thanou’

(16) Λεκανοπέδιο της Αττικής

Basin of Attica

‘Attica Basin’

(17) Τράπεζα της Ελλάδας

Bank of Greece

‘Bank of Greece’

(18) Μητροπόλεως 1, Θεσσαλονίκη, Τ.Κ. 54124

Mitropoleos 1 Thessaloniki P.S. 54124

‘1, Mitropoleos St. Thessaloniki, P.S. 54124’

On one hand, the great number of existing NE and the constant creation of new NE and on the other hand their high degree of possible combinations make them difficult to identify. Moreover, the delimitation of a NE is often difficult to achieve. In order to facilitate their identification, NE are classified into categories (and sub-categories). The following example comprises NE of four different categories:

(19) Η αεροπορική επίθεση προκάλεσε τουλάχιστον 11 νεκρούς και 12 τραυματίες

The air strike caused at least 11 dead and 12 wounded

σε αεροδρόμιο της Υεμένης την Τρίτη 12 Μαρτίου

in airport of Yemen on Tuesday 12 of March

‘The air strike caused at least 11 deaths and 12 casualties in Yemen airport Tuesday, 12th March.’

The table below displays the NE of the previous example (19) according to their category:

Named entities	EXTRACTED INFORMATION
DATE	την Τρίτη 12 Μαρτίου on Tuesday 12 of March ‘Tuesday, 12 th March’
PLACE	σε αεροδρόμιο της Υεμένης in airport of Yemen ‘in Yemen airport’
EVENT	αεροπορική επίθεση air strike ‘air strike’
IMPACT	τουλάχιστον 11 νεκρούς και 12 τραυματίες at least 11 dead and 12 wounded ‘at least 11 deaths and 12 casualties’

Table 1 | Named entities of the example 19

However, a NE could comprise other NE. In particular, NE describing location could be part of an organization, e.g.:

- (20) Συγκοινωνιακό Κέντρο **Αθηνών** και **Κιάτου**
 Transportation Center of Athens and of Kiato
 ‘Transportation Center of Athens and Kiato’

Moreover, organization NE may comprise person NE either simple (only the first or only the last name) or compound (first and last names), e.g.:

- (21) εταιρία του **Καρτσωνάκη**
 company of Kartsonakis (last name)
 ‘Kartsonakis’ company’

- (22) εταιρία του **Ιάκωβου Καρτσωνάκη**
 ‘company of Iakovos (first name) Kartsonakis (last name)’
 ‘Iakovos Kartsonakis’ company’

Organizations may also comprise person and location names such as the organization:

- (23) Περιφερειακό Γενικό Αντικαρκινικό Νοσοκομείο **Μεταξά**
 ‘Regional General Anticancer Hospital **of Metaxas** (Greek last name)
 του Πειραιά
 of Piraeus (location name)
 ‘Regional General Anticancer Hospital Metaxas of Piraeus’

Complex segments that describe organization NE may be formed by two or more complex segments. For example, the NE:

- (24) Ενεργειακό Κέντρο της Περιφέρειας Κρήτης
 Energy Center of the Region of Crete
 ‘Energy Center of the Region of Crete’

is formed by two nominal segments: Ενεργειακό Κέντρο ‘Energy Center’ and Περιφέρεια Κρήτης ‘Region of Crete’. In that case, the second segment is linked to the first with the genitive case (της Περιφέρειας Κρήτης ‘of the Region of Crete’). Equally, the following example also consists of two organizations Πανεπιστημιακή Παιδιατρική Κλινική ‘University Pediatric Clinic’ και Πανεπιστήμιο Αθηνών ‘University of Athens’:

- (25) Πανεπιστημιακή Παιδιατρική Κλινική του Πανεπιστημίου Αθηνών
 University Pediatric Clinic of the University of Athens’
 ‘University Pediatric Clinic of the University of Athens’

NE are susceptible to various types of combinations and their constituents could be found in various forms. For instance, dates in Modern Greek may be formed in various ways:

a) day in letters-number-month in letters, e.g.:

- (26) Πέμπτη 31 Ιουλίου
 Thursday 31st of July

‘Thursday 31st July’;

b) definite determiner-year in numbers, e.g.:

(27) το 1987
in 1987
‘in 1987’;

c) definite determiner-month in letters, e.g.:

(28) τον Μάρτιο
in March
‘in March’

d) day in letters-number-month in letters-year in numbers, e.g.:

(29) Τετάρτη 5 Ιουλίου 1961
Wednesday 5th of July 1961
‘Wednesday 5th July 1961’, etc.

