



FACHLITERATUR  
EDITION ROMIOSINI  
ΕΠΙΣΤΗΜΗ



ICGL12 | 12<sup>th</sup> 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 12<sup>TH</sup> INTERNATIONAL  
CONFERENCE ON GREEK LINGUISTICS**

**ΠΡΑΚΤΙΚΑ ΤΟΥ 12<sup>ΟΥ</sup> ΣΥΝΕΔΡΙΟΥ ΕΛΛΗΝΙΚΗΣ  
ΓΛΩΣΣΟΛΟΓΙΑΣ**

VOL. 1



© 2017 Edition Romiosini/CeMoG, Freie Universität Berlin. Alle Rechte vorbehalten.  
Vertrieb und Gesamtherstellung: Epubli ([www.epubli.de](http://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](http://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, Ναπολέον Κάτσος, Ευαγγελία Κορδώνη, Αμαλία Μόζερ, Ελένη Μπουτουλούση, Κική Νικηφορίδου, Αγγελική Ράλλη, Άννα Ρούσσου, 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 Σηφιανού: <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>Γατρικές 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 L2</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 Bacskai-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 Fragkakis, 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: Veritäten - Interferenz</i> .....	405
Μαρία Ιακώβου, Ηριάννα Βασιλειάδη-Λιναρδάκη, Φλώρα Βλάχου, Όλγα Δήμα, Μαρία Καββαδία, Τατιάνα Κατσίνα, Μαρίνα Κουτσομπού, 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 & Δώρας 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 mīpos</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 Chatzioannidis, 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 Τσαμαδοῦ- Jacobberger & Μαρία Ζέρβα: <i>Εκμάθηση ελληνικών στο Πανεπιστήμιο Στρασβούργου: κίνητρα και αναπαραστάσεις</i> ...	1051
Stavroula Tsiplakou & Spyros Armostis: <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:	
<i>Η μεθοδολογική προσέγγιση της εικόνας της Γερμανίας στις ελληνικές εφημερίδες .....</i>	1095
Anastasia Tzilinis:	
<i>Begründendes Handeln im neugriechischen Wissenschaftlichen Artikel: Die Situierung des eigenen Beitrags im Forschungszusammenhang.....</i>	1109
Κυριακούλα Τζωρτζάτου, Αργύρης Αρχάκης, Άννα Ιορδανίδου & Γιώργος Ι. Ευδόπουλος:	
<i>Στάσεις απέναντι στην ορθογραφία της Κοινής Νέας Ελληνικής: Ζητήματα ερευνητικού σχεδιασμού .....</i>	1123
Nicole Vassalou, Dimitris Papazachariou & Mark Janse:	
<i>The Vowel System of Mišótika Cappadocian .....</i>	1139
Marina Vassiliou, Angelos Georganas, Prokopis Prokopidis & Haris Papageorgiou:	
<i>Co-referring or not co-referring? Answer the question!.....</i>	1155
Jeroen Vis:	
<i>The acquisition of Ancient Greek vocabulary.....</i>	1171
Christos Vlachos:	
<i>Mod(aliti)es of lifting wh-questions.....</i>	1187
Ευαγγελία Βλάχου & Κατερίνα Φραντζή:	
<i>Μελέτη της χρήσης των ποσοδεικτών λίγο-λιγάκι σε κείμενα πολιτικού λόγου .....</i>	1201
Madeleine Voga:	
<i>Τι μας διδάσκουν τα ρήματα της ΝΕ σχετικά με την επεξεργασία της μορφολογίας.....</i>	1213
Werner Voigt:	
<i>«Σεληνάκι μου λαμπρό, φέγγε μου να περπατώ ...» oder: warum es in dem bekannten Lied nicht so, sondern eben φεγγαράκι heißt und ngr. φεγγάρι .....</i>	1227
Μαρία Βραχιονίδου:	
<i>Υποκοριστικά επιρρήματα σε νεοελληνικές διαλέκτους και ιδιώματα .....</i>	1241
Jeroen van de Weijer & Marina Tzakosta:	
<i>The Status of *Complex in Greek.....</i>	1259
Theodoros Xioufis:	
<i>The pattern of the metaphor within metonymy in the figurative language of romantic love in modern Greek.....</i>	1275

ΑΠΟ ΤΗΝ «ΕΚΦΡΑΣΗ» ΣΤΟ «ΠΟΛΥΤΡΟΠΟ»:  
ΣΧΕΔΙΑΣΜΟΣ ΚΑΙ ΟΡΓΑΝΩΣΗ ΕΝΟΣ  
ΕΝΝΟΙΟΛΟΓΙΚΟΥ ΛΕΞΙΚΟΥ

Αγγελική Φωτοπούλου & Βούλα Γιούλη

ΙΕΛ/Ε.Κ «Αθηνά»

afotop@ilsp.athena-innovation.gr, voula@ilsp.athena-innovation.gr

*Abstract*

*The paper presents work aimed at the development of a conceptual lexicon of Modern Greek. The lexicon builds on the dictionary infrastructure defined within the project EKFRASIS and is organized as a linguistic ontology. We present the main structure of the POLYTROPON lexicon, and we focus on the methodological considerations thereof.*

*Λέξεις-κλειδιά: εννοιολογικό λεξικό, γλωσσική οντολογία, σημασιολογικά πεδία, σημασιολογικές σχέσεις, λεξικοί πόροι*

