



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. 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 12<sup>TH</sup> INTERNATIONAL  
CONFERENCE ON GREEK LINGUISTICS**

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

**VOL. 2**

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

Online-Bibliothek der Edition Romiosini:  
[www.edition-romiosini.de](http://www.edition-romiosini.de)

## ΠΕΡΙΕΧΟΜΕΝΑ

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

|                                                                                                                                                                                                     |     |
|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|-----|
| Γεωργία Ανδρέου & Ματίνα Τασιούδη:<br><i>Η ανάπτυξη του λεξιλογίου σε παιδιά με Σύνδρομο Απνοιών στον Ύπνο</i> .....                                                                                | 175 |
| Ανθούλα- Ελευθερία Ανδρεσάκη:<br><i>Ιατρικές μεταφορές στον δημοσιογραφικό λόγο της κρίσης: Η οπτική γωνία των Γερμανών</i> .....                                                                   | 187 |
| Μαρία Ανδριά:<br><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:<br><i>Mastering word-initial syllable onsets by Cypriot Greek toddlers with and without early language delay</i> .....                                             | 215 |
| Julia Bacskai-Atkari:<br><i>Ambiguity and the Internal Structure of Comparative Complements in Greek</i> .....                                                                                      | 231 |
| Costas Canakis:<br><i>Talking about same-sex parenthood in contemporary Greece: Dynamic categorization and indexicality</i> .....                                                                   | 243 |
| Michael Chiou:<br><i>The pragmatics of future tense in Greek</i> .....                                                                                                                              | 257 |
| Maria Chondrogianni:<br><i>The Pragmatics of the Modern Greek Segmental Markers</i> .....                                                                                                           | 269 |
| Katerina Christopoulou, George J. Xydopoulos & Anastasios Tsangalidis:<br><i>Grammatical gender and offensiveness in Modern Greek slang vocabulary</i> .....                                        | 291 |
| Aggeliki Fotopoulou, Vasiliki Foufi, Tita Kyriacopoulou & Claude Martineau:<br><i>Extraction of complex text segments in Modern Greek</i> .....                                                     | 307 |
| Αγγελική Φωτοπούλου & Βούλα Γιούλη:<br><i>Από την «Έκφραση» στο «Πολύτροπο»: σχεδιασμός και οργάνωση ενός εννοιολογικού λεξικού</i> .....                                                           | 327 |
| Marianthi Georgalidou, Sofia Lampropoulou, Maria Gasouka, Apostolos Kostas & Xanthippi Foulidi:<br><i>“Learn grammar”: Sexist language and ideology in a corpus of Greek Public Documents</i> ..... | 341 |
| Maria Giagkou, Giorgos Fraggakis, Dimitris Pappas & Harris Papageorgiou:<br><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:<br><i>The Diachronic Corpus of Greek of the 20th century: Design and compilation</i> .....                                                                                                    | 369 |
| Kleanthes K. Grohmann & Maria Kambanaros:<br><i>Bilectalism, Comparative Bilingualism, and the Gradience of Multilingualism:<br/>A View from Cyprus</i> .....                                                                                                                         | 383 |
| Günther S. Henrich:<br>„Γεωγραφία νεωτερική“ στο Λίβιστρος και Ροδάμνη: μετατόπιση ονομάτων βαλτικών<br>χωρών προς την Ανατολή; .....                                                                                                                                                 | 397 |
| Noriyo Hoozawa-Arkenau & Christos Karvounis:<br><i>Vergleichende Diglossie - Aspekte im Japanischen und Neugriechischen:<br/>Veriäten - Interferenz</i> .....                                                                                                                         | 405 |
| Μαρία Ιακώβου, Ηριάννα Βασιλειάδη-Λιναρδάκη, Φλώρα Βλάχου, Όλγα Δήμα, Μαρία<br>Καββαδία, Τατιάνα Κατσίνα, Μαρίνα Κουτσομπού, Σοφία-Νεφέλη Κύτρου, Χριστίνα<br>Κωστάκου, Φρόσω Παππά & Σταυριαλένα Περρέα:<br><i>ΣΕΠΙΜΕ2: Μια καινούρια πηγή αναφοράς για την Ελληνική ως Γ2</i> ..... | 419 |
| Μαρία Ιακώβου & Θωμαΐς Ρουσουλιώτη:<br><i>Βασικές αρχές σχεδιασμού και ανάπτυξης του νέου μοντέλου αναλυτικών<br/>προγραμμάτων για τη διδασκαλία της Ελληνικής ως δεύτερης/ξένης γλώσσας</i> .....                                                                                    | 433 |
| Μαρία Καμηλάκη:<br>«Μαζί μου ασχολείσαι, πόσο μαλάκας είσαι!»: Λέξεις-ταμπού και κοινωνιογλωσσικές<br>ταυτότητες στο σύγχρονο ελληνόφωνο τραγούδι .....                                                                                                                               | 449 |
| Μαρία Καμηλάκη, Γεωργία Κατσούδα & Μαρία Βραχιονίδου:<br><i>Η εννοιολογική μεταφορά σε λέξεις-ταμπού της ΝΕΚ και των νεοελληνικών<br/>διαλέκτων</i> .....                                                                                                                             | 465 |
| Eleni Karantzola, Georgios Mikros & Anastassios Papaioannou:<br><i>Lexico-grammatical variation and stylometric profile of autograph texts<br/>in Early Modern Greek</i> .....                                                                                                        | 479 |
| Sviatlana Karpava, Maria Kambanaros & Kleanthes K. Grohmann:<br><i>Narrative Abilities: MAINing Russian–Greek Bilingual Children in Cyprus</i> .....                                                                                                                                  | 493 |
| Χρήστος Καρβούνης:<br><i>Γλωσσικός εξαρχαϊσμός και «ιδεολογική» νόρμα: Ζητήματα γλωσσικής διαχείρισης<br/>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:<br><i>Contrasting the Conceptual Lexicon of ILSP (CL-ILSP) with major lexicographic examples</i> ..... | 709 |
| Γεώργιος Μαρκόπουλος & Αθανάσιος Καρασίμος:<br><i>Πολυεπίπεδη επισημείωση του Ελληνικού Σώματος Κειμένων Αφασικού Λόγου</i> .....                               | 725 |
| Πωλίνα Μεσηνιώτη, Κατερίνα Πούλιου & Χριστόφορος Σουγανίδης:<br><i>Μορφοσυντακτικά λάθη μαθητών Τάξεων Υποδοχής που διδάσκονται την Ελληνική ως Γ2</i> .....    | 741 |
| Stamatia Michalopoulou:<br><i>Third Language Acquisition. The Pro-Drop-Parameter in the Interlanguage of Greek students of German</i> .....                     | 759 |
| Vicky Nanousi & Arhonto Terzi:<br><i>Non-canonical sentences in agrammatism: the case of Greek passives</i> .....                                               | 773 |
| Καλομοίρα Νικολού, Μαρία Ξεφτέρη & Νίτσα Παραχεράκη:<br><i>Το φαινόμενο της σύνθεσης λέξεων στην κυκλαδοκρητική διαλεκτική ομάδα</i> .....                      | 789 |
| Ελένη Παπαδάμου & Δώρης Κ. Κυριαζής:<br><i>Μορφές διαβαθμιστικής αναδίπλωσης στην ελληνική και στις άλλες βαλκανικές γλώσσες</i> .....                          | 807 |
| Γεράσιμος Σοφοκλής Παπαδόπουλος:<br><i>Το δίπολο «Εμείς και οι Άλλοι» σε σχόλια αναγνωστών της Lifo σχετικά με τη Χρυσή Αυγή</i> .....                          | 823 |
| Ελένη Παπαδοπούλου:<br><i>Η συνδυαστικότητα υποκοριστικών επιθημάτων με β' συνθετικό το επίθημα -άκι στον διαλεκτικό λόγο</i> .....                             | 839 |
| Στέλιος Πιπερίδης, Πένυ Λαμπροπούλου & Μαρία Γαβριλίδου:<br><i>clarin:el. Υποδομή τεκμηρίωσης, διαμοιρασμού και επεξεργασίας γλωσσικών δεδομένων</i> .....      | 851 |
| Maria Pontiki:<br><i>Opinion Mining and Target Extraction in Greek Review Texts</i> .....                                                                       | 871 |
| Anna Roussou:<br><i>The duality of mipos</i> .....                                                                                                              | 885 |