We have to point out that in texts, dates do not always follow the same word order; the month could precede the day as it is displayed in the example 31:

(30) Τα μαθήματα θα ξεκινήσουν τη Δευτέρα 11 Σεπτεμβρίου
‘The lessons will start on Monday 11th of September.
‘The lessons will start on Monday 11th September’.

(31) Τα μαθήματα θα ξεκινήσουν στις 11 Σεπτεμβρίου, Δευτέρα
The lessons will start on 11th of September, Monday’
‘The lessons will start on 11th September, Monday’.

According to the above classification, a system would recognize as dates the structures Δευτέρα 11 Σεπτεμβρίου ‘Monday 11th September’, 11 Σεπτεμβρίου ‘11th September’, Δευτέρα ‘Monday’. However, in order to achieve an accurate extraction of dates, articles should also be considered as constituents of the NE (τη Δευτέρα 11 Σεπτεμβρίου ‘the Monday 11th of September’, στις 11 Σεπτεμβρίου ‘on the 11th September’).³

3 Literal translation.

Frequently, a NE is displayed in the form of an acronym (Anastasiadis-Simeonidis 1986, Clairis and Babiniotis 2009), e.g.:

- (32) Δ.Ν.Τ. - Διεθνές Νομισματικό Ταμείο
'IMF - International Monetary Fund'

3.3. *Compound verbal forms*

Verbal forms could also be found in complex segments, i.e. complex forms of verbs with compound tense, negation, embedded modifiers. In the examples presented below, other lexical forms occur between the constituents of the verbal forms. More precisely, in the example 33 an adverb (ήδη 'already') has intervened between the auxiliary and the main verb. Additionally, in the example 34 two adverbial modifiers κατηγορηματικά 'positively' and επανειλημμένα 'repeatedly' joined by the conjunction και 'and' are found:

- (33) Ο Γιώργος έχει **ήδη** κάνει τα μαθήματά του
The George has already done the homework his
'George has already done his homework.'

- (34) Ο Π. Εφθymiou έχει **κατηγορηματικά** και **επανειλημμένα** δηλώσει πως...
The P. Efthimiou has positively and repeatedly declared that...
'P. Efthimiou has positively and repeatedly declared that...'

In the following example (35), a whole adverbial phrase occurs between the auxiliary and the main verb (καταφέρει 'managed'):

- (35) Ο Γιώργος έχει **προς μεγάλη έκπληξη των γονέων του** καταφέρει να επιτύχει
The George has to the surprise of parents his managed to succeed
'George has to the surprise of his parents managed to succeed.'

Furthermore, the interference of clitics between the constituents of a verbal form may impede the accurate recognition:

- (36) Η Μαρία δεν θα **τους το** είχε ποτέ αποκαλύψει, αν δεν την ανάγκαζαν
 The Maria not would them it have never revealed, if not her forced'
 'Maria would have never revealed it to them, if they did not force her'

All these complex forms that cannot be predicted and listed in dictionaries, complicate the identification and extraction of verbs in texts.

4. Tools and resources

In order to identify complex segments in corpora, we construct linguistic resources. The introduced resources are built using Unitex, an open source, cross-platform and multilingual (French, English, Greek and other 18 languages) corpus processing suite. Unitex tools are designed to perform several NLP tasks on a textual corpus relying on linguistic resources such as electronic dictionaries and local grammars.