## **1. Εισαγωγή**

Η αποσαφήνιση της σημασίας των λέξεων και η επίλυση της λεξικής αμφισημίας (word sense disambiguation) εξαρτάται από το συγκεκριμένο και - κατ' επέκταση - από το θέμα που πραγματεύεται το κείμενο. Οποιοδήποτε δε θέμα ενός κειμένου απαιτεί ένα συγκεκριμένο λεξιλόγιο. Αυτή η κυκλική καταρχήν σχέση μάς οδηγεί στην θέση ότι η κατάλληλη ομαδοποίηση των λέξεων βάσει γενικότερων θεμάτων μπορεί να είναι πολλαπλώς χρήσιμη. Σ' αυτή την απλή αλλά πολύ επιτυχημένη ιδέα στηρίζονται τα εννοιολογικά ή ονομαστικά λεξικά (ή θησαυροί), στα οποία οι λέξεις ομαδοποιούνται γύρω από γενικότερου χαρακτήρα σημασιολογικές κατηγορίες ή έννοιες. Για παράδειγμα, οι λέξεις *εργασία*, *μισθωτή εργασία*, *επάγγελμα*, *βιοπορισμός*, *δημο-*

σιογραφία, επαγγελματική στέγη, επιτήδευμα, εργολαβία, μαστοράτσα, οικοβιοτεχνία ανήκουν όλες στο θεματικό πεδίο που αναφέρεται στο «επάγγελμα».

Στο άρθρο αυτό παρουσιάζουμε τον σχεδιασμό και την οργάνωση ενός υπολογιστικού εννοιολογικού λεξικού, το οποίο αναπτύχθηκε στο πλαίσιο του έργου «ΠΟΛΥΤΡΟΠΟΝ» και το οποίο έχει ως αφετηρία το οντολογικό μοντέλο που προτείνει το λεξικό *ΕΚΦΡΑΣΗ*, (Μαρκαντωνάτου & Φωτοπούλου, 2007). Βασική σχεδιαστική αρχή και των δύο λεξικών είναι η εξής: από μία τυχαία λέξη ενός σημασιολογικού πεδίου να καθίσταται δυνατή η πρόσβαση στις άλλες λέξεις του πεδίου μέσω σημασιολογικών και μορφολογικών σχέσεων. Η απαίτηση αυτή οδηγεί στο να υιοθετηθούν ως βασικές μονάδες ταξινόμησης τα λεγόμενα *σημασιολογικά πεδία*. Στην συνέχεια θα αναφερθούμε σε σχετικές εργασίες (ενότητα 2), θα περιγράψουμε την δομή του λεξικού (ενότητα 3), την μεθοδολογία ανάπτυξής του (ενότητα 4), την επιλογή του υλικού (ενότητα 5) και τέλος τα συμπεράσματα (ενότητα 6).

## 2. Σχετικές εργασίες

Το υπολογιστικό εννοιολογικό λεξικό *ΠΟΛΥΤΡΟΠΟΝ* έχει ως αφετηρία και έμπνευση τους θησαυρούς Roget's και «Ονομαστικών» Βοσταντζόγλου ενώ οργανώνεται κατά το πρότυπο των οντολογιών, υιοθετώντας τις αρχές της οντολογίας OntoClean (Guarino & Welty 2004). Στην παρούσα υλοποίηση, αξιοποιήθηκαν, επίσης χαρακτηριστικά επιλεγμένων υπολογιστικών λεξικών (Wordnet, Framenet), γεγονός που καθιστά τον πόρο συμβατό με υφιστάμενους λεξιλογικούς πόρους. Στη συνέχεια, θα αναφέρουμε τους γλωσσικούς πόρους οι οποίοι αποτέλεσαν πηγή αναφοράς και για το δικό μας λεξικό.

Βασικός πόρος αναφοράς αποτελεί ο θησαυρός Roget's (Roget's Thesaurus): το πρώτο σχέδιασμα του του λεξικού συντάχθηκε το 1805 αλλά ο θησαυρός ολοκληρώθηκε και εκδόθηκε για πρώτη φορά 47 χρόνια αργότερα, το 1852. Βασικό χαρακτηριστικό του θησαυρού είναι ο τρόπος σχεδιασμού του. Το ενδιαφέρον του λεξικογράφου δεν είναι πια τι σημαίνει η κάθε λέξη, αλλά ποιες λέξεις έχουν περίπου την ίδια σημασία, ποια είναι αυτή η σημασία και αν μπορεί αυτή να ομαδοποιηθεί μαζί με άλλες.

Στην ίδια λογική, το Αντιλεξικόν Βοσταντζόγλου (1962) αποτελεί ένα καινοτόμο για την εποχή του εννοιολογικά οργανωμένο λεξικό της ελληνικής, το οποίο συνδυάζει ερμηνεύματα με γραμματικο-συντακτικές πληροφορίες (μέρος του λόγου, βασική ρηματική σύνταξη). Αποτελεί προσπάθεια να γίνει μια «γενική απογραφή του

λεκτικού πλούτου της νεοελληνικής γλώσσας, καθαρεύουσας και δημοτικής, και μια λογική διάταξη της ύλης με τρόπο που εφαρμόζεται για πρώτη φορά» προκειμένου να γίνει ακριβέστερος ο διαχωρισμός των εννοιών και ευκολότερη η εξεύρεση των λέξεων μέσα στο κείμενο.

