



FACHLITERATUR
EDITION ROMIOSINI
ΕΠΙΣΤΗΜΗ



ICGL12 | 12th INTERNATIONAL CONFERENCE
ON GREEK LINGUISTICS
16 – 19 SEPTEMBER 2015
FREIE UNIVERSITÄT BERLIN, CEMOG

Proceedings of the ICGL12

vol. 2

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. 2

© 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 978-3-946142-35-5
Printed in Germany

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

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>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 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: Veriäten - 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 Περγέα: <i>ΣΕΠΙAME2: 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 NEK 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 Kyrioti: <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 Γαβριλίδου: <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 Gemimates</i>	1007
Liana Tronci: <i>At the lexicon-syntax interface Ancient Greek constructions with <i>ἔχειν</i> 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 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 Georgharas, 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

CONTRASTING THE CONCEPTUAL LEXICON OF ILSP (CL-ILSP) WITH MAJOR LEXICOGRAPHIC EXAMPLES

Georgia Maniati, Voula Gotsoulia & Stella Markantonatou
Institute for Language and Speech Processing/“Athena” RC
ge.maniati@gmail.com, vgotsoulia@gmail.com, marks@ilsp.gr

Περίληψη

Αντιπαραβάλλουμε το εννοιολογικό λεξικό της Ελληνικής (CL-ILSP) που αναπτύσσει το Ινστιτούτο Επεξεργασίας του Λόγου με τρία λεξικογραφικά παραδείγματα: το «Αντιλεξικόν», το WordNet και το FrameNet. Δείχνουμε πώς το εννοιολογικό μοντέλο του CL-ILSP συνδυάζει τα πλεονεκτήματα των πιο πάνω πόρων αντλώντας την οργανωτική προοπτική του από τη διάκριση μεταξύ Σημαίνοντος και Σημαινομένου. Κεντρικό ρόλο έχει η αναπαράσταση της σημασιολογικής συνάφειας των λέξεων μέσω της εκφραστικής δύναμης των οντολογιών. Η σύγκριση βασίζεται στα σημασιολογικά πεδία Περί Υγείας, Περί Τέχνης και Περί Εμπορικής Συναλλαγής που έχουν ήδη κωδικοποιηθεί και περιέχουν περισσότερες από 2000 λέξεις.

Keywords: lexicology, lexicography, lexical semantics, conceptual lexicon, ontology, NLP

1. Introduction

Conceptually organised lexica, such as Roget’s Thesaurus of English Words (Roget 1911) have been highly influential in the international lexicography. In Greek lexicography, Antilexikon by Vostantzoglou (1962) is a prominent example of a lexical resource linking semantically related words. Driven by advances in (psycho)-linguistic and NLP research in the 20th century, WordNet (Fellbaum 1998) and FrameNet

(Baker et al. 1998) were developed independently of each other and constitute major paradigms in the domain of conceptually organised lexica.

Conceptually organised lexica are useful for looking up words with few clues, solving the “tip of the tongue problem”, e.g. one can look up the word “surgery” starting from the verb “to injure”. In addition, various types of NLP applications (e.g. information retrieval, question answering, etc.) can benefit from modeling lexical similarity at a semantic level. To this end, besides hand-coded resources, various lexical similarity models have been developed; latent semantic analysis (Landauer and Dumais 1997), probabilistic topic models (Griffiths et al. 2007), distributional semantic models mapping words to vectors capturing their linguistic context, i.e. co-occurrences with other words (Padó and Lapata 2007, Baroni and Lenci 2010).

As regards Modern Greek, significant alphabetically organised lexical resources have been published relatively recently. Also, a small number of lexical resources that attempt a conceptual organisation, mostly restricted to relations such as synonymy and antonymy, have been made available; Enniolexon (Malakos 2005), Lexiko Synonymon-Antonymon (Babiniotis 2011), Neurolingo (Tsalidis et al. 2009) and the ILSP/“Athena” RC conceptually organised lexicon (CL-ILSP from now on) (Markantonatou and Fotopoulou 2007, Markantonatou et al. 2010), among others.