|                                                                                                                                                                                                                          |      |
|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|------|
| Stathis Selimis & Demetra Katis:<br><i>Reference to static space in Greek: A cross-linguistic and developmental perspective of poster descriptions</i> .....                                                             | 897  |
| Evi Sifaki & George Tsoulas:<br><i>XP-V orders in Greek</i> .....                                                                                                                                                        | 911  |
| Konstantinos Sipitanos:<br><i>On desiderative constructions in Naousa dialect</i> .....                                                                                                                                  | 923  |
| Eleni Staraki:<br><i>Future in Greek: A Degree Expression</i> .....                                                                                                                                                      | 935  |
| Χριστίνα Τακούδα & Ευανθία Παπαευθυμίου:<br><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:<br><i>Corpus C58 and the interface between intra- and inter-sentential linguistic information</i> .... | 961  |
| Arhonto Terzi & Vina Tsakali:<br><i>The contribution of Greek SE in the development of locatives</i> .....                                                                                                               | 977  |
| Paraskevi Thomou:<br><i>Conceptual and lexical aspects influencing metaphor realization in Modern Greek</i> .....                                                                                                        | 993  |
| Nina Topintzi & Stuart Davis:<br><i>Features and Asymmetries of Edge Gemimates</i> .....                                                                                                                                 | 1007 |
| Liana Tronci:<br><i>At the lexicon-syntax interface Ancient Greek constructions with ἔχειν and psychological nouns</i> .....                                                                                             | 1021 |
| Βίλλυ Τσάκωνα:<br><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:<br><i>Εκμάθηση ελληνικών στο Πανεπιστήμιο Στρασβούργου: κίνητρα και αναπαραστάσεις</i> ...                                                                                   | 1051 |
| Stavroula Tsiplakou & Spyros Armotistis:<br><i>Do dialect variants (mis)behave? Evidence from the Cypriot Greek koine</i> .....                                                                                          | 1065 |
| Αγγελική Τσόκογλου & Σύλα Κλειδή:<br><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

CLARIN:EL  
ΥΠΟΔΟΜΗ ΤΕΚΜΗΡΙΩΣΗΣ, ΔΙΑΜΟΙΡΑΣΜΟΥ  
ΚΑΙ ΕΠΕΞΕΡΓΑΣΙΑΣ ΓΛΩΣΣΙΚΩΝ ΔΕΔΟΜΕΝΩΝ

Στέλιος Πιπερίδης, Πένυ Λαμπροπούλου & Μαρία Γαβριηλίδου

Ε.Κ. Αθηνά/ΙΕΛ

spip@ilsp.gr, penny@ilsp.gr, maria@ilsp.gr

*Abstract*

*This paper presents clarin:el (www.clarin.gr), a research infrastructure intended to support multi-disciplinary research conducted through language. clarin:el, the Greek member of the European network CLARIN ERIC, brings together Language Resources and Technologies in an integrated and interoperable environment where providers can deposit and share their resources and consumers can access and use them through dedicated Language Processing Web Services. The paper discusses the principles and objectives of the infrastructure, its architecture and functionalities, and the network enabling its operation and parallel activities.*

*Λέξεις-κλειδιά: υποδομή γλωσσικών πόρων, γλωσσικοί πόροι, τεχνολογίες γλωσσικής επεξεργασίας, διαδικτυακές υπηρεσίες, ανοιχτή πρόσβαση*

## **1. Εισαγωγή**

Η αύξηση του όγκου των ψηφιακών δεδομένων δεν εξασφαλίζει ότι τα δεδομένα αυτά είναι προσβάσιμα, διαθέσιμα και αξιοποιήσιμα από όλους τους ενδιαφερόμενους χρήστες. Στο (πολύ συνηθισμένο, πλέον) σενάριο κατά το οποίο ένας ερευνητής αναζητά γλωσσικά δεδομένα για την έρευνά του, έρχεται αντιμέτωπος με εμπόδια σε πολλαπλά επίπεδα: ενημέρωσης, εντοπισμού, τεχνικής αλλά και νομικής διαλειτουργικότητας.

Είναι γεγονός ότι υπάρχουν πολλές αρχειακές συλλογές σε πανεπιστήμια, ερευνητικά ιδρύματα και άλλους φορείς· όπως υπάρχουν και δεδομένα, εργαλεία επεξεργασίας και σχετικές διαδικτυακές υπηρεσίες που είναι γνωστά μόνο σε ορισμένες ερευνητικές κοινότητες. Δεδομένου ότι τα περισσότερα παραμένουν ασύνδετα μεταξύ τους, δυσχεραίνεται ο εντοπισμός και η αξιοποίησή τους σε ένα ευρύτερο διεπιστημονικό ψηφιακό περιβάλλον. Πρόσθετη πρόκληση αποτελεί η υιοθέτηση διαφορετικών προτύπων τεκμηρίωσης και μορφοποίησης, η οποία οδηγεί σε ασυμβατότητα γλωσσικών δεδομένων και υπηρεσιών/εργαλείων επεξεργασίας τους. Η υπολογιστική ισχύς και ο χώρος που απαιτούνται για την αποθήκευση και την επεξεργασία των γλωσσικών δεδομένων αποτελούν ακόμη έναν αποθαρρυντικό παράγοντα. Τέλος, ο ερευνητής (πάροχος ή καταναλωτής πόρων) αντιμετωπίζει ένα νομικό κυκεώνα σχετικά με τα δικαιώματα και τις υποχρεώσεις του. Όλοι αυτοί οι παράγοντες δυσχεραίνουν τη διάθεση και τον διαμοιρασμό πόρων καθώς και τη συνεργατική έρευνα.

Μια προσέγγιση στη λύση αυτών των προβλημάτων για τον ελληνικό χώρο σκοπεύει να παράσχει η Ερευνητική Υποδομή *clarin:el*.

## 2. Η Ερευνητική Υποδομή CLARIN

Η ευρωπαϊκή Ερευνητική Υποδομή CLARIN ([www.clarin.eu](http://www.clarin.eu))<sup>1</sup> αποτελεί μία ολοκληρωμένη και διαλειτουργική ερευνητική υποδομή Γλωσσικών Πόρων και Τεχνολογιών, που φιλοδοξεί να καταπολεμήσει την ισχύουσα αποσπασματικότητα και να προσφέρει ένα σταθερό, συνεπές, εύχρηστο και επεκτάσιμο περιβάλλον πρόσβασης σε γλωσσικά δεδομένα προς όφελος των ερευνητών όλων των επιστημών. Δημιούργησε το πανευρωπαϊκό δίκτυο οργανισμών CLARIN ERIC, μέσω του οποίου συγκεντρώνονται, τεκμηριώνονται, συντηρούνται και διαμοιράζονται Γλωσσικοί Πόροι, Τεχνολογίες και Υπηρεσίες Γλωσσικής Επεξεργασίας. Οι Γλωσσικοί Πόροι που ενσωματώνει η υποδομή CLARIN<sup>2</sup> περιλαμβάνουν ψηφιακό περιεχόμενο κάθε είδους (κείμενο, ήχο, εικόνα, βίντεο), όπως σύνολα κειμένων πρωτογενών ή και επισημειωμένων με δομική ή γραμματική επισημείωση, λεξικά, ορολογικούς θησαυρούς, οντολογίες, γλωσσικά μοντέλα κτλ. Τα Εργαλεία Γλωσσικής Τεχνολογίας που προσφέρονται ως υπηρεσίες γλωσσικής επεξεργασίας είναι (ενδεικτικά) εργαλεία αναγνώρισης φωνής, λημματο-

---

1 Βλ. Hinrichs et al (2014).