Σε άλλη κατεύθυνση, το υπολογιστικό λεξικό WordNet (Fellbaum 1998) είναι οργανωμένο σημασιολογικά, και περιέχει ουσιαστικά, ρήματα, επίθετα και επιρρήματα. Η βασική σημασιολογική μονάδα του λεξικού είναι το σύνολο συνωνύμων (synset). Υπάρχει διαχωρισμός εννοιών σε σημασίες (senses), ενώ ουσιαστικά, ρήματα, επίθετα και επιρρήματα ομαδοποιούνται σε σύνολα συνωνύμων με διακριτή σημασία. Τα σύνολα συνωνύμων συνάπτουν μεταξύ τους σχέσεις με κυρίαρχη αυτή της υπερωνυμίας-υπωνυμίας, η οποία αποτελεί τον βασικό άξονα ανάπτυξης του λεξικού. Παράλληλα, τα σύνολα συνωνύμων αναπτύσσουν μεταξύ τους και άλλου είδους σχέσεις (όπως αυτή της αντωνυμίας, της μερωνυμίας, της μορφολογικά παραγόμενης συγγένειας κ.λπ.).

Η λεξιλογική βάση VerbNet (Kipper-Schuler 2005), αποτελεί ένα ρηματικό λεξικό, συμβατό με το WordNet, στο οποίο, όμως, δηλώνονται ρητά οι συντακτικές και σημασιολογικές πληροφορίες κάθε ρηματικής εγγραφής. Η σημασιολογική αναπαράσταση αποτελεί συνδυασμό παραδοσιακής σημασιολογικής πληροφορίας και σύγχρονης συντακτικής προσέγγισης (π.χ. με την παράθεση θεματικών ρόλων, πλαισίων υποκατηγοριοποίησης, επιλογικών περιορισμών), απεικονίζοντας τη σημασιολογική συμπεριφορά των ρημάτων. Με τον τρόπο αυτό, επιτυγχάνεται η αναπαράσταση σύνθετων σημασιολογικών σχημάτων.

Τέλος, ο λεξιλογικός πόρος FrameNet (Fillmore et al. 2001), αναπτύχθηκε και οργανώθηκε στη βάση της θεωρίας των σημασιολογικών πλαισίων (Frame Semantics). Έτσι, η λεξική σημασία στο FrameNet κωδικοποιείται βάσει γνωσιακών δομών που συνδέονται μεταξύ τους με τέτοιο τρόπο, ώστε προκειμένου να κατανοήσει κανείς μια ορισμένη έννοια είναι απαραίτητο να κατανοήσει ολόκληρο το σύστημα.

Ως προς τις οντολογίες και τα οντολογικά σχήματα αναφέρουμε ενδεικτικά :

- i) το OntoWordnet Project (Gagnemi et al 2003), στόχος του οποίου είναι η σαφής δήλωση των δεσμεύσεων (commitments) που εκφράζονται στο WordNet και η συνεπής εφαρμογή τους. Έχει γίνει επίσης προσπάθεια να εμπλουτιστούν οι σαφώς ορισμένες σχέσεις του WordNet με τις επιπλέον σχέσεις που εννοούνται στο κείμενο ορισμού (gloss),

- ii) η οντολογία SUMO (Niles & Pease 2000, Pease 2011), η οποία αποτελείται από ένα βασικό οντολογικό σχήμα ανωτέρου επιπέδου και πλήθος οντολογιών πεδίου. Συνιστά έτσι μία από τις μεγαλύτερες τυπικές οντολογίες που χρησιμοποιείται σε ποικίλες εφαρμογές Επεξεργασίας Φυσικής Γλώσσας,
- iii) η οντολογία Mikrokosmos (Mahesh & Nirenburg 1995a, Mahesh & Nirenburg 1995b), η οποία είναι ανεξάρτητη γλώσσας (language-independant) και δε βασίζεται σε κάποια οντολογική θεωρία αλλά προέκυψε από την προσπάθεια επίλυσης πρακτικών ζητημάτων εννοιολογικής αναπαράστασης,
- iv) η οντολογία SENSUS (Corcho et al 2006) προέκυψε από την ανάγκη για την κατανόηση κειμένων ως προϋπόθεση για τη μετάφρασή τους μέσω εργαλείων μηχανικής μετάφρασης, και για την αποδοτική εξαγωγή και ανάκτηση πληροφορίας (information retrieval) από αυτά. Η οντολογία αυτή αποτελεί προέκταση του WordNet έχοντας στοιχεία και από το Penmann Upper Model.

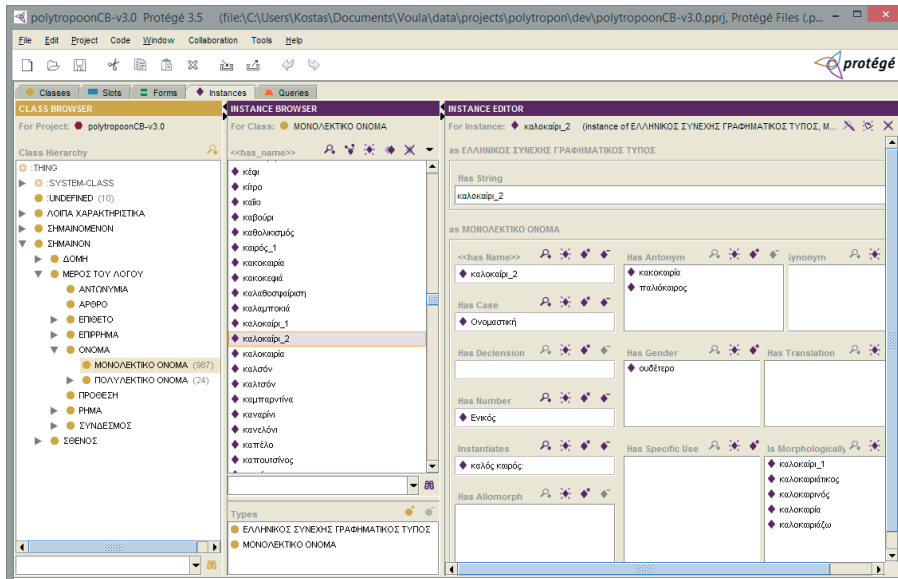
Κυρίως, όμως, το εν λόγω λεξικό ακολουθεί το οντολογικό μοντέλο που προτείνει το λεξικό *ΕΚΦΡΑΣΗ*, και το οποίο - με τη σειρά του - στηρίζεται στην ιδέα του *γλωσσικού σημείου*, ως αδιάρρηκτης ένωσης μιας έννοιας (*ΣΗΜΑΙΝΟΜΕΝΟ*) με μια ακουστική εικόνα (*ΣΗΜΑΙΝΟΝ*). Το οντολογικό αυτό μοντέλο θα περιγράψουμε στην ενότητα που ακολουθεί.