Conceptually organised lexical resources are evaluated against language coverage and organisation of the lexical material. CL-ILSP’s strengths, compared to key lexical semantic resources that group together words with similar meanings, lie at its organisation principles. CL-ILSP both amplifies WordNet’s descriptive power and maps FrameNet’s complex semantic networks to simpler conceptual structures. Below, we describe CL-ILSP’s organisational perspective and emphasize on the modelling of the notion of semantic similarity/relatedness. CL-ILSP is encoded with the Protégé ontology-editing tool (Noy et al. 2003), so that relations among words that are only implicit in printed lexica become explicit.

2. Background: key lexical semantic resources

2.1. Antilexikon

Antilexikon (Vostantzoglou 1962), that draws heavily on Roget (1911), is the only conceptually organised Modern Greek dictionary. It is an onomasiological dictionary (it

leads the user from relatively known to relatively unknown words) contrary to traditional dictionaries that use known words to interpret unknown ones (Hartmann 1998).

Antilexikon is structured in hierarchically organised thematic fields. Thematic fields are represented as nodes in the hierarchy; they are labeled with one word, one “local top concept”, and subsume semantically related words and phrases. Under each concept node, the words are further classified into “modules” containing synonyms and related words. The semantic relation among these “modules” is meaning similarity whose precise nature is left unspecified. Antilexikon groups together all semantically related words and phrases regardless of lexical category; it clearly is much more than a lexicon of synonyms/antonyms.

Antilexikon consists of three main parts: the local top concept hierarchy, the main part and an index of the linguistic material. At the concept hierarchy, Vostantzoglou followed the organisation and classification of the concepts made by Roget (Trapalis et al. 2005). In the main part, linguistic material is grouped into modules subsuming semantically related words, distinguished in terms of individual senses. Finally, linguistic material is organised in alphabetical order in the index, where for each word the number of the concept under which it has been classified is given.

Concepts are bifunctional in Antilexikon (Trapalis et al. 2005). They function as semantic nodes under which the linguistic material is organized, and they form the liaison between the concept hierarchy and the main part of the dictionary that contains the lexical entries.

Antilexikon’s aim was educational (i.e. to make speakers aware of the expressive richness of Modern Greek) as well as facilitative to speakers who sought for the right word in each communication situation. Although Antilexikon can be used as the basis for the creation of any conceptual lexicon for Modern Greek, Trapalis et al. (2005) note that its material needs to be updated. We would add that relations among words need to be made explicit to facilitate both humans and the machines.

2. 2. *WordNet*

WordNet (Fellbaum 1998) is the largest database of lexical sense relations for English. It can be viewed as a database of non-domain specific open class words, consisting of three individual databases; one for nouns, one for verbs and one for both adjectives and adverbs.

Each of the three databases contains sets of words, where each word is annotated

with a set of senses (Jurafsky and Martin 2009). Each sense has a dictionary style definition (gloss), a list of near synonymous words (synset) and, in some cases, examples of usage. WordNet's representational unit is the synset, that is a list of synonymous words representing a concept. Each word is classified into one or more synsets (e.g. nouns have an average of 1.24 synsets) and in each synset it has a unique sense, e.g. the word "trade" is part of the WordNet synset {deal, trade, business deal} with the gloss "a particular instance of buying or selling" and of the synset {trade wind, trade} with the gloss "steady winds blowing from east to west above and below the equator" among others. As the databases are split according to lexical category, each synset contains only words of the same lexical category: WordNet resembles a lexicon of synonyms.

WordNet differs from an onomasiological dictionary as it explicitly specifies a set of relations, namely semantic relations between word meanings and lexical relations between word forms. Semantic relations are defined among synsets, with the hyperonymy/hyponymy or *Is-a* relation (Miller 1995) being the backbone of the database's organisation. That is, each synset is related to its immediately more general and more specific synsets and eventually all synsets are connected to the root (Entity for the noun hierarchy). Other important relations are meronymy (*Part-whole*) and antonymy (*Opposites*).