2 Βλ. Wittenburg et al (2010), Van Uytvanck et al (2010).

ποιητές, εργαλεία μορφοσυντακτικής επισημείωσης, εξαγωγής όρων, περίληψης, ανagnώρισης κυρίων ονομάτων κτλ. Οι πόροι αποθηκεύονται σε ένα δίκτυο αποθετηρίων (με εθνικά υπο-δίκτυα) που μεριμνούν για την έρευνα και την ψηφιακή προσαρμογή και ετοιμότητα των διαφόρων γλωσσών.

Στόχος της υποδομής είναι να συγκεραστούν οι ανάγκες των τελικών χρηστών για εύκολη και αποτελεσματική ανεύρεση πόρων κατάλληλων για την έρευνά τους με τις ανάγκες των παρόχων πόρων για τεκμηρίωση, συντήρηση και νόμιμη διάθεση των πόρων τους με ταυτόχρονο έλεγχο της αδειοδότησης και χρήσης τους. Η υποδομή στοχεύει, μέσω των υπηρεσιών που παρέχει, να βελτιωθούν ο εντοπισμός ενός πόρου (discoverability), η τεκμηρίωση (documentation) και συντήρησή του (curation and preservation), η αναγνωρισιμότητα και ταυτοποίησή του (identification), η διαθεσιμότητα (availability) και η προσβασιμότητά του (accessibility). Έτσι, η υποδομή συνδυάζει δεδομένα και εργαλεία/υπηρεσίες με σκοπό τη διάθεση (κυρίως) ανοικτά προσβάσιμων πόρων σε ευρεία κλίμακα. Μέσω της υποδομής<sup>3</sup>, οι ερευνητές μπορούν με μία πιστοποιημένη εγγραφή στην υποδομή

- να αναζητήσουν, να βρουν και να πάρουν τη νόμιμη έγκριση να χρησιμοποιήσουν πόρους και εργαλεία που βρίσκονται αποθηκευμένα στο αποθετήριο του δικού τους οργανισμού ή στα αποθετήρια άλλων οργανισμών-μελών του δικτύου
- να επιλέξουν, ή να δημιουργήσουν, τα σύνολα δεδομένων που τους ενδιαφέρουν
- να χρησιμοποιήσουν εργαλεία γλωσσικής ανάλυσης που προσφέρονται από άλλα μέλη του δικτύου για να επεξεργαστούν το επιλεγμένο σύνολο δεδομένων
- να χρησιμοποιήσουν την υπολογιστική ισχύ ενός άλλου υπολογιστικού κέντρου-μέλους του δικτύου προκειμένου να εκτελέσουν τις υπηρεσίες επεξεργασίας
- να αποθηκεύσουν τη διαδικασία και τα αποτελέσματα της επεξεργασίας και
- να τα μοιραστούν περαιτέρω με άλλους συνεργάτες.

Σήμερα, το ευρωπαϊκό δίκτυο αριθμεί περισσότερα από 34 μέλη από 18 χώρες. Η Ελλάδα έγινε μέλος του CLARIN ERIC τον Φεβρουάριο του 2015. Η υποδομή **clarin:el** ([www.clarin.gr](http://www.clarin.gr)) είναι το ελληνικό σκέλος της ευρωπαϊκής υποδομής, το οποίο υλοποιείται από το εθνικό δίκτυο Γλωσσικών Πόρων και Τεχνολογιών.

---

3 Βλ. Odiijk (2016) για μελέτες γλωσσικής έρευνας που αξιοποίησαν την υποδομή CLARIN.



### 3. Η ελληνική υποδομή clarin:el

Η υποδομή clarin:el προσφέρει πρόσβαση σε πόρους και υπηρεσίες που διαθέτουν οι πάροχοι-μέλη του δικτύου, και συγκεκριμένα, σε ψηφιακούς πόρους δεδομένων, οι οποίοι μπορεί να περιλαμβάνουν αρχεία διαφόρων τύπων, όπως κείμενο, ήχο, εικόνα και βίντεο, καθώς και σε εργαλεία και υπηρεσίες γλωσσικής τεχνολογίας, όπως ληματοποιητές, επισημειωτές μορφοσυντακτικής πληροφορίας, εργαλεία στατιστικής ανάλυσης κειμένων κτλ., με τη μορφή εκτελέσιμων προγραμμάτων ή/και διαδικτυακών υπηρεσιών.

Η υποδομή αποτελείται από ένα δυναμικό περιβάλλον εργασίας, το οποίο περιλαμβάνει δύο στενά διασυνδεδεμένα υποσυστήματα:

- ένα σύστημα τεκμηρίωσης, αποθήκευσης, διαμοιρασμού, αναζήτησης, ανάκτησης και καταφόρτωσης γλωσσικών πόρων, και
- ένα σύστημα επεξεργασίας γλωσσικών δεδομένων μέσω διαδικτυακών υπηρεσιών γλωσσικής επεξεργασίας και παραγωγής νέων δεδομένων.

Μέσω αυτών των υποσυστημάτων, ο πάροχος πόρων τεκμηριώνει, αποθηκεύει, διαχειρίζεται και διαθέτει τους πόρους του, ενώ ο καταναλωτής πόρων έχει πρόσβαση μέσω του διαδικτύου σε ένα εικονικό αποθετήριο πόρων (το οποίο συγκεντρώνει τους πόρους των δικτυωμένων κέντρων), από το οποίο μπορεί να επιλέξει τον πόρο που τον ενδιαφέρει και είτε να τον καταφορτώσει στον υπολογιστή του είτε να τον επεξεργαστεί με κάποια από τις προσφερόμενες υπηρεσίες γλωσσικής επεξεργασίας και να καταφορτώσει τα αποτελέσματα στον υπολογιστή του.

#### 3.1. Αρχές της υποδομής

Επιγραμματικά οι βασικές λειτουργικές και τεχνικές αρχές της υποδομής είναι οι ακόλουθες:

- Ανοιχτή και κατανεμημένη υποδομή, με βέλτιστη γεωγραφική κατανομή.
- Ανοιχτή πρόσβαση σε πόρους δεδομένων και διαδικτυακές υπηρεσίες, μέσω πιστοποίησης-εξουσιοδότησης των χρηστών και σύμφωνα με τα δικαιώματα χρήσης των πόρων.
- Πόροι και διαδικτυακές υπηρεσίες προσφέρονται, ελέγχονται, ενημερώνονται,

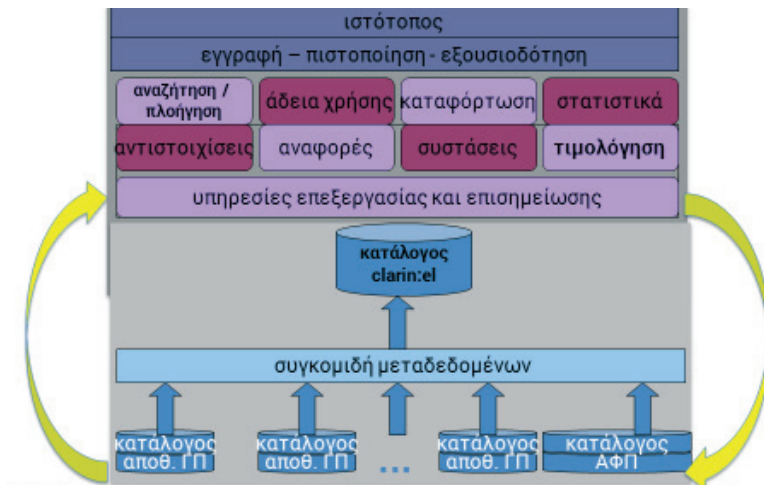
κτλ. από τους παρόχους-μέλη του δικτύου.