### 3. Περιγραφή Λεξικού

Η βασική δομή της οντολογίας στην *ΕΚΦΡΑΣΗ*, η οποία υιοθετείται με κάποιες αποκλίσεις και στην παρούσα υλοποίηση, ξεκινά από δύο βασικούς ανεξάρτητους κλάδους: (α) τον κλάδο *ΣΗΜΑΙΝΟΝ*, με πραγματώσεις λεξήματα, και (β) τον κλάδο *ΣΗΜΑΙΝΟΜΕΝΟ*, με πραγματώσεις έννοιες. Μονάδα οργάνωσης του λεξικού αποτελεί η *λεξική εγγραφή*, που μοντελοποιεί αυτή την ιδέα. Σύμφωνα με το οντολογικό μοντέλο που υιοθετήθηκε, ο όρος εγγραφή, παραπέμπει στην κατηγορία *ΛΕΞΙΚΗ ΜΟΝΑΔΑ* της οντολογίας του λεξικού, η οποία είναι, με τη σειρά της, υποκατηγορία της κλάσης *ΣΗΜΑΙΝΟΝ*. Κάθε λεξική εγγραφή αποτελεί μοναδικό συνδυασμό μορφής και σημασίας, ενώ το λεξικό περιλαμβάνει εγγραφές μονολεκτικές και πολυλεκτικές. Οι μονολεκτικές εγγραφές αποτελούν στιγμιότυπα της υποκατηγορίας *ΜΟ-*

ΝΟΛΕΚΤΙΚΗ ΛΕΞΙΚΗ ΜΟΝΑΔΑ ενώ οι πολυλεκτικές εγγραφές αποτελούν στιγμιότυπα της υποκατηγορίας ΠΟΛΥΛΕΚΤΙΚΗ, ΛΕΞΙΚΗ ΜΟΝΑΔΑ. Βασική ιδιότητα της εγγραφής είναι η συσχετίση της με μία έννοια της κλάσης ΣΗΜΑΙΝΟΜΕΝΟΝ. Ένας γραφηματικός τύπος (μονολεκτικός ή πολυλεκτικός) συνδέεται με μία έννοια μέσω της σχέσης πραγμάτωσης (*instantiates*(X,Y)).

Η γλωσσική πληροφορία η οποία αφορά στη μορφή, κωδικοποιείται υπό τον κλάδο ΣΗΜΑΙΝΟΝ ενώ στον κλάδο ΣΗΜΑΙΝΟΜΕΝΟΝ συστηματοποιούνται οι σημασιολογικές σχέσεις. Η πληροφορία στον κλάδο ΣΗΜΑΙΝΟΝ αφορά στοιχεία όπως: γραμματική κατηγορία, στοιχεία συμφωνίας (γένος, αριθμός, κλπ). Επιπλέον, στο επίπεδο αυτό κωδικοποιήθηκαν λεξικές σημασιολογικές σχέσεις (συνωνυμία, αντωνυμία), αλλά και λοιπές λεξικές σχέσεις που σχετίζονται με τη μορφολογία (αλλομορφία, μορφολογικά συγγενείς λέξεις) μεταξύ των εγγραφών. Τέλος, πληροφορία αναφορικά με το επίπεδο ύφους (register) δόθηκε σε λεξικές εγγραφές όπου απαιτείται. Παράδειγμα κωδικοποίησης εγγραφής στην κλάση ΣΗΜΑΙΝΟΝ δίνεται στην εικόνα 1 παρακάτω. Αντίστοιχα, στον κλάδο ΣΗΜΑΙΝΟΜΕΝΟΝ κωδικοποιούνται σχέσεις σημασιολογικές.



Σχήμα 1 | Παράδειγμα κωδικοποίησης εγγραφής στον κλάδο ΣΗΜΑΙΝΟΝ

Σε αντίθεση με την ΕΚΦΡΑΣΗ, στην παρούσα υλοποίηση δόθηκε έμφαση στον προσδιορισμό και την αυστηρή δόμηση των κλάδων ΕΙΔΟΣ Χ και ΠΕΡΙ Χ του ΣΗΜΑΙΝΟΜΕΝΟΥ. Έτσι, στον κλάδο ΕΙΔΟΣ Χ έχουν υπαχθεί έννοιες οι οποίες πρωτίστως συνδέονται μεταξύ τους με τη σχέση *is\_a*, ή σχέση υπωνυμίας – υπερωνυμίας. Υπό την έννοια αυτή, ο κλάδος ΕΙΔΟΣ Χ εμφανίζει μία ιεραρχική δόμηση. Αντιθέτως, υπό τον κλάδο ΠΕΡΙ Χ οι έννοιες οργανώνονται σε σημασιολογικά πεδία και συνδέονται μεταξύ τους με *σχέσεις συνάφειας*. Υπό την έννοια αυτή τα σημασιολογικά πεδία αποτελούν σύνολα σημασιών τα οποία αντιμετωπίζονται ως ανεξάρτητες οντότητες στις οποίες εγκλείονται έννοιες ή σημασίες, οι οποίες συνδέονται μεταξύ τους με ένα πυκνό δίκτυο σχέσεων. Οι σχέσεις συνάφειας δεν είναι κοινές για όλα τα πεδία, και προσδιορίζονται ανάλογα με τις σημασίες του εκάστοτε σημασιολογικού πεδίου. Έτσι, για κάποια σημασιολογικά πεδία, έχουν οριστεί σχέσεις οι οποίες είναι εγγενείς και τις χαρακτηρίζουν.