WordNet is commonly used for word sense disambiguation, among other tasks.

2. 3. *FrameNet*

The FrameNet project (Baker et al. 1998) is a lexical database for English that models lexical meaning in terms of the semantics of events. It is based on the theory of Frame Semantics (Fillmore 1982), according to which every word evokes a semantic frame, and it has been built drawing on annotated example sentences from large corpora.

FrameNet's representational unit is the semantic frame, a linguistic entity that structures experiential entities and corresponds to descriptions of prototypical events and their participants. Frame evoking elements (words) have similar argument structures as well as number and type of semantic arguments. Roles specific to a frame are called frame elements (FEs). These can be core elements, that are always present when a specific frame is evoked distinguishing it from others, the so-called non-core elements, that can be realised in more than one semantic frames.

For example, the Commerce-Buy frame subsumes the words "buy.v", "buyer.n", "client.n", "purchase.n", "purchase.v", "purchaser.n" and is defined as follows (Fillmore 2006):

These are words describing a basic commercial transaction involving a *Buyer* and a *Seller* exchanging *Money* and *Goods*, taking the perspective of the *Buyer*. The words vary individually in the patterns of frame element realization they allow. For example, the typical pattern for the verb BUY: Buyer buys Goods from Seller for Money.

Words grouped within the same frame share aspects of their meaning. Verbs, nouns, adjectives, adverbs as well as multiword expressions illustrate the event, and the thematic roles illustrate the semantic relation of a predicate with its arguments. Inheritance is the strongest relation between frames and corresponds to a non-strict version of the *Is-a* ontological relation.

FrameNet furnishes training data for semantic role labeling algorithms, by specifying where the arguments of each predicate are likely to be found (Baker et al. 1998). It also makes a valuable lexicographic contribution by documenting the use of individual words.

3. CL-ILSP

3.1. CL-ILSP's conceptual model: classes and relations

CL-ILSP distinguishes the word level from the semantic level. Grammatical and functional properties of lexical items are represented by the ontology branch rooted at the Signifier class, i.e. the word level. Word meanings are represented by the ontology branch rooted at the Signified class, i.e. the semantic level, and are linked with lexical items at the Signifier branch (Figure 1).

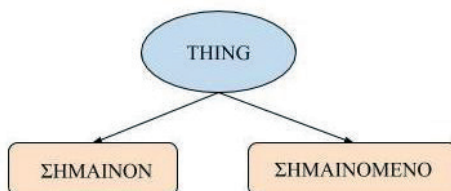


Figure 1 | The upper part of the ontology: the root *Thing* expands into the *Signifier* (ΣΗΜΑΙΝΟΝ) and the *Signified* (ΣΗΜΑΙΝΟΜΕΝΟ) branches.

At the word (i.e. the Signifier) level (Figure 2), lexical items (including collocations and multiword expressions) are represented as instances of classes, linked to one another in terms of relations encoding different types of properties. Lexical relations in CL-ILSP include morphological relations (*Has allomorph*, *Is morphologically related to*, etc.), lexical semantic relations (*Has synonym*, *Has antonym*), functional relations (*Has translation*, *Has specific use*). Via the *Instantiates* relation, lexical items are linked to instances of semantic classes (word meanings) rooted at the Signified. In other words, the *Instantiates* relation relates the signifier with the signified.