4.1. *Electronic dictionaries*

Unitex supports the DELA (Dictionnaires Électroniques du LADL 'LADL electronic dictionaries') standard of lexicon. The electronic dictionaries built for Modern Greek comprise approximately 252 000 entries (~2 000 000 inflected forms) of simple words and around 40 000 compound entries (~162 000 inflected forms). A typical DELA entry (Courtois 1990), shown below, is composed by: 1 inflected form, 2 ,canonical form, 3 .grammatical category, 4 +semantic attributes, 5 :inflectional information (case, number, gender):

- (40) δικηγόρος, δικηγόρος.N+Hum+Profession:Nms:Nfs
 lawyer lawyer

The inflected form is separated by the canonical form (or lemma) by a comma. The canonical form is followed by a full stop and the grammatical category (here N for noun). The syntacticosemantic information follows linked with the symbol '+'. The feature Hum corresponds to human and the additional information about the profession is also given. It is also indicated that that inflected form corresponds to the

nominative case (N), to masculine and feminine gender (m, f) and to singular (s). Compound nouns are displayed in the dictionary in a similar way.

4.2. Local grammars

Graphical representations of local grammars, called graphs under Unitex, are composed by a set of linked boxes. Graph outputs can be displayed as an interactive concordancer or even they can be added or substituted in the recognized text segments.

The graph presented in figure 1 identifies Greek organizations. According to the grammar, an organization in Modern Greek could be formed by common nouns such as οργανισμός 'organization', εταιρία 'company', σύλλογος 'club', τράπεζα 'bank', ομοσπονδία 'federation', ένωση 'corporation', πανεπιστήμιο 'university', όμιλος 'group' that could be determined by adjectives (<A>) and/or followed by location nouns

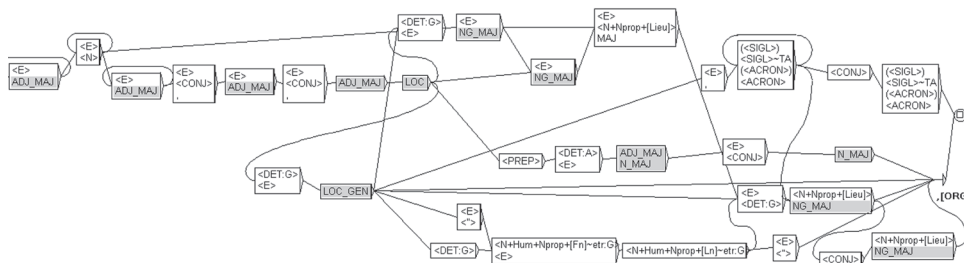


Figure 1 | Graph for the recognition of Greek organizations

νοσηλεύεται εκτός κινδύνου στο [Γενικό Κρατικό Νοσοκομείο της Νίκαιας](#).[\[ORG\]](#),
να καθίσει μαζί τους στο εδώλιο του [Τριμελούς Εφετείου Κακουρηγμάτων Αθηνών](#).[\[ORG\]](#) στις 3
Μαρτίου 2003.
Σύμφωνα με στοιχεία του [Εθνικού Ινστιτούτου για την Πνευματική Υγεία](#).[\[ORG\]](#),
{S} Το [Τεχνολογικό Εκπαιδευτικό Ίδρυμα Δυτικής Μακεδονίας](#).[\[ORG\]](#)
{S} Εισήχθη στο [Περιφερειακό Γενικό Αντικαρκινικό Νοσοκομείο Μεταξά του Πειραιά](#).[\[ORG\]](#),
τον Ιούλιο του 1996.
του [Εθνικού Καρδιολογικού, Πνευμονολογικού και Αιματολογικού Ινστιτούτου των ΗΠΑ](#).[\[ORG\]](#)
την [Εφορεία Αγαθωγών Καταστημάτων της Μητρόπολης Ιωαννίνων](#).[\[ORG\]](#)
{S} Η [Βάλρι Μαργέλ](#) του [Βασιλικού Ινστιτούτου Διεθνών Υποθέσεων του Λονδίνου](#).[\[ORG\]](#)
στη [Θεραπευτική Κλινική της Ιατρικής Σχολής του Πανεπιστημίου Αθηνών](#).[\[ORG\]](#)

Figure 2 | Extract of the identified Greek organizations

(LOC), person names (N+Hum+Nprop), abbreviations (<SIGLES>), acronyms (<ACRON>). This graph also functions as a transducer, i.e. it adds the label ORG (organization) to the extracted segments as it is shown in figure 2.