Εφόσον στο λεξικό οργανώνονται κάτω από κάθε κλάση εγγραφές με κοινό σημασιολογικό παρονομαστή, οι σχέσεις που προβλέπονται για κάθε κλάση συνδέουν εγγραφές που ανήκουν σε διαφορετικές γραμματικές κατηγορίες (π.χ. τα ουσιαστικά με τα ρήματα, επίθετα και επιρρήματα) καθώς επίσης και εγγραφές που ανήκουν σε διαφορετικά αλλά σχετιζόμενα σημασιολογικά πεδία (π.χ. επαγγελμα με τον τόπο άσκησης του επαγγέλματος, κλπ). Στην κατεύθυνση αυτή, έχει οριστεί μία σειρά σχέσεων οι οποίες συνδέουν εγγραφές που ανήκουν σε διαφορετικές κατηγορίες (π.χ., η σημασιολογική σχέση μεταξύ ουσιαστικών και ρημάτων, επιθέτων και ουσιαστικών, κλπ). Για παράδειγμα, η σχέση *is\_the\_action\_of* συνδέει ένα ρήμα με το ουσιαστικό, ενώ η σχέση *is\_the\_property\_of* συνδέει ουσιαστικό με επίθετο:

*is\_the\_action\_of* (οδηγώ, οδήγηση)

*is\_the\_property\_of* (εξυπνάδα,έξυπνος)

Δίνουμε ακόμα κάποια παραδείγματα για το πώς προβλέπονται οι σχέσεις που οργανώνουν τις εγγραφές σε σύνολα και υποσύνολα. Πρόκειται για σχέσεις απλού εγκλεισμού και σχέσεις κληρονομικότητας, παρόμοια με την Inheritance στο Framenet. Προκειμένου για τη διευκόλυνση του λεξικογράφου, όταν ορίζεται μία σχέση, ορίζεται ταυτόχρονα και μία ή περισσότερες κλάσεις (δηλαδή ομάδες εννοιών ή γραμματικών κατηγοριών, κλπ) στις οποίες μπορούν να ανήκουν τα μέλη που συμμετέχουν στην σχέση αυτή:

1. *is\_made\_of* (ξύλινος, ξύλο) σχέση στο σημασιολογικό πεδίο MATERIAL

2. `is_related_to.feels_emotion`(έκπληκτος, έκπληξη) σχέση στο σημασιολογικό πεδίο EMOTION
3. `is_the_emotion_of`(χαίρομαι, χαρά) σχέση στο σημασιολογικό πεδίο EMOTION
4. `is_the_state_of`(κοιμάμαι, ύπνος) σχέση στο σημασιολογικό πεδίο STATE
5. `causes_emotion`(λυπητερός, λύπη) σχέση στο σημασιολογικό πεδίο EMOTION
6. `is_the_manifestation_of`(δάκρυ, λύπη) σχέση στο σημασιολογικό πεδίο EMOTION
7. `is_the_manifestation_of`(δάκρυ, συγκίνηση) σχέση στο σημασιολογικό πεδίο EMOTION

Όπως είναι φυσικό, κάποιες από τις σχέσεις αυτές μπορούν να οριστούν και αντιστρόφως, και αυτή η λειτουργικότητα δίνεται με τον ορισμό αντίστροφων σχέσεων (`inverse_relations`):

`is_the_property_of` ↔ `has_property`  
`is_the_quality_of` ↔ `has_the_quality`  
`needs_tool` ↔ `is_used_in`  
`is_practiced_by` ↔ `practices`  
`uses_tool` ↔ `is_used_by`  
`works_in` ↔ `is_the_workplace_of`

Πιο συγκεκριμένα, το σύνολο των σχέσεων στα διάφορα επίπεδα δημιουργεί ένα πυκνό δίκτυο συνδεόμενων εγγραφών, που θα επιτρέψει την επιδιωκόμενη εύκολη και συστηματική μετάβαση από μία τυχαία λέξη ενός σημασιολογικού πεδίου στις άλλες λέξεις του πεδίου. Να σημειωθεί, τέλος, ότι κάθε στιγμιότυπο-έννοια αναπαρίσταται μέσω σύντομου ορισμού (`gloss`), και εν συνεχεία συμπληρώνονται οι μεταβλητές για τις εκάστοτε σχέσεις.

Συνολικά, το λεξικό στην παρούσα φάση αποτελείται από ~16,5K εγγραφές. Στον κλάδο ΕΙΔΟΣ X έχουν οριστεί συνολικά 114 κλάσεις οι οποίες αντιστοιχούν σε σημασιολογικά πεδία, ενώ στην κλάση ΠΕΡΙ X ορίστηκαν ~7000 έννοιες. Αριθμητικά στοιχεία για τη σύσταση του λεξικού δίνονται στον πίνακα 1 που ακολουθεί.