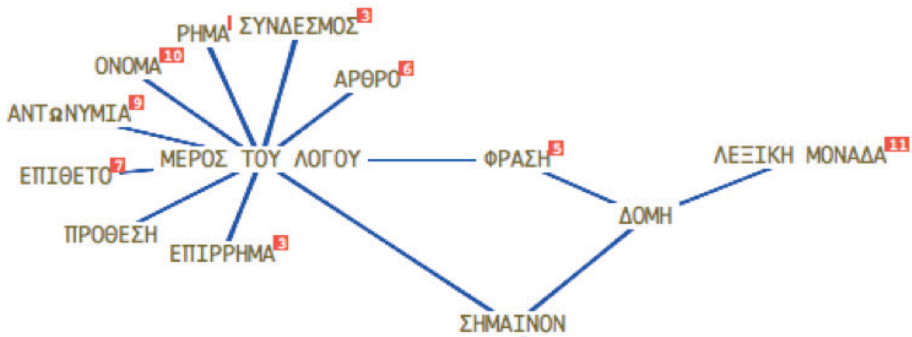


Figure 2 | The Signifier (ΣΗΜΑΙΝΟΝ) level of the ontology

At the semantic (i.e. the Signified) level, semantic classes (called Concepts) directly subsume other semantic classes, capturing the semantics of events or states-of-affairs (soas) and participating entities. Instances/sub-concepts of the semantic classes, called Lexical Classes are related with instances of the Signifier and may group together words that are interchangeable in some contexts. They represent word senses and provide the means for representing synonymy.

As illustrated in Figure 3 below, Περί Εμπορικής Συναλλαγής (About Commercial Transaction) and Περί Εμπόρου (About Merchant) are Concepts, i.e. classes at the Signified branch of the hierarchy. «Κάθε φυσικό πρόσωπο που ασχολείται με το εμπόριο» (any individual who is occupied within the field of trade) is an instance of Περί Εμπόρου. «Εμπορας», «εμπορικός», «εμπορευόμενος» (merchant) are instances of the Signifier branch that are related with «κάθε φυσικό πρόσωπο που ασχολείται με το εμπόριο» with the relation *Instantiates/Instantiated in Language by*. Similarly,

«εμπορικός» (commercial) is an instance of Signifier that is related with *Instantiates/Instantiated in Language* by relation with «που σχετίζεται με τον έμπορο» that is an instance of the class Περί Εμπορού.



Figure 3 | The lexical class of trader within the Commercial Transaction field: (i) classes of Signified (upper case characters), (ii) instances of Signified (lower case letters without a red index) (iii) instances of Signifier (lower case characters with a red index) related with instances of Signified.

There are two kinds of relations at the Signified level: those linking classes across the various branches of the hierarchy, and the *Is about* (*Περί*) relation that links concepts in a mother-daughter relationship. The former relations represent lexical semantic relations such as *Has hypernym*, *Has hyponym*, *Afflicts entity*, *Entity afflicted by*, *Entity that causes*, *The effect of*, *Has the quality of*, *Is the quality of*, *Part of*, *Consists of*, *Is term of*, *Is the action of*. The latter (*Is about*) represents the notion of semantic relatedness. Semantic relatedness in the lexicon is structured in terms of the semantics of events and state-of-affairs, and entities participating in these (see example Figure 4).

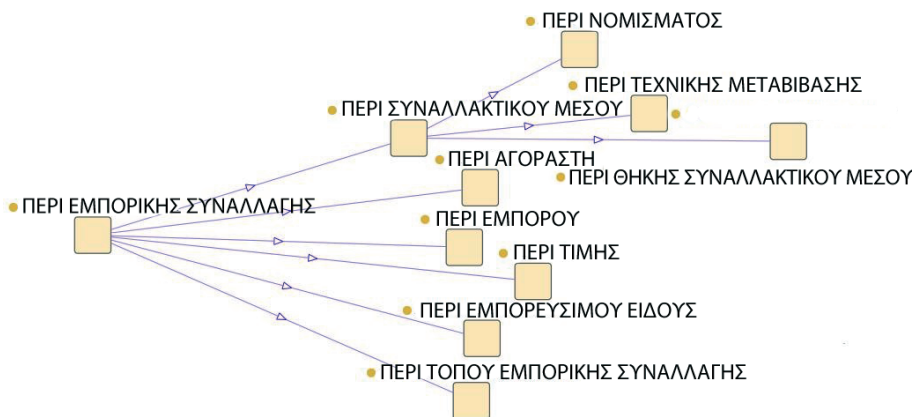


Figure 4 | The structure of the Commercial Transaction field (ΠΕΡΙ ΕΜΠΟΡΙΚΗΣ ΣΥΝΑΛΛΑΓΗΣ). The semantic classes (concepts) are linked to it via the Is about relation.