- Οι πόροι συνοδεύονται από μεταδεδομένα, τα οποία είναι συμβατά με διεθνή πρότυπα και ανοιχτά σε διαδικασίες συγκομιδής (harvesting).
- Η τεκμηρίωση γίνεται βάσει ενιαίας τυπολογίας πόρων και υπηρεσιών και ενιαίου μοντέλου μεταδεδομένων.
- Όλα τα ζητήματα πνευματικών δικαιωμάτων πρέπει να έχουν εκκαθαριστεί προκειμένου ένας πόρος να συμπεριληφθεί στην υποδομή.
- Ιδανικά, οι πόροι διατίθενται με ανοιχτές άδειες χρήσης. Εφόσον, όμως, υπάρχουν περιορισμοί στον τύπο των χρηστών ή στο είδος της χρήσης, αυτοί είναι ρητοί και γίνονται σεβαστοί.

### 3.2. Αρχιτεκτονική της υποδομής

Η υποδομή clarin:el λειτουργεί ως ένα κατανεμημένο δίκτυο αποθετηρίων και περιλαμβάνει έναν Κεντρικό Συσσωρευτή και Τοπικά Αποθετήρια (εικόνα 1). Ο Κεντρικός Συσσωρευτής είναι το κεντρικό Αποθετήριο της Υποδομής, βασική αρμοδιότητα του οποίου είναι η συγκομιδή των μεταδεδομένων από τα λοιπά Αποθετήρια, η ενοποίηση και προβολή τους σε ενιαίο κατάλογο και, τέλος, η διάθεση των πόρων στους χρήστες.

Τα Τοπικά Αποθετήρια της υποδομής clarin:el είναι δύο ειδών: τα Ιδρυματικά Αποθετήρια και το (ειδικό) Αποθετήριο Φιλοξενούμενων Πόρων.



Εικόνα 1 | Η αρχιτεκτονική της υποδομής clarin:el

Αντιστοιχεί ένα Ιδρυματικό Αποθετήριο σε κάθε φορέα-μέλος του δικτύου, ο οποίος είναι υπεύθυνος για τη λειτουργία του. Τα τοπικά αποθετήρια περιλαμβάνουν τον κατάλογο και τις περιγραφές (μεταδεδομένα τεκμηρίωσης) των πόρων και υπηρεσιών που διαθέτει ο φορέας μέσω της υποδομής· επίσης, αποθηκεύουν τους ίδιους τους πόρους και τα εργαλεία επεξεργασίας/ τις υπηρεσίες του φορέα. Δυνατότητα τεκμηρίωσης και αποθήκευσης πόρων και εργαλείων επεξεργασίας/υπηρεσιών στο Ιδρυματικό Αποθετήριο κάθε φορέα έχουν μόνο τα μέλη του συγκεκριμένου φορέα.

Στην υποδομή, ωστόσο, μπορούν να συμμετάσχουν και φορείς χωρίς αποθετήριο. Στο Αποθετήριο Φιλοξενούμενων Πόρων αποθηκεύονται πόροι και εργαλεία επεξεργασίας/υπηρεσίες που παρέχονται από φορείς-παρόχους που δεν διατηρούν Αποθετήριο, καθώς και τα μεταδεδομένα των πόρων αυτών. Εκτός από τους φορείς, πόρους, τεχνολογίες και υπηρεσίες μπορούν να διαθέσουν μέσω της υποδομής και άτομα (Συνεργαζόμενοι Πάροχοι), οι οποίοι αποθηκεύουν, τεκμηριώνουν και διαχειρίζονται τους πόρους τους στο Αποθετήριο Φιλοξενούμενων Πόρων.

Οι Πάροχοι είναι υπεύθυνοι για την επιμέλεια των Πόρων και των Υπηρεσιών τους, την τεκμηρίωσή τους και την εκκαθάριση των πνευματικών δικαιωμάτων που συνδέονται με αυτούς.

Όλα τα υποσυστήματα είναι εγκατεστημένα στην υπολογιστική υποδομή του Εθνικού Δικτύου Έρευνας και Τεχνολογίας (ΕΔΕΤ, [www.grnet.gr](http://www.grnet.gr)), το οποίο παρέχει στην υποδομή εικονικές μηχανές, αποθηκευτικό χώρο και υπολογιστική ισχύ. Η ρύθμιση αυτή εξασφαλίζει την ασφάλεια και τη συντήρηση των πόρων και των μεταδεδομένων τους, αλλά και την απρόσκοπτη λειτουργία της υποδομής.

### *3.3. Τεκμηρίωση και μεταφόρτωση πόρων*

Οι πάροχοι τεκμηριώνουν τους πόρους που αποθηκεύουν στο αποθετήριό τους μέσω του ειδικού περιβάλλοντος που παρέχει η υποδομή, το οποίο τους καθοδηγεί ώστε να δημιουργούν και να επεξεργάζονται τις περιγραφές των πόρων τους και να μεταφορτώνουν τα δεδομένα τους. Οι πόροι που διατίθενται μέσω της υποδομής πρέπει να τηρούν ορισμένες προδιαγραφές που έχουν στόχο να βελτιώσουν την προσβασιμότητά τους, και συγκεκριμένα, να έχουν τεκμηριωθεί με το μοντέλο τεκμηρίωσης που έχει υιοθετήσει η υποδομή (σχήμα META-SHARE v.3.0.2<sup>4</sup>) και να δηλώνουν με σαφήνεια τους όρους αδειοδότησής τους.

---

4 Το σχήμα σε μορφή XSD διατίθεται στο: <https://inventory.clarin.gr/META-XMLSchema/v3.0.2/>

Οι πόροι, όπως περιγράφεται στο Gavrilidou et al. (2012), ταξινομούνται βάσει δύο κριτηρίων: το μέσο πόρου (media type) και τον τύπο πόρου (resource type). Έτσι, ως προς το μέσο πόρου, διακρίνονται σε κειμενικό, ηχητικό ή πολυμεσικό πόρο και ως προς τον τύπο πόρου, προβλέπονται οι εξής κατηγορίες:

- σώμα δεδομένων (κειμενικά, ηχητικά, πολυμεσικά κτλ. δεδομένα)
- λεξικός/εννοιολογικός πόρος (π.χ. λεξικό, ορολογικός πόρος, κατάλογος λέξεων, οντολογία κτλ.)
- γλωσσική περιγραφή (π.χ. γλωσσικό μοντέλο, υπολογιστική γραμματική κτλ.)
- εργαλείο/τεχνολογία/διαδικτυακή υπηρεσία (π.χ. λημματοποιητής, επισημειωτής, εργαλείο αναγνώρισης κυρίων ονομάτων, εξαγωγής όρων, περίληψης κτλ.).

Οι πόροι δεδομένων και οι τεχνολογίες γλωσσικής επεξεργασίας οφείλουν να πληρούν συγκεκριμένες προδιαγραφές προκειμένου να ενταχθούν στην υποδομή, κυρίως για να εξασφαλίζουν την μεταξύ τους διαλειτουργικότητα. Αυτό σημαίνει ότι η υποδομή μπορεί να φιλοξενήσει οποιονδήποτε τύπο πόρου, ωστόσο, για να μπορεί να εγγυηθεί την επεξεργασιμότητά του, ο πόρος πρέπει να ακολουθεί συγκεκριμένες προδιαγραφές. Με τον τρόπο αυτό, η υποδομή γνωρίζει αν μία διαδικτυακή υπηρεσία (π.χ. ένας μορφοσυντακτικός αναλυτής) μπορεί να επεξεργαστεί ένα σύνολο δεδομένων (π.χ. ένα σώμα ελληνικών νομικών κειμένων). Οι προδιαγραφές για την εξασφάλιση της δυνατότητας επεξεργασίας ενός κειμενικού πόρου αφορούν:

- τη γλώσσα: μόνο Ελληνικά και Αγγλικά κείμενα (μονόγλωσσα ή/και δίγλωσσα παράλληλα) είναι προς το παρόν επεξεργάσιμα. Σύντομα θα προσφέρονται υπηρεσίες και για Γερμανικά, Πορτογαλικά και άλλες γλώσσες. Σε παλαιότερες μορφές της Ελληνικής (π.χ. Μεσαιωνική) ή σε διαλεκτικά κείμενα οι παρεχόμενες τεχνολογίες γλωσσικής επεξεργασίας έχουν μικρότερο βαθμό επιτυχίας, ενώ προς το παρόν δεν επεξεργάζονται κείμενα της αρχαίας ελληνικής.
- το μορφότυπο: μόνο αρχεία τύπου txt ή xces<sup>5</sup>, ή παράλληλα κειμενικά αρχεία τύπου tmx<sup>6</sup> ή mooses<sup>7</sup> είναι επεξεργάσιμα από τα εργαλεία, ενώ αρχεία pdf ή σαρωμένα σε μορφή εικόνας τύπου jpeg/png/bmp, δεν μπορούν να αποτελέσουν αντικείμενο επεξεργασίας.