	Μονολεκτικές εγγραφές	Πολυλεκτικές εγγραφές	Σύνολο
<b>Επίθετα</b>	3259	41	3300



<b>Ουσιαστικά</b>	8514	573	9087
<b>Ρήματα</b>	2179	687	2866
<b>Επιρρήματα</b>	868	209	1077
<b>Σύνολο</b>	14820	1510	16330

Πίνακας 1 | Κατανομή εγγραφών του λεξικού

Τέλος, ένα δίκτυο το οποίο συνέχουν ~70 σημασιολογικές σχέσεις, ενώνει τις εγγραφές του λεξικού στο επίπεδο της σημασίας.

#### 4. Μεθοδολογία ανάπτυξης του λεξικού

Ακολουθώντας σύγχρονες λεξικογραφικές πρακτικές, η ανάπτυξη του εν λόγω λεξικού στηρίχθηκε σε αυστηρώς καθορισμένες μεθοδολογικές αρχές. Επιγραμματικά, οι αρχές στις οποίες υπόκειται η ανάπτυξη του λεξικού συνίσταντα στα ακόλουθα: (α) λεπτομερής σχεδιασμός των βημάτων συλλογής υλικού, και της εννοιολογικής οργάνωσης καθώς επίσης και λήψη αποφάσεων για τα κριτήρια επιλογής και τον τρόπο ταξινόμησης του λεξικού υλικού, (β) στήριξη στα εμπειρικά δεδομένα (corpus-based lexicography), και τέλος (γ) ακρίβεια στην κωδικοποίηση.

Οι αρχές αυτές εξειδικεύονται περαιτέρω με μία σειρά από βήματα τα οποία συνιστούν μια υβριδική μεθοδολογία ανάπτυξης του λεξικού από-κάτω-προς-τα-πάνω (*bottom-Up*) αλλά και από πάνω-προς-τα-κάτω (*Top-down*) για την ολοκλήρωση της ερευνητικής εργασίας. Ειδικότερα, η *bottom-Up* προσέγγιση ξεκινά από το λημματολόγιο και καταλήγει στη δόμηση των σημασιολογικών πεδίων. Οι εργασίες εδώ συνίστανται στα ακόλουθα: (α) κατάρτιση λημματολογίου και διάκριση σημασιών, (β) στοίχιση σημασιών με εγγραφές του λεξικού (επιλογή σημασιών), (γ) ταξινόμηση εγγραφών σε σημασιολογικά πεδία, (δ) εύρεση σχέσεων μεταξύ εγγραφών εντός και εκτός του σημασιολογικού πεδίου, και (ε) ιεραρχική οργάνωση σημασιολογικών πεδίων με στόχο τον ορισμό ενός οργανωμένου ταξινομικού σχήματος. Αντιθέτως, η *top-down* προσέγγιση λαμβάνει ως αφετηρία τα σημασιολογικά πεδία και καταλήγει στις εγγραφές. Η μεθοδολογία αυτή αποσκοπεί στην οργάνωση και τον εμπλουτισμό των επιμέρους σημασιολογικών πεδίων, αλλά και στον εντοπισμό σχέσεων συνάφειας

μεταξύ εγγραφών διαφορετικών σημασιολογικών πεδίων. Σύμφωνα με αυτή, για κάθε πεδίο χωριστά, ακολουθούνται τα παρακάτω βήματα: (α) οργάνωση εγγραφών ανάλογα με τη γραμματική κατηγορία/σημασία, (β) ιεραρχική οργάνωση εγγραφών μέσω σχέσεων υπωνυμίας, υπερωνυμίας, (γ) εξαντλητική μελέτη και εύρεση σχέσεων μεταξύ εγγραφών εντός του πεδίου, (δ) εξαντλητική μελέτη και εύρεση σχέσεων μεταξύ εγγραφών εκτός του πεδίου, και τέλος (ε) εμπλουτισμός με νέες εγγραφές προκειμένου για την πληρέστερη και ακριβέστερη απεικόνιση κάθε πεδίου.

Σε γενικές γραμμές, η πρώτη προσέγγιση (bottom-up) συντελεί περισσότερο στη δόμηση του κλάδου ΕΙΔΟΣ\_X της οντολογίας ενώ η δεύτερη (top-down) βοηθά στη δόμηση του κλάδου ΠΕΡΙ\_X και στον ορισμό των σχέσεων συνάφειας οι οποίες συνδέουν στιγμιότυπα διαφορετικών κλάσεων ΕΙΔΟΣ\_X μέσα στο ΠΕΡΙ\_X. Η συγκεκριμένη υβριδική προσέγγιση προσφέρει τη δυνατότητα μίας συνολικής θεώρησης του λεξικού, ενώ ταυτόχρονα επιτρέπει την περαιτέρω ολοκληρωμένη ανάπτυξη των επιμέρους πεδίων. Στη διάρκεια του έργου, πραγματοποιήθηκαν πολλαπλές επαναλήψεις της ανωτέρω διαδικασίας για την διενέργεια ελέγχων και διορθώσεων.