4.3. Dictionary graphs

The new version of Unitex enables users to develop dictionary graphs, i.e. local grammars which dynamically generate dictionary entries in DELA format. A dictionary graph could be applied into texts like any other electronic dictionary and it generates dynamically new entries based on the text. The following dictionary graph in figure 3 generates date dictionary entries.

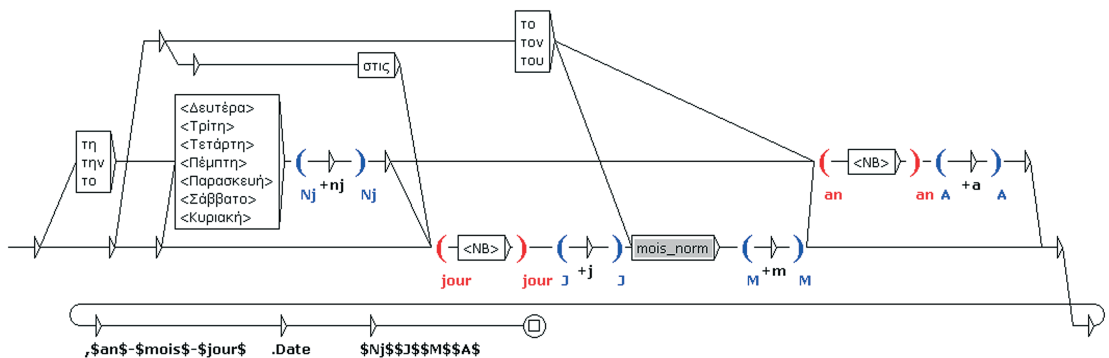


Figure 3 | Dictionary graph

A node in this graph recalls a sub-graph (mois_norm) where the months are normalized. Normalization is the process during which the segments of a text are transformed in structured data. Normalization consists in the attribution of a single representation for each structure (here for dates). In order to achieve that, we have used the Unitex function which recalls information from the dictionaries that are incorporated in the system, with the use of variables. These variables appear in graphs in red (jour 'day', an 'year'). The graph displayed in figure 4 presents all possible forms of months (in numbers or letters) and performs their normalization in numbers.

The dictionary graph after being applied into a text generates dynamically the dictionary of the given text. An extract of such a dictionary is shown in the figure 5.

δεν θα έχει ολοκληρωθεί, **ολοκληρώνομαι**.FV+fl=θα έχει ολοκληρωθεί+fl
 δεν θα έχει ολοκληρώσει, **ολοκληρώνω**.FV+fl=θα έχει ολοκληρώσει+Neg+fl
 δεν θα έχουν επιλέξει, **επιλέγω**.FV+fl=θα έχουν επιλέξει+Neg=δεν+neg
 δεν θα είχα κάνει, **κάνω**.FV+fl=θα είχα κάνει+Neg=δεν+neg
 δεν θα είχα πιάσει, **πιάνω**.FV+fl=θα είχα πιάσει+Neg=δεν+neg
 δεν θα είχαμε χάσει, **χάνω**.FV+fl=θα είχαμε χάσει+Neg=δεν+neg
 δεν θα είχαν προχωρήσει, **προχωρώ**.FV+fl=θα είχαν προχωρήσει+Neg=δεν+neg
 δεν θα είχε υπογράψει, **υπογράφω**.FV+fl=θα είχε υπογράψει+Neg=δεν+neg
 θα έχει ήδη προσφέρει, **προσφέρω**.FV+fl=θα έχει προσφέρει+mod=ήδη
 θα έχει δουλέψει, **δουλεύω**.FV+fl=θα έχει δουλέψει
 θα έχετε πληροφορηθεί, **πληροφορούμαι**.FV+fl=θα έχετε πληροφορηθεί
 θα έχουμε πλέον σωθεί, **σώνομαι**.FV+fl=θα έχουμε σωθεί+mod=πλέον
 θα έχουν ήδη πραγματοποιήσει, **πραγματοποιώ**.FV+fl=θα έχουν πραγματοποιήσει
 θα έχω οριστικά αποφασίσει, **αποφασίζω**.FV+fl=θα έχω αποφασίσει+mod=οριστικά
 θα είχα ίσως ξοδέψει, **ξοδεύω**.FV+fl=θα είχα ξοδέψει+mod=ίσως
 θα είχα σίγουρα πάθει, **παθαίνω**.FV+fl=θα είχα πάθει+mod=σίγουρα
 θα είχα υποβάλει, **υποβάλλω**.FV+fl=θα είχα υποβάλει
 θα είχαμε προσφέρει, **προσφέρνω**.FV+fl=θα είχαμε προσφέρει
 θα είχαν ήδη παραιτηθεί, **παραιτούμαι**.FV+fl=θα είχαν παραιτηθεί+mod=ήδη
 θα είχατε ακούσει, **ακούω**.FV+fl=θα είχατε ακούσει
 θα είχε ήδη παραιτηθεί, **παραιτούμαι**.FV+fl=θα είχε παραιτηθεί+mod=ήδη
 θα είχες μετανοήσει, **μετανοώ**.FV+fl=θα είχες μετανοήσει