5 Πρότυπο κειμενικής αναπαράστασης (<http://www.xces.org/>).

6 Translation Memory Exchange (<http://www.gala-global.org/lisa-oscar-standards>).

7 <http://www.statmt.org/moses/>.

- την άδεια διάθεσης: έχει αποφασιστεί να αποτελούν αντικείμενο επεξεργασίας μόνο πόροι με ανοιχτή άδεια χρήσης που επιτρέπει παράγωγα (derivatives).

## 4. Χρήση της υποδομής: η αλληλεπίδραση με τον χρήστη

### 4.1. Βασικές λειτουργίες της υποδομής

Μέσω της υποδομής<sup>8</sup>, ο χρήστης μπορεί να αναζητήσει πόρους, να πλοηγηθεί στον κατάλογο πόρων, να επιλέξει έναν πόρο και να δει την πλήρη περιγραφή του και, εφόσον τον ενδιαφέρει, να τον καταφορτώσει ή να χρησιμοποιήσει διαδικτυακές υπηρεσίες για να τον επεξεργαστεί.

Με τη μορφή διαδικτυακών υπηρεσιών διατίθενται εργαλεία και τεχνολογίες γλωσσικής επεξεργασίας, τα οποία μπορεί ο χρήστης να χρησιμοποιεί για την επεξεργασία του πόρου που επιθυμεί. Με αυτόν τον τρόπο, ο χρήστης αποφεύγει την καταφόρτωση του εργαλείου στον υπολογιστή του, χρησιμοποιεί την υπολογιστική ισχύ και τους χώρους αποθήκευσης που του παρέχει η υποδομή και απλώς μεταφορτώνει τα αποτελέσματα της επεξεργασίας στον υπολογιστή του (ή εφόσον θέλει, χρησιμοποιεί και γι' αυτό τον σκοπό την υποδομή).

Εναλλακτικά, ο χρήστης μπορεί να χρησιμοποιήσει τις υπηρεσίες γλωσσικής επεξεργασίας της υποδομής για να επεξεργαστεί δικούς του πόρους. Στην περίπτωση αυτή, πρέπει να αποθηκεύσει προσωρινά στην υποδομή τον πόρο του ώστε να μπορέσει να τον επεξεργαστεί η υπηρεσία που επιλέγει. Για τα αποτελέσματα της επεξεργασίας (τα οποία αποτελούν νέο πόρο) δίνονται δύο δυνατότητες: να αποθηκευτούν στην υποδομή (και επομένως να είναι διαθέσιμα σε όλους) ή στον υπολογιστή του χρήστη. Εφόσον ο χρήστης επιλέξει την αποθήκευση στην υποδομή, τα μεταδεδομένα τεκμηρίωσης του νέου πόρου δημιουργούνται αυτόματα από την υποδομή, συνδυάζοντας τα μεταδεδομένα του πρωτογενούς πόρου και της υπηρεσίας επεξεργασίας που χρησιμοποιήθηκε.

Στη συνέχεια παρουσιάζονται αναλυτικά οι δυνατότητες που παρέχει η υποδομή στον χρήστη.

<sup>8</sup> Σημειώνεται ότι το περιβάλλον της υποδομής διατίθεται στην αγγλική γλώσσα ώστε να είναι προσβάσιμο και από χρήστες της ευρωπαϊκής υποδομής που δεν γνωρίζουν ελληνικά. Η τεκμηρίωση των πόρων (π.χ. τίτλος, περιγραφή του πόρου) γίνεται εν μέρει και στην ελληνική.

#### *4.2. Φυλλομέτρηση του καταλόγου της υποδομής*

Η φυλλομέτρηση (browse) του κεντρικού καταλόγου επιτρέπει στον χρήστη να δει τους τίτλους και ένα υποσύνολο βασικών χαρακτηριστικών των πόρων που περιλαμβάνει η υποδομή. Ο κεντρικός κατάλογος συγκεντρώνει τους πόρους όλων των αποθετηρίων και ενημερώνεται αυτόματα κάθε φορά που προστίθεται νέος πόρος σε κάποιο από τα αποθετήρια του δικτύου. Όπως φαίνεται στην εικόνα 2 στην επόμενη σελίδα, ο χρήστης μπορεί να φιλτράρει τον κατάλογο με βάση ορισμένα κριτήρια, όπως είναι η γλώσσα (π.χ. για να δει μόνο Ελληνικά κείμενα), ο τύπος πόρου (π.χ. για να δει μόνο λεξικά), το μέσο πόρου (π.χ. για να δει μόνο πολυμεσικούς πόρους), η άδεια χρήσης (για να δει μόνο πόρους πλήρως ανοιχτής πρόσβασης) κ.ο.κ.

#### *4.3. Αναζήτηση πόρων με λέξεις*

Εναλλακτικά, ο χρήστης μπορεί να αναζητήσει πόρους με ελεύθερο κείμενο (free text query) με λέξη ή συνδυασμό λέξεων. Η υποδομή θα ανασύρει και θα του επιστρέψει τον κατάλογο των πόρων που πληρούν τα κριτήρια που έθεσε. Στην εικόνα 3 παρουσιάζεται αναζήτηση με βάση το κείμενο “Ελληνικά νομικά κείμενα”.

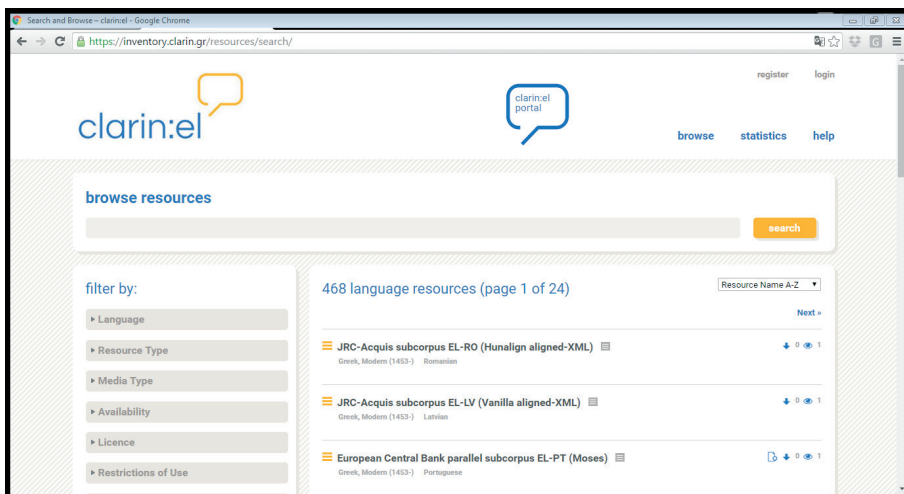
Η υποδομή θα επιστρέψει όλους τους πόρους που είναι τεκμηριωμένοι ως κείμενα που ανήκουν στο νομικό τομέα και είναι (και) στα ελληνικά, δηλαδή, αν υπάρχουν δίγλωσσα κείμενα Ελληνικά-Χ γλώσσα, θα ανασυρθούν και αυτά, δεδομένου ότι το κριτήριο γλώσσας ερμηνεύεται ως «κείμενα που περιλαμβάνουν ελληνικά» χωρίς να διευκρινίζεται αν πρέπει ή όχι να περιλαμβάνεται και άλλη γλώσσα. Εάν ο χρήστης επιθυμεί αποκλειστικά μονόγλωσσα ελληνικά κείμενα, μπορεί να επιλέξει το σχετικό φίλτρο ώστε να περιορίσει την αναζήτηση (εικόνα 4).