Up to the present, more than 2000 words have been encoded in CL-ILSP, covering 4 semantic fields (Table 1).

Semantic field	Concepts	Lexical classes	Lexical items	Relations per Lexical class (AVG)	Relations per Lexical item (AVG)
Commercial transaction	9	412	648	3,12	1,51
Health	17	560	1.028	3,22	1,72
Art	11	193	317	3,13	1,47
Body parts^{1*}	0	51	69	2,5	1,66
Total	37	1.216	2.062	3,14	1,6

Table 1 | CL-ILSP's semantic fields, population and relations.

1 The "Body parts" semantic field has not been fully structured yet.

3.2 Creation and population of the About Commercial Transaction semantic class

For the development and the population of the classes of CL-ILSP, a mixture of top-down and bottom-up procedures has been reported (Tzortzi and Markantonatou 2014). We follow the same technique while developing the About Commercial Transaction field.

The definition of the top concept About Commercial Transaction was made on the basis of Antilexikon's organisation, where the top concept Transaction (Δοσοληψία) is situated under the general node Values (top-down procedure). From this point on, the procedure was corpus-driven (bottom-up), and resulted in the definition of 6 "lower" concepts and the encoding of 648 words.

The corpus-driven population of the concept was made on the basis of a small number of verbs undoubtedly relevant to the event of commercial transaction: «αγοράζω» (to buy), «ψωνίζω» (to shop), «ξοδεύω» (to spend), «πουλάω» (to sell), «εμπορεύομαι» (to merchandise/to trade). This set of verbs is used as a "seed" in order to delineate and populate the semantic field with concepts. Note that the first three seed verbs were chosen so as to take the buyer's perspective on the event of commercial transaction while the last two adopt the seller's view of the event.

A large number of sentences containing the seed verbs were retrieved from the Hellenic National Corpus (HNC)² and retrieved examples were annotated for the semantic function of the verbal complements. For example, given the buyer's perspective, the agent semantic function "points" to the type of entity Buyer, while patient "points" to the type of entity Commodity. Eventually, the roles/semantic functions that systematically appeared in the seed corpus sentences, formed the About Commercial Transaction sub-concepts. As illustrated in Figure 4 above, these are: About Buyer, About Merchant, About Commodity, About Means of Transaction, About Case/Receptacle of Means of Transaction, About Currency, About Method of Transfer, About Place of Commercial Transaction.

Based on our corpus driven research, we ended up incorporating a number of distinct Antilexikon classes within the general concept About Commercial Transaction. In Antilexikon, these classes are situated "close" to the Transaction class and under the general node Values. Among them are: Price, High Price, Low Price, Commodity

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

(Εμπόρευμα), Merchant (Εμπορος), Buy(n) (Αγορά), Sale (Πώληση), Place of Transaction (Τόπος Συναλλαγών), Money (Χρήμα), Currencies (Νομίσματα), Treasury (Ταμείο). We also incorporated lexical units of at least 21 separate Antilexikon classes (professionals and people who provide goods or services in exchange for money).

3.3. Evaluation of CL-ILSP with respect to key lexicographic examples

3.3.1. Antilexikon

In Antilexikon, semantic relations among concepts are left implicit. In CL-ILSP, semantic relations are explicitly encoded at the Signified branch. In particular, the explicit *Is about* semantic relation among the concepts facilitates both humans and machines, because it provides a consistent modeling of the unspecified notion of meaning similarity. In Antilexikon, the meaning similarity relation that connected the individual “modules” under each general concept was left unspecified. Furthermore, in CL-ILSP, as in Anilexikon, a word can “be part of” different classes, through multiple inheritance.

Finally, in comparison to Antilexikon, that was last edited in 1962, CL-ILSP’s developed semantic fields reflect the current state of the Greek language. For the population of CL-ILSP, Antilexikon was used as a starting point, but obsolete (no-longer in use) vocabulary was omitted and new words were obtained from a combination of resources (*Lexiko tis Kinis Neoellinikis*, Babiniotis 2002, Kriaras 1995).