## 5. Επιλογή και ταξινόμηση του γλωσσικού υλικού

Όπως αναφέρθηκε παραπάνω, βασικό στοιχείο της παρούσας προσέγγισης είναι η ανάπτυξη του εννοιολογικού λεξικού στη βάση εμπειρικών δεδομένων κατά το πρότυπο της σύγχρονης Λεξικογραφικής πρακτικής που βασίζεται σε Σώματα Κειμένων (Corpus-based Lexicography). Για τον λόγο αυτό, χρησιμοποιήθηκαν δύο είδη γλωσσικών πόρων: (α) ένα γενικό Σώμα Κειμένων (στο εξής ΣΚ) και (β) μικρότερα ειδικά ΣΚ (specialized corpora) τα οποία συνελέγησαν από ετερογενείς πηγές. Ειδικότερα, για την ανάπτυξη του λεξικού χρησιμοποιήθηκε εκτενώς ο Εθνικός Θησαυρός της Ελληνικής Γλώσσας (ΕΘΕΓ)<sup>1</sup> (Hatzigeorgiou et al 2000), ένα ΣΚ αναφοράς (~47Μ λέξεις), το οποίο έχει αναπτυχθεί στο ΙΕΛ με απώτερο στόχο την καταγραφή της σύγχρονης ελληνικής γλώσσας. Επιβοηθητικά, συγκροτήθηκαν μικρότερα ΣΚ με προέλευση από το διαδίκτυο: ιστολόγια, forums, κείμενα έκφρασης γνώμης (π.χ. επιστολές αναγνωστών σε διαδικτυακές εκδόσεις εφημερίδων), κείμενα life-style, σχολιασμός ειδήσεων, κ.τ.λ., καθώς επίσης και από μεταγραμμένα κείμενα προφορικού λόγου. Η εργασία αυτή, όπως αναμενόταν, έδωσε γλωσσικά δεδομένα τα οποία περιελάμβαναν λήμματα τα

1 <http://hnc.ilsp.gr>

οποία δεν είναι υψίσυχνα στον ΕΘΕΓ, χρησιμοποιούνται, όμως, ευρύτατα στον καθημερινό λόγο. Επίσης, από τα ειδικά κείμενα εξήχθησαν εγγραφές για τον εμπλουτισμό των επιμέρους σημασιολογικών πεδίων με τα οποία ασχοληθήκαμε (υγεία, εκπαίδευση, τέχνη). Τέλος, τα κειμενικά αυτά δεδομένα χρησιμοποιήθηκαν για τον με ημι-αυτόματο τρόπο εμπλουτισμό του εννοιολογικού λεξικού με πολυλεκτικές εγγραφές.

### *5.1. Ταξινόμηση εγγραφών σε θεματικά πεδία*

Στην παρούσα εργασία, η έννοια του θεματικού πεδίου ορίστηκε ως σημασιολογικό ή λεξιλογικό πεδίο. Στην βιβλιογραφία χρησιμοποιείται ο όρος λεξιλογικό πεδίο (από τον γερμανικό όρο Wortfeld) για να δηλώσει το σύνολο των λέξεων, οι οποίες βρίσκονται σε στενή αλληλεξάρτηση. Κάθε λέξη προσδιορίζεται σημασιολογικά από τις σχέσεις της προς τις συγγενείς λέξεις και το σύνολο. Σύμφωνα με την άποψη αυτή, το Λεξιλογικό πεδίο αποτελείται από λεξιλογικές μονάδες συνδεδεμένες με στενότερη σημασιολογική σχέση, ενώ καθορίζονται και διακρίνονται με κριτήρια επιλογής και αλληλοαποκλεισμού, όπως φαίνεται στα παραδείγματα (1) και (2) αντιστοίχως:

- α) Θα πάω στη Ρόδο σε {μέρες, μήνες, χρόνια...}
- β) το φόρεμά μου είναι πράσινο {κόκκινο, μπλε, μαύρο...}

Επιπλέον, το σημασιολογικό πεδίο ορίζεται ως το σύνολο των όρων που συνδέονται μεταξύ τους με μια δομή σχέσεων, μορφολογικών, σημασιολογικών / εννοιολογικών / νοητικών, μορφο-σημασιολογικών.

Τα παραπάνω κριτήρια ελήφθησαν υπόψη κατά τη διαδικασία ταξινόμησης των εγγραφών σε σημασιολογικά πεδία. Ειδικότερα, αρχικά το σύνολο των επιλεγμένων εγγραφών ομαδοποιήθηκε με βάση έναν κοινό σημασιολογικό παρονομαστή. Εν συνεχεία, ακολούθησε η εργασία για την εύρεση του αρχιλεξήματος, του χαρακτηριστικού ή αντιπροσωπευτικού, δηλαδή λεξήματος για τον χαρακτηρισμό του θεματικού πεδίου. Στο επόμενο στάδιο, η εργασία ολοκληρώθηκε με την εύρεση και τον εντοπισμό σε πρωτογενή δεδομένα (σώματα κειμένων) και σε άλλα έργα αναφοράς (π.χ. λεξικά) άλλων εγγραφών-λεξημάτων καθώς και η ταξινόμηση αυτών στα θεματικά πεδία που προέκυψαν. Απώτερο στόχο της εργασίας αυτής αποτελεί η κατά το δυνατόν πληρέστερη δόμηση των θεματικών πεδίων με την υπαγωγή σε αυτά ικανού αριθμού εγγραφών.

Οι ανωτέρω εργασίες πραγματοποιήθηκαν από 2 γλωσσολόγους οι οποίοι επισημείωσαν τις επιλεγμένες εγγραφές αναφορικά με το θεματικό πεδίο με τρόπο συστηματικό. Τακτικοί έλεγχοι για την εύρεση του ποσοστού συμφωνίας μεταξύ των γλωσσολόγων (inter-annotator agreement) συνετέλεσαν στον επαναπροσδιορισμό και περαιτέρω ταξινόμηση. Τόσο κατά την διάρκεια των παραπάνω διαδικασιών (επιλογή και κωδικοποίηση εγγραφών), όσο και μετά το πέρας αυτών, προβλέφθηκε η διενέργεια πολλαπλών ελέγχων τόσο τυπικών (formal validation) όσο και περιεχομένου (content validation) με στόχο την – κατά το δυνατόν διασφάλιση τόσο της ορθότητας της γλωσσικής πληροφορίας που κωδικοποιείται, όσο και της συνέπειας και συστηματικότητας στη χρήση των σχέσεων, ιδιοτήτων, κλπ. Εννοείται ότι η αναδιοργάνωση των πεδίων και ο εμπλουτισμός τους με νέα δεδομένα συνεχίζεται.