#### *4.4. Επιλογή θέασης συγκεκριμένου πόρου*

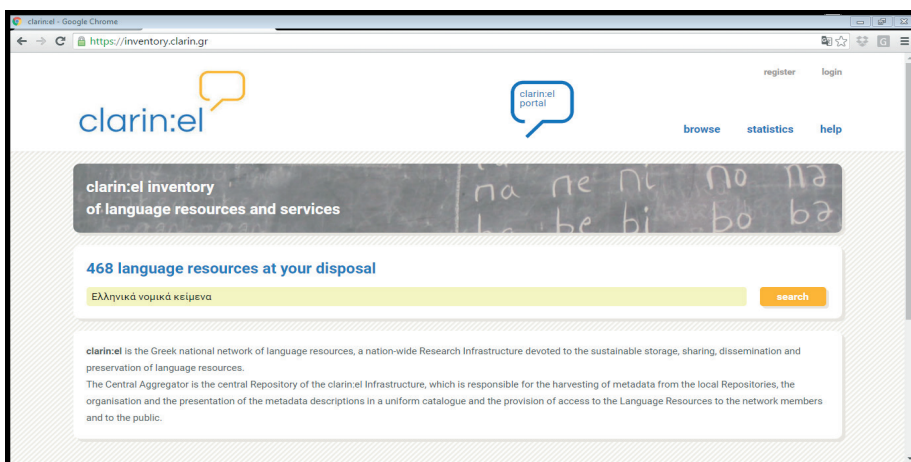
Επιλέγοντας κάποιο πόρο από τον κατάλογο, ο χρήστης μπορεί να δει την πλήρη τεκμηρίωσή του (εικόνα 5).

#### *4.5. Ενημέρωση για τους όρους χρήσης του πόρου*

Πριν προχωρήσει σε οποιαδήποτε επιλογή χρήσης του πόρου (είτε αποφασίσει να τον καταφορτώσει τοπικά, είτε επιλέξει να τον επεξεργαστεί με μία από τις διαθέσιμες



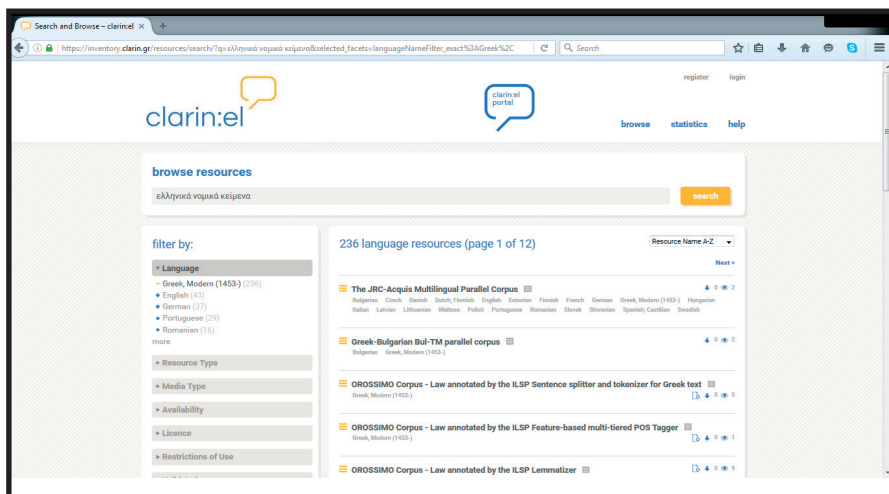
Εικόνα 2 | Φυλλομέτρηση πόρων υποδομής



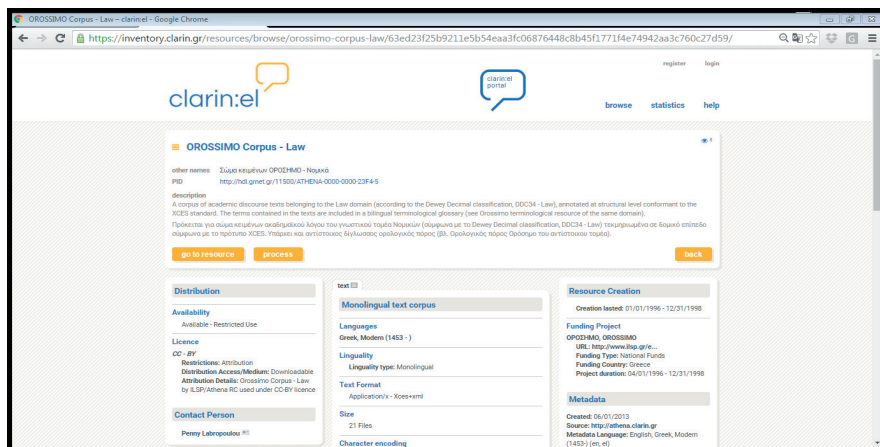
Εικόνα 3 | Αναζήτηση πόρων με ελεύθερο κείμενο

υπηρεσίες) ο χρήστης ενημερώνεται για την άδεια και τους όρους χρήσης του πόρου (εικόνα 6).

Προκειμένου να προχωρήσει στο επόμενο βήμα, ο χρήστης πρέπει να δηλώσει ότι δέχεται τους όρους χρήσης που έχει θέσει ο πάροχος του πόρου, σημειώνοντας ☒ στο σχετικό πεδίο.



Εικόνα 4 | Φιλτραρισμένα αποτελέσματα αναζήτησης



Εικόνα 5 | Θέαση τεκμηρίωσης πόρου

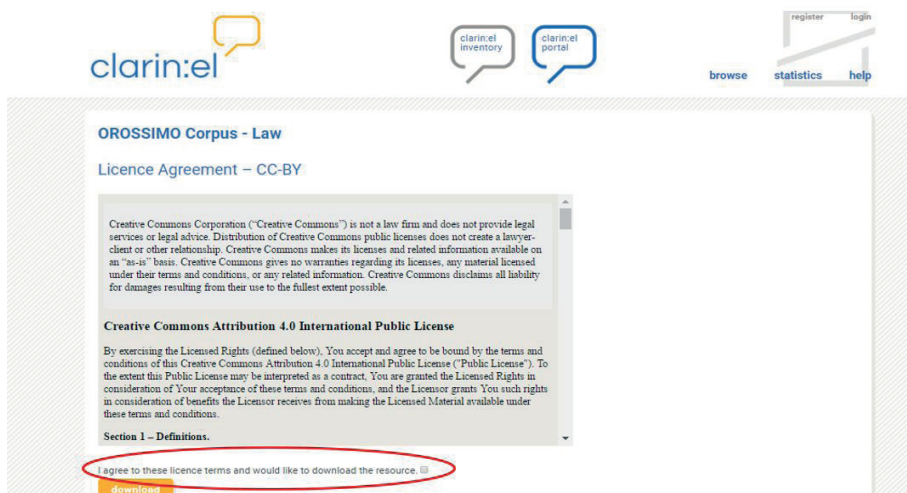
## 4.6 Καταφόρτωση του πόρου

Αφού αποδεχτεί τους όρους χρήσης, ο χρήστης μπορεί να προχωρήσει στο επόμενο βήμα. Στην περίπτωση που απλώς επιθυμεί την καταφόρτωση του πόρου στον



υπολογιστή του, επιλέγει «download» και ο πόρος καταφορτώνεται σε μορφή συμπίεσμένου αρχείου (εικόνα 7).