Figure 7 | Extracted verbal forms

4.4. “Expert” grammars

The above dictionary graphs developed for the identification of dates and verbal forms could be transformed into more generic grammars and be applied to all categories of NE and to possible combinations of verbal forms.

The graph presented below in figure 8 recalls lexical masks so as to identify various complex segments. Each lexical mask recalls the corresponding text dictionary of the complex segments ‘person’, ‘verbal form (FV)’ and ‘date’. The presence of features leads to the identification of person NE consisting of a first name (prenom) and a last name (nom), a negative form (neg) of a verb (FV) and a date comprising a day (j) and a month (m):

Following this approach, users are capable of reformulating the lexical masks and enriching the resources by improving the corresponding dictionary graphs. We have to

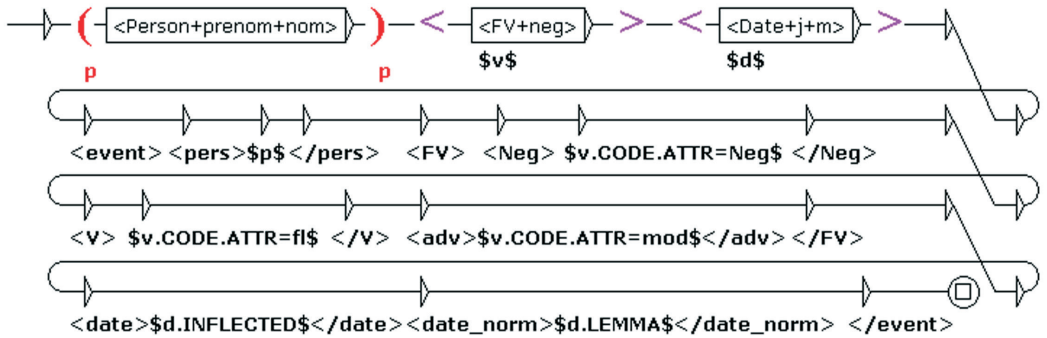


Figure 8 | Graph for the identification of complex segments

point out that the creation of new features does not influence completely the resources that were developed previously.

This way of enriching a local grammar enables the construction of «expert grammars» that recognize complex segments encouraging the elaboration of libraries of resources in a more flexible and modular way. For instance, the analysis of the phrase

- (41) Ο Αλέξης Τσίπρας δεν θα έχει ακόμα επιστρέψει στις 20 Αυγούστου 2015
 The Alexis Tsipras not will have yet returned on the 20th of August 2015
 ‘Alexis Tsipras will not have returned yet on 20th August 2015’

could be fulfilled via the grammar presented above. The analysis will be displayed as in figure 9.

5. Conclusion

The identification of complex segments in corpora is a paramount but also demanding procedure due to their heterogenous, numerous and unpredictable nature. The use of graphs and mostly dictionary graphs that we propose in our method gives the possibility to combine the power and versatility of the local grammars and the expressivity of the electronic dictionaries. It seems to be an effective method to promote the adaptability, reusability and modularity of electronic linguistic resources. Additionally,

```

<event>
    <pers>Αλέξης Τσίπρας</pers>
    <FV>
        <Neg>δεν</Neg>
        <V>θα έχει επιστρέψει</V>
        <adv>ακόμα</adv>
    </FV>
    <date>στις 20 Αυγούστου 2015</date>
    <date_norm>2015-08-20</date_norm>
</event>

```

Figure 9 | Analysis of a phrase

the construction of expert grammars reduces the number of separate graphs which is important for the reuse of the existing resources at all levels (testing, improvement, enrichment etc.). Therefore, quantitative, but mostly qualitative analysis is accurately achieved.