## 6. Συμπεράσματα

Στόχος του παρόντος κειμένου, όπως είπαμε και στην αρχή, ήταν η παρουσίαση του εννοιολογικού λεξικού που αναπτύχθηκε στο Ινστιτούτο Επεξεργασίας του Λόγου. Στα αποτελέσματα της ερευνητικής εργασίας καταγράφονται τα πρωτογενή δεδομένα (Σώματα Κειμένων), τα οποία συνελέγησαν για την ανάπτυξη του λεξικού, το οντολογικό σχήμα, η οντολογία, και – φυσικά – το υπολογιστικό λεξικό (~16,5 εγγραφές μέχρι τώρα). Μελλοντική εργασία προβλέπει τον εμπλουτισμό του λεξικού με νέες εγγραφές, στα υπάρχοντα αλλά και σε νέα θεματικά πεδία καθώς επίσης και τον ορισμό νέων σχέσεων μεταξύ δύο ή περισσότερων λέξεων. Αυτό το υπολογιστικό εννοιολογικό λεξικό, βασική υποδομή για την ελληνική γλώσσα, μπορεί να αξιοποιηθεί οριζόντια σε διάφορες εφαρμογές και γλωσσικές τεχνολογίες, όπως: αποσαφήνιση σημασίας λεκτικών μονάδων (WSD), διασύνδεση εννοιών (Concept Linking), αλλά μπορεί επίσης με την κατάλληλη διεπαφή να αποτελέσει εργαλείο για εκπαιδευτικές και παντός τύπου, συγγραφικές, εφαρμογές.

Το παρόν έργο χρηματοδοτήθηκε από το έργο ΚΡΗΠΙΣ - Αναλυτική Περιεχομένου και Πολυτροπική Επικοινωνία (ΠΟΛΥΤΡΟΠΟΝ), ΟΠΣ : 448306).

## Βιβλιογραφία

- Βοσταντζόγλου, Θεολόγος. 1962. «Αντιλεξικόν ή ονομαστικόν της Νεοελληνικής γλώσσης». Αθήνα, Επανεκδοση 2008.
- Corcho, Oscar, Pinar Alper, Ioannis Kotsiopoulos, Paolo Missier, Sean Bechhofer, and Carole Goble. 2006. "An overview of S-OGSA: A Reference Semantic Grid Architecture".
- Fellbaum, Christiane (ed.). 1998. *WordNet: An Electronic Lexical Database*. The MIT Press.
- Fotopoulou, Aggeliki, Stella Markantonatou, and Voula Giouli. 2014. "Encoding MWEs in a conceptual lexicon". In *Proceedings of the 10<sup>th</sup> Workshop on Multiword Expressions (MWE)*, European Association of Computational Linguistics (EACL), Gothenburg, Sweden.
- Kipper-Schuler, Karin. 2005. *VerbNet: A broad-coverage, comprehensive verb lexicon*. PhD diss. Computer and Information Science Dept., University of Pennsylvania, Philadelphia, PA, June.
- Fillmore, Charles J., Charles Wooters, and Collin F. Baker. 2001. "Building a Large Lexical Databank Which Provides Deep Semantics". In *Proceedings of the Pacific Asian Conference on Language, Information and Computation*. Hong Kong.
- Available from: <http://framenet.icsi.berkeley.edu/papers/dsemlex16.pdf>  
Accessed 21 June 2007
- Gangemi, Aldo, Nicola Guarino, Claudio Masolo, and Alessandro Oltramari. 2003. "Sweetening WordNet with DOLCE". *AI Magazine* 24(3): Fall 2003, 13–24.
- Guarino, Nicola and Christopher A. Welty. 2004. "An Overview of OntoClean". Στο Staab, Steffen, and Rudi Studer (επιμ.) *The Handbook on Ontologies*. 151–172. Berlin:Springer-Verlag.
- Markantonatou, Stella, and Aggeliki Fotopoulou. 2007. "The tool Ekfrasi". In *Proceedings of the 8th International Conference on Greek Linguistics, The Lexicography Workshop*. Ioannina, Greece (in greek).
- Niles, Ian and Adam Pease. 2001. "Origins of the IEEE, Standard Upper Ontology". In *Working Notes of the IJCAI-2001 Workshop on the IEEE Standard Upper Ontology*, AAAI Press.
- Mahesh, Kavi and Sergei Nirenburg. 1995a. "A Situated Ontology for Practical NLP". *Proceedings of the Workshop on Basic Ontological Issues in Knowledge Sharing*, International Joint Conference on Artificial Intelligence. Montreal, Canada. August.

- Mahesh, Kavi and Sergei Nirenburg. 1995b. "Semantic Classification for Practical Natural Language Processing". In *Proceedings of the 6th ASIS SIG/CR Classification Research Workshop: An Interdisciplinary Meeting*. Chicago, IL.
- Pease, Adam. 2011. *Ontology: A Practical Guide*. Articulate Software Press, Angwin, CA.
- Hatzigeorgiu, Nikos, Maria Gavrilidou, Stelios Piperidis, Georgios Carayannis, Anna Papakostopoulou, Anna Spiliotopoulou, Anna Vacalopoulou, Penny Labropoulou, Elena Mantzari, Haris Papageorgiou, and Iason Demiros. 2000. "Design and implementation of the online ILSP Greek Corpus." In *Proceedings of the LREC'00 Conference*, 1737–1742.