#### 4.7. Επεξεργασία πόρου

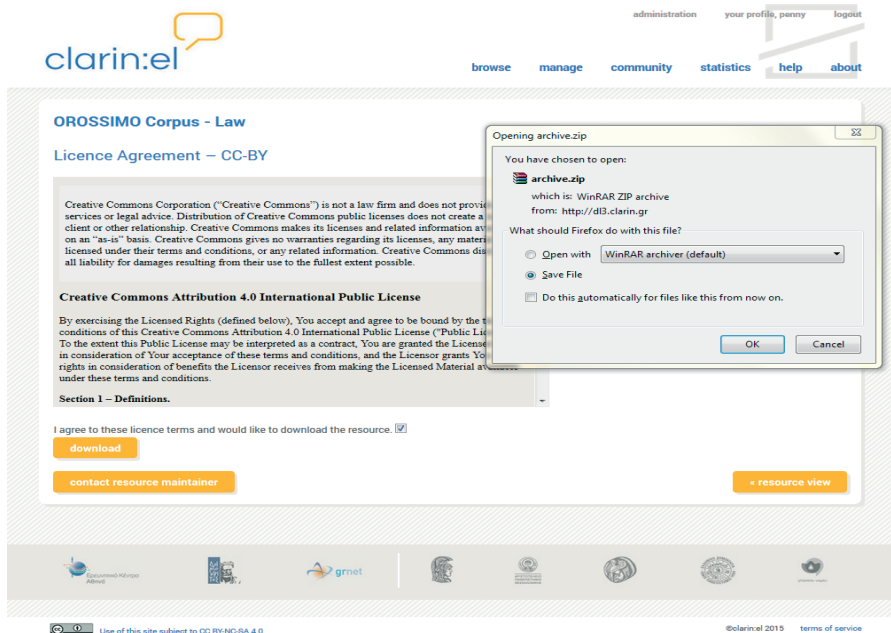


Εικόνα 6 | Άδεια χρήσης πόρου

Ιδιαίτερα σημαντική λειτουργία της υποδομής είναι η παροχή εργαλείων γλωσσικής επεξεργασίας με τη μορφή διαδικτυακών υπηρεσιών. Μέσω αυτής της λειτουργίας, ο χρήστης μπορεί να επιλέξει εκείνη που εξυπηρετεί τον σκοπό της έρευνάς του, να επεξεργαστεί τον πόρο που επιθυμεί και να καταφορτώσει τα αποτελέσματα της επεξεργασίας. Οι υπηρεσίες, όπως και τα δεδομένα, παρέχονται από τα μέλη του δικτύου.

Στην παρούσα φάση (2016) προσφέρονται οι εξής υπηρεσίες επεξεργασίας μονό-γλωσσων σωμάτων κειμένων καθώς και δίγλωσσων σωμάτων για το ζεύγος Ελληνικά-Αγγλικά

- αναγνώριση λέξεων και προτάσεων (tokenisation and sentence splitting)
- μορφοσυντακτική ανάλυση (part of speech tagging)
- λημματοποίηση (lemmatisation)
- συντακτική ανάλυση (syntactic parsing)



Εικόνα 7 | Καταφόρτωση πόρου

- αναγνώριση και εξαγωγή ορολογίας (term recognition and extraction)
- αναγνώριση ονοματικών οντοτήτων (named entity recognition)

Κρίσιμη για τη διαδικασία αυτή είναι η διαλειτουργικότητα δεδομένων και υπηρεσιών<sup>9</sup>: να μπορεί, δηλαδή, να εντοπίσει η υποδομή εάν ένας πόρος επιτρέπει την επεξεργασία του από τεχνικής και νομικής απόψεως, ποιοι τύποι επεξεργασίας διατίθενται για το συγκεκριμένο είδος πόρου και ποια εργαλεία προσφέρονται για τον σκοπό αυτό. Η διαλειτουργικότητα αυτή επιτυγχάνεται μέσω της τεκμηρίωσης που συνοδεύει κάθε πόρο (όπως ήδη αναφέρθηκε στην Ενότητα 3.3).

Συγκεκριμένα, βάσει της τεκμηρίωσης πόρων και υπηρεσιών, η υποδομή μπορεί να εντοπίσει τα επεξεργάσιμα δεδομένα, δηλαδή εκείνα που ικανοποιούν τις προδιαγραφές δεδομένων και υπηρεσιών σε σχέση με τη γλώσσα (ελληνικά, αγγλικά) και το αναμενόμενο μορφότυπο εισόδου, και να τα συνδυάσει με τις υπηρεσίες που αντιστοι-

9 Βλ. Piperidis *et al.* (2015)

χούν στην απαίτηση του χρήστη. Έτσι, αν ο χρήστης επιλέξει να κάνει μορφοσυντακτική επεξεργασία ενός ελληνικού κειμένου, η υποδομή θα επιλέξει από τον κεντρικό κατάλογο μόνο ελληνικά κείμενα και μόνο τις υπηρεσίες μορφοσυντακτικής επεξεργασίας που χειρίζονται ελληνικά και θα τα προτείνει στον χρήστη για να επιλέξει.

Τα επίπεδα γλωσσικής επεξεργασίας και, επομένως, και οι αντίστοιχες υπηρεσίες είναι οργανωμένες με τη μορφή ενός κατευθυνόμενου ακυκλικού γράφου (directed acyclic graph), δηλαδή ενός δένδρου, όπου κάθε κόμβος αντιστοιχεί σε ένα επίπεδο επεξεργασίας. Η εκτέλεση μιας υπηρεσίας επεξεργασίας αντιστοιχεί στη διάσχιση του γράφου με κόμβο εκκίνησης εκείνον που αντιστοιχεί στο επίπεδο επεξεργασίας του αρχικού πόρου και τερματικό κόμβο εκείνον που αντιστοιχεί στο επίπεδο επεξεργασίας που επιθυμεί ο χρήστης. Με αυτόν τον τρόπο, η υποδομή είναι σε θέση να αναγνωρίζει τις αλληλεξαρτήσεις μεταξύ των διαφορετικών επιπέδων επεξεργασίας. Για παράδειγμα, για τη λημματοποίηση ενός κειμένου απαιτείται να έχει προηγηθεί μορφοσυντακτική ανάλυση του κειμένου, έτσι ώστε ο λημματοποιητής να επιλέξει το σωστό λήμμα που αντιστοιχεί στο συγκεκριμένο λεκτικό τύπο στο δεδομένο συγκεκριμένο (π.χ. για τον λεκτικό τύπο “εξουσιοδοτήσεις”, θα επιλέξει “εξουσιοδοτώ”, εάν είναι ρήμα, ή “εξουσιοδότηση”, εάν είναι ουσιαστικό). Επίσης, για τη συντακτική ανάλυση απαιτείται να έχει προηγηθεί μορφοσυντακτική επεξεργασία. Και βέβαια, για τα επίπεδα επεξεργασίας και των δύο παραδειγμάτων απαιτείται να έχει προηγηθεί κατάτμηση του κειμένου σε λεκτικές μονάδες και προτάσεις.

Οι αλληλεξαρτήσεις μεταξύ επιπέδων και υπηρεσιών επεξεργασίας καθορίζονται από τον πάροχό τους και εξασφαλίζουν τη διαλειτουργικότητα μεταξύ των υπηρεσιών. Η ρύθμιση αυτή σημαίνει ότι μόνο υπηρεσίες που ανήκουν στην ίδια οικογένεια ή στο ίδιο πλαίσιο επεξεργασίας (processing framework) μπορούν να διαλειτουργήσουν, ενώ ο συνδυασμός υπηρεσιών από διαφορετικά πλαίσια καθιστά αναγκαία τη δημιουργία εργαλείων μετατροπής που αναλαμβάνουν να μετατρέψουν το μορφότυπο (format) της εξόδου μιας υπηρεσίας σε μορφή κατάλληλη ώστε να αποτελέσει την είσοδο μιας άλλης υπηρεσίας επεξεργασίας.