However, given on one hand that the remarks made in this article are part of an on-going study and that there does not exist a reference corpus manually annotated, it is not possible to make a precision and recall evaluation.

References

- Anastasiadis-Simeonidis, Anna. 1986. *I Neologia sti Nea Elliniki*. PhD diss., Thessaloniki: Aristotle University of Thessaloniki.
- Bakakou-Orfanou, Aikaterini. 2005. "I Leksi tis Neas Ellinikis sto Glossiko Sistima kai sto Kimeno". In *Parousia* 65.
- Chinchor, Nancy. 1998. "MUC-7 Named Entity Task Definition (version 3.5)". In *Proceedings of the 7th Message Understanding Conference (MUC-7)*, 19 April-1 May 1998, Fairfax, VA, <http://www.aclweb.org/>

anthology/M98-1028.

- Courtois, Blandine. 1990. “Un Système de Dictionnaires Électroniques pour les Mots Simples du Français”. In *Langue Française* 87, 11–22. Paris : Larousse.
- Daltas, Periklis. 1995. *Dokimia Didaktikis tis Metafrasis. Pros mia Kinonioglossologiki Theorisi*. Athens: Epikairotita.
- Fotopoulou, Aggeliki. 1993. *Une Classification des Phrases à Compléments Figés en Grec Moderne : Étude Morphosyntaxique des Phrases Figées*. PhD diss., Paris : Université Paris VIII.
- Fotopoulou, Aggeliki, Giorgos Giannopoulos, Maria Zourari, and Marianna Mini. 2008. “Aftomati Anagnorisi kai Eksagogi Polilektikon Onomatikon Ekfraseon apo Somata Kimenon”. In *Μελέτες για την Ελληνική Γλώσσα* 29 [*Studies in Greek Linguistics* 29]: 620–633. Thessaloniki: Aristotle University of Thessaloniki.
- Fotopoulou, Aggeliki, Stella Markantonatou, and Voula Giouli. 2014. “Encoding MWEs in a Conceptual Lexicon”. In *MWEs Proceedings of the 10th Workshop on Multiword Expressions (MWE), European Association of Computational Linguistics (EACL)*, 43–47. Gothenburg, Sweden.
- Foufi, Vasiliki. 2014. *Morfologiki, Simasiologiki kai Sintaktiki Perigrifi ton Polilektikon Sintheton Monadon me ti Morfi Epitheto + Ousiastiko. Efarmogi sti Didaktiki tis Ellinikis os Ksenis/Defteris Glossas*. PhD diss., Thessaloniki: Copy City.
- Frantzi, Katerina. 2003. “Updating LSP Dictionaries with Collocational Information”. In *UCREL Technical Papers – Special Issue: Corpus Linguistics 2003*, edited by Dawn Archer, Paul Rayson, Andrew Wilson and Tony Mc Enery, 219–226.
- Goldman, Jean-Philippe, Nerima, Luka, and Eric Wehrli. 2001. “Collocation Extraction Using a Syntactic Parser”. In *Proceedings of the ACL Workshop on Collocation: Computational Extraction, Analysis and Exploitation*, 61–66. Toulouse.
- Holton, David, Peter Mackridge, and Eirini Filippaki-Warburton. 1997. *Greek: A Comprehensive Grammar of the Modern Language*. London: Routledge.
- Iordanidou, Anna, Mavina Pantazara, Eleni Mantzari, Georgios Orfanos, Aristeidis Vaggelatos, and Vasilis Papapanagiotou. 2007. “Zitimata Anagnorisis ton Polilektikon Simplokon Oron ston Tomea tis Vioiatrikis”. In *Anakinoseis tou 6^{ou} Sinedriou “Elliniki Glosa kai Orologia”*, 178–191. Athens: Elliniki Etairia Orologias.
- Kalampokis, Alkis, and Tita Kyriacopoulou. 2006. “Vers une Extraction Automatique des Mots Composés en Grec Moderne: Problèmes et

Perspectives/Approche Linguistique et Statistique”. Oral presentation, 25^e Colloque International sur le Lexique et la Grammaire, Palerme, Italie.