3.3.2. WordNet

CL-ILSP’s instances of the Signified that are instantiated by lexical items (instances of the Signifier) are roughly equivalent to WordNet synsets. CL-ILSP’s strength, compared to WordNet, comes from the combination of fine-grained lexical semantic relations, such as hyponymy, with a general, conceptual relation, the *Is about* relation. It is the *Is about* relation that yields a much richer semantic organisation of words and their meanings.

In WordNet, for example, the word “trader” is associated with the words “bargainer”, “dealer”, “monger” (members of the same synset). The word “merchant” forms a synset with the word “merchandise”. None of these words is associated with lexical items such as “contraband”, “bargain”, “cost”, “bank”, “save money”, “bidder”, etc. CL-ILSP’s con-

ceptual model, by contrast, allows such connections. The semantic classes instantiated by lexical items such as «λαθρεμπόριο» (contraband), «κοστίζω» (to cost), «τράπεζα» (bank), «αποταμιεύω/βάζω λεφτά στην άκρη» (to make savings), «πλειοδότης» (bidder) are allowed to be linked, via the *Is about* relation, to the class Περί Εμπορικής Συναλλαγής (About Commercial Transaction), or to the classes Περί Εμπόρου (About Merchant) and Περί Συναλλακτικού Μέσου (About Means of Commercial Transaction), that are in turn related with Περί Εμπορικής Συναλλαγής.

CL-ILSP currently lacks some of the (semantic) relations employed in WordNet. Some of these relations, such as the troponym relation, are dealt with the *Is about* relation. The *Is about* relation seems to be more appropriate than troponymy because it allows for a relation among events and their subevents, and among events and their main types of participants. It is not an *Is-a* relation that poses problems when it is applied on events. Furthermore, its ontology of lexical and semantic relations is consistently structured to enable the distinction of different levels of linguistic analysis: for example, the lexical relation *Is Morphologically Related to*, that links words such as «θεραπεύω» (to cure) and «θεραπεία» (treatment), has semantic counterparts, e.g. the relation *Is the Action of*, linking the semantic classes instantiated by «θεραπεύω» and «θεραπεία» (i.e. «ανάκτηση της υγείας» (recovery) *Is the Action of* «αποκαθιστώ μια βλάβη ή ανωμαλία του οργανισμού» (to restore a fault or abnormality of the organism)).

3.3.3. FrameNet

In terms of associated words, concepts in CL-ILSP may be equivalent to a frame in FrameNet. For instance, the class Περί Θεραπείας roughly corresponds to the Cure frame. More frequently, words associated with a concept in CL-ILSP will evoke more than one FrameNet frames, that may or may not be connected to one another with some kind of frame relation. For instance, classes instantiated by the words «πουλάω» (to sell), «αγοράζω» (to buy), «εξάγω» (to export), «εμπόριο» (trade), «πώληση» (sale), «αγορά» (purchase), «εξαγωγικό εμπόριο» (export trade) are all encoded under the concept Περί Εμπορικής Συναλλαγής (About Commercial Transaction). In FrameNet, the English equivalents of the words above evoke a number of distinct frames, i.e. Commerce Buy, Commerce Sell, Commerce Scenario, and Exporting. Frame definitions are semantic interpretations of words' argument structure properties, and frame-to-frame relations are bound to these definitions.

The above frames are connected in terms of the following relations (3):

- Commerce Buy *Is Perspective on* Commerce Goods Transfer
- Commerce Sell *Is Perspective on* Commerce Goods Transfer
- Commerce Goods Transfer *Is Subframe of* Commercial Transaction
- Commercial Transaction *Is Subframe of* Commerce Scenario
- Exporting *Uses* Commerce Sell

Thus, relating the words “buy” and “commerce” requires the successive establishment of frame-to-frame relations as depicted in the above complex structure. CL-ILSP’s strength lies in that its *Is about* relation (roughly equivalent to FrameNet’s *Uses* relation) implements a much simpler and intuitive semantic organisation of words related to events and their participants, leaving frames’ semantic components (underlying FrameNet’s complex network of relations) out of its scope or, put differently, purposefully underspecified.