Σε σχέση με τους νομικούς περιορισμούς και τις αλληλεπιδράσεις, η υποδομή υιοθετεί την απλή προσέγγιση της επεξεργασίας αποκλειστικά πόρων που διατίθενται με ανοιχτή πρόσβαση. Έτσι, αντικείμενο επεξεργασίας μπορούν να αποτελέσουν μόνο πόροι ανοιχτής πρόσβασης που επιτρέπουν παράγωγα έργα και στις προτεινόμενες υπηρεσίες συμμετέχουν μόνο οι ανοικτά διατιθέμενες.

Τα προς επεξεργασία δεδομένα μπορούν να διατίθενται από την υποδομή ή να φορτωθούν στην υποδομή από τον χρήστη. Στη δεύτερη περίπτωση η υποδομή ελέγχει τη

συμβατότητα των πόρων με τις προδιαγραφές, ώστε να διαπιστώσει τη διαλειτουργικότητά τους με τις υπηρεσίες επεξεργασίας.

Τα επεξεργασμένα δεδομένα συνιστούν νέα δεδομένα, που αποθηκεύονται στην υποδομή, αποκτούν αυτομάτως μεταδεδομένα τεκμηρίωσης τα οποία, όπως προαναφέρθηκε, παράγει η υποδομή συνδυάζοντας τα μεταδεδομένα του πόρου και της υπηρεσίας που χρησιμοποιήθηκαν και τέλος, μέσω ειδικών μεταδεδομένων συσχετίζονται τα επεξεργασμένα δεδομένα με τα αρχικά (π.χ. συνδέεται ο μορφοσυντακτικά επισημειωμένος πόρος με τον αρχικό πόρο από τον οποίο προήλθε).

Αν τα πρωτογενή δεδομένα φορτώθηκαν από τον χρήστη δίνεται ένα περιθώριο 48 ωρών για την καταφόρτωσή τους, και στη συνέχεια διαγράφονται. Υπάρχει ωστόσο η δυνατότητα να αποθηκευτούν μόνιμα και να είναι διαθέσιμα σε όλα τα μέλη του δικτύου, εφόσον ο χρήστης είναι ήδη ή επιλέξει να γίνει μέλος του δικτύου.

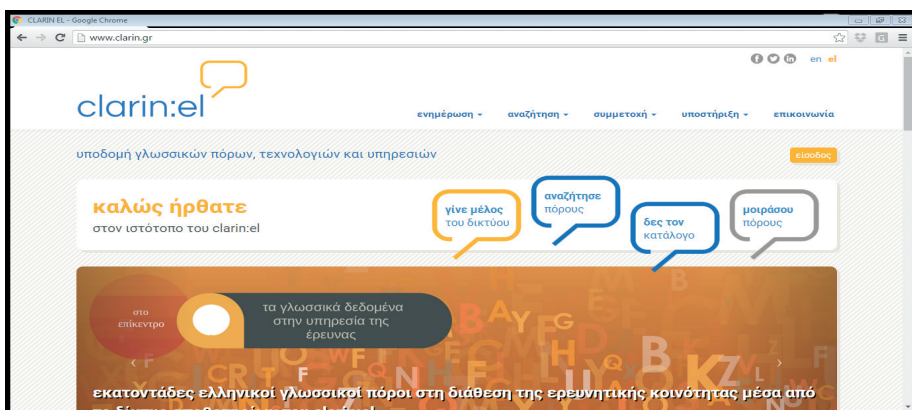
## **5. Ανοιχτή πρόσβαση**

Η πρόσβαση στην υποδομή clarin:el είναι ελεύθερη για όλους τους φορείς ή/και ιδιώτες που ασχολούνται με τη Γλωσσική Τεχνολογία, τους Γλωσσικούς Πόρους και τη μελέτη της γλώσσας, για τους ερευνητές όλων των επιστημών που επιθυμούν να διεξαγάγουν την έρευνά τους σε γλωσσικό υλικό αξιοποιώντας τους διατιθέμενους πόρους ή/και τα εργαλεία γλωσσικής τεχνολογίας, αλλά και για το ευρύ κοινό. Όλα τα μεταδεδομένα τεκμηρίωσης είναι ελεύθερα προσβάσιμα σε όλους με άδεια Creative Commons Attribution 4.0, ενώ οι πόροι και οι υπηρεσίες διατίθενται με άδεια χρήσης που καθορίζεται από τον πάροχο.

## **6. Διαδικτυακή Πύλη**

Η Διαδικτυακή Πύλη ([www.clarin.gr](http://www.clarin.gr)) αποτελεί το κεντρικό σημείο εισόδου του χρήστη (είτε καταναλωτή είτε παρόχου πόρων) στην υποδομή. Μέσω αυτού του σημείου εισόδου ο χρήστης έχει πρόσβαση στον ιστότοπο του clarin:el, ο οποίος είναι σε δύο γλώσσες (ελληνικά και αγγλικά) και περιλαμβάνει

- όλο το πληροφοριακό περιεχόμενο, δηλαδή το ενημερωτικό υλικό, τη βιβλιοθήκη με τις σχετικές δημοσιεύσεις, άρθρα, εγχειρίδια κτλ., σελίδες βοήθειας,



Εικόνα 8 | Η διαδικτυακή πύλη της υποδομής clarin:el

τα νέα, τις εκδηλώσεις, καθώς και στατιστικές πληροφορίες για τη χρήση της υποδομής

- σύνδεση με την υποδομή, δηλαδή
  - με τον Κεντρικό Συσσωρευτή και
  - με όλα τα αποθετήρια της εθνικής υποδομής, και
- σύνδεση με την ευρωπαϊκή υποδομή.

Η εγγραφή, σύνδεση και πιστοποίηση των χρηστών γίνεται μέσω της διαδικτυακής πύλης και του κεντρικού συσσωρευτή. Οι ακαδημαϊκοί χρήστες έχουν πρόσβαση τόσο στον Κεντρικό Συσσωρευτή όσο και στο Ιδρυματικό Αποθετήριο (εφόσον υπάρχει) του οργανισμού στον οποίο ανήκουν. Χρήστες που δεν ανήκουν σε ακαδημαϊκό οργανισμό αλλά θέλουν να χρησιμοποιήσουν την υποδομή χρησιμοποιούν εγγραφή με απλό λογαριασμό και έχουν πρόσβαση μόνο στον Κεντρικό Συσσωρευτή και, εφόσον διαθέσουν πόρους, στο Αποθετήριο Φιλοξενούμενων Πόρων.

## 7. Μέλη του Δικτύου

Το δίκτυο clarin:el είναι το δίκτυο των νομικών προσώπων που συμμετέχουν και συνεργάζονται για τη δημιουργία, λειτουργία και υποστήριξη της υποδομής clarin:el.

Στον πυρήνα του δικτύου βρίσκεται η κοινοπραξία του έργου, δηλαδή οι φορείς που υλοποιούν την υποδομή, οι οποίοι αποτελούν και τα ιδρυτικά μέλη του δικτύου:

το Ερευνητικό Κέντρο «Αθηνά» με το Ινστιτούτο Επεξεργασίας του Λόγου, φορέα συντονιστή, και το Ινστιτούτο Πληροφοριακών Συστημάτων, το Ινστιτούτο Πληροφορικής και Τηλεπικοινωνιών του ΕΚΕΦΕ "ΔΗΜΟΚΡΙΤΟΣ" και η ΕΔΕΤ.

Εκτός από τα ιδρυτικά μέλη, συμμετέχουν ως μέλη με ιδρυματικά αποθετήρια οι εξής ακαδημαϊκοί φορείς: Πανεπιστήμιο Αθηνών, Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης, Πανεπιστήμιο Αιγαίου, Ιόνιο Πανεπιστήμιο και Κέντρο Ελληνικής