- Klairs, Christos, and Georgios Babiniotis. 2009. *Grammatiki tis Neas Ellinikis. Domolitourgiki-Epikinoniaki*. Athens: Ellinika Grammata.
- Kyriacopoulou, Tita. 2005. *L'Analyse Automatique des Textes Écrits : Le Cas du Grec Moderne*. Thessaloniki: University Studio Press.
- Kyriacopoulou, Tita, Claude Martineau, and Thanassis Mavropoulos. 2011. “Les Noms Propres de Personne en Français et en Grec : Reconnaissance, Extraction et Enrichissement de Dictionnaire”. In *Penser le Lexique-Grammaire, Perspectives Actuelles*, edited by Fryni Kakoyianni-Doa, 467–479. Éditions Honoré Champion.
- Kyriakopoulou, Anthoula. 2011. *Les Noms Composés du Type N (E+DET:G) N:G en Grec Moderne*. PhD diss., Paris : Université Paris-Est Marne-la-Vallée.
- Linardaki, Evita, Carlos Ramisch, Aline Villavicencio, and Aggeliki Fotopoulou. 2010. “Towards the Construction of Language Resources for Greek Multiword Expressions: Extraction and Evaluation”. In *Proceedings of the LREC Workshop on Exploitation of Multilingual Resources and Tools for Central and (South) Eastern European Languages (ELRA)*, 1–8.
- Panayotopoulou, Styliani. 1992. *Les Noms Composés du Type Adjectif Nom (AN) en Grec Moderne*. Mémoire de DEA de Linguistique, France : Université Paris VIII and C.E.R.I.L.
- Paumier, Sébastien. 2003. *De la Reconnaissance de Formes Linguistiques à l'Analyse Syntaxique*. PhD diss., Paris : Université de Marne-la-Vallée.
- Ralli, Angela. 1991. “Leksiki Frasi : Antikimeno Morfologikou Endiaferontos”. In *Meletes gia tin Elliniki Glossa [Studies in Greek Linguistics] 1990*:139–158.
- Ralli, Angela. 2008. “I Morfologia os Aftonomo Tmima tis Grammatikis”. In *Glossis Xarin*, edited by Amalia Mozer, Ekaterini Bakakou-Orfanou, Christoforos Xaralambakis, Despoina Xeila-Markopoulou, 141–156.
- Ralli, Angela, and Melita Stavrou. 1995. “A/N Compounds vs. A/N Constructs in Greek.” In *Proceedings of the 2nd International Meeting of Greek Linguistics*, 97–107.
- Ralli Angela, and Melita Stavrou. 1998. “Morphology-Syntax Interface: A/N Compounds versus A/N Constructs in Modern Greek.” In *Yearbook of Morphology 1997*:243–264.
- Samaridi, E. Nikoleta, and Stella Markantonatou. 2014. “Parsing Modern Greek Verb MWEs with LFG/XLE Grammars.” In *Proceedings of the 10th*

- Workshop on Multiword Expressions (MWE 2014). Association for Computational Linguistics, 33–37.
- Sklavounou, Elsa. 1993. *Un Lexique-Grammaire Trilingue de Noms Composés (grec-français-anglais)–Application au Vocabulaire Spécialisé du Tennis*. D.E.A. Université Paris VIII, France.
- Thomou, Paraskevi. 2006. *Lexilogikes Sinapsis (Lexical Collocations) sti Nea Elliniki os Kseni Glossa*. PhD diss., Crete: University of Crete.
- Tsopanakis, Antonis. 1998. *Neoelliniki Grammatiki* (3rd edition). Thessaloniki: Kyriakidis.
- Wehrli, Eric. 2000. “Parsing and Collocations.” In *Natural Language Processing–NLP 2000*, edited by Dimitris Christodoulakis, 272–282. Springer Verlag.
- Xydopoulos, Georgios. 2008. *Leksikologia. Eisagogi stin Analisi tis Leksis kai tou Leksikou*. Athens: Patakis.