Furthermore, words such as «έμπορος» (merchant) and «εμπόριο» (trade) that in CL-ILSP are subsumed by the concept Περί Εμπορικής Συναλλαγής (About Commercial Transaction), seem to remain unrelated in FrameNet. In FrameNet, the word “trader” evokes the People by Vocation frame that currently bears no relation to any of the frames in (3) or other frames related to the commercial transaction event. Encoding “trader” both in People by Vocation and e.g. in the Commerce Scenario frame would wrongly indicate that the word is associated with two distinct senses, as the meaning of “trader” is the same in both frames. In CL-ILSP such examples are modeled with multiple inheritance relations: that is, the word «έμπορος» (merchant) is associated with the single low-level semantic class «κάθε φυσικό πρόσωπο που ασχολείται με το εμπόριο» (any individual who is occupied in the sector of trade), but this class, that captures the word sense, is allowed to be linked with more than one Concepts (i.e. one related to professions and another related to the commercial transaction event) as illustrated in Figure 3.

4. Conclusions and future research directions

To conclude, we note that while CL-ILSP combines the strengths of both WordNet and FrameNet (WordNet-like fine-grained lexical relations, FrameNet-like rich semantic organisation), clearly adding to the descriptive efficiency of WordNet and mapping FrameNet’s complex semantic networks to simpler conceptual structures, its organi-

sational perspective is different than that of either of these resources. Relying on the expressive power of ontologies, it explicitly sets out to model the notion of semantic relatedness of words, treating it as ontologically separate from linguistic properties of words, such as argument structure, that are modelled at the lexical, not the semantic level.

Following this organisational evaluation of the resources, future directions lie into evaluating the semantic model of CL-ILSP in terms of a specific natural language processing application (e.g. question answering).

As CL-ILSP's development and lexical coverage expansion is in progress, a reengineering of the lexicon is being designed with the emphasis to be given on enhancements of its semantic organisation. Distributional semantic models are to be explored so as to automatically acquire CL-ILSP's concepts and speed up the process of populating the lexicon, while any morpho-syntactic information is to be automatically acquired from other databases conversing with the dictionary, as has been done for the "Idion" MWE database (Markantonatou et al. 2015).

References

- Babinotis, Georgios. 2011. *Lexiko Synonymon-Antonymon tis Neas Ellinikis Glossas. Lexilogikos Thisavros*. Athens: Centre of Lexicology. http://www.lexicon.gr/lexica/fbook_syn/index.html.
- Babinotis, Georgios. 2002. *Lexiko tis Neas Ellinikis Glossas*. Athens: Centre of Lexicology.
- Baker, Collin F., Charles J. Fillmore, and John B. Lowe. 1998. "The Berkeley FrameNet Project." In *Proceedings of the 17th International Conference on Computational Linguistics* 1: 86–90.
- Baroni, Marco, and Alessandro Lenci. 2010. "Distributional Memory: A general Framework for Corpus-based Semantics." *Computational Linguistics* 36(4): 673–721.
- Blacoe, William, and Mirella Lapata. 2012. "A Comparison of Vector-based Representations for Semantic Composition." In *Proceedings of the 2012 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning* 546–556.

- Fellbaum, Christiane. 1998. *WordNet: An Electronic Lexical Database*. Cambridge, Massachusetts: MIT Press.
- Fillmore, Charles J. 1982. "Frame Semantics." In *Linguistics in the Morning Calm* 111–137. Seoul: Hanshin Publishing Co.
- Fillmore, Charles J. 2006. "FrameNet. Website of Berkeley FrameNet Project. Frame Index: Commerce_buy." Accessed June 25, 2016. https://framenet2.icsi.berkeley.edu/fnReports/data/frameIndex.xml?frame=Commerce_buy.
- Griffiths, Thomas L., Mark Steyvers, and Joshua B. Tenenbaum. 2007. "Topics in Semantic Representation." *Psychological Review* 114(2):211–244.
- Hartmann, Reinhard R. K., and James, Gregory. 1998. *Dictionary of Lexicography*. London: Routledge.
- Jurafsky, Daniel, and James H. Martin. 2009. *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Prentice Hall, Second Edition.
- Kriaras, Emmanouil. 1995. *Neo Elliniko Lexiko tis Sinxronis Dimotikis Glossas*. Athens: Ekdotiki Athinon.
- Landauer, Thomas K., and Susan T. Dumais. 1997. "A Solution to Plato's Problem: The Latent Semantic Analysis Theory of Acquisition, Induction and Representation of Knowledge." *Psychological Review* 104:211–240.
- Lexiko tis Kinis Neoellinikis*. Thessaloniki: Institutou Neoellinikon Spoudon.
- Malakos, Georgios. 2005. *Enniolexo*. Athina: Lexigram. <http://apps.lexigram.gr/elli-nognosiademo/ennoiolekso.php#Hist0>.
- Markantonatou, Stella, Erasmia Koletti, Elpiniki Margariti, Panagiotis Minos, Amalia Stripeli, Georgios Zakis, and Niki Samaridi. 2015. "Lexical Resource for free Subject Verb MWEs." Poster presented at 4th PARSEME General Meeting, Valetta, Malta, March 19-20.
- Markantonatou, Stella, Aggeliki Fotopoulou, Maria Mini, and Maria Alexopoulou. 2010. "In Search of the Right Word". In *Proceedings of Cogalex-2: Cognitive Aspects of the Lexicon, 2nd SIGLEX endorsed Workshop* 66-74.
- Markantonatou, Stella, and Aggeliki Fotopoulou. 2007. "The tool Ekfrasi". Paper presented at 8th International Conference on Greek Linguistics, The Lexicography Workshop, Ioannina, Greece, August 29-September 2.
- Miller, George A. 1995. "WordNet: A Lexical Database for English". *Communications of the ACM* 38(11):39–41.
- Mitchell, Jeff, and Mirella Lapata. 2010. "Composition in Distributional Models of

- Semantics. *Cognitive Science* 34(8):1388–1429.
- Noy, Natalya F., Monica Crubézy, Ray W. Ferguson, Holger Knublauch, Samson W. Tu, Jennifer Vendetti, and Mark A. Musen. 2003. “Protégé-2000: An Open-source Ontology-Development and Knowledge-Acquisition Environment”. In *AMIA Annual Symposium Proceedings* 953.
- Padó, Sebastian, and Mirella Lapata. 2007. “Dependency-based Construction of Semantic Space Models.” *Computational Linguistics* 33(2):161–199.
- Roget, Peter Mark. 1911. *Roget’s Thesaurus of English Words and Phrases*. TY Crowell Company.
- Trapalis, Giorgos, Stella Markantonatou, Maria Alexopoulou, Aggeliki Fotopoulou, and Yianis Maistros. 2005. “To Antilexikon tou Th.Vostantzoglou ke i Dimiourgia mias mikris klimakas Simasiologika Organomenis Vasis Dedomenon tis Sinxronis Ellinikis”. Paper presented at 7th International Conference on Greek Linguistics, York, UK, September 8-10.
- Tsalidis, Christos, Mavina Pantazara, Panagiotis Minos, and Elena Mantzari. 2009. “NLP Tools for Lexicographic Applications in Modern Greek”. In *eLexicography in the 21st Century: New Challenges, New Applications* 457–462.
- Tzortzi, Katerina, and Stella Markantonatou. 2014. “Development of a Conceptual Lexicon with Ontological Techniques”. Paper presented at 8th Conference on Terminology and Ontology: Theories and Applications (TOTh 2014), Chambéry, France, June 12-13.
- Vostantzoglou, Theologos. 1962. *Antilexikon i Onomastikon tis Neas Ellinikis Glossis*. Athens: Domi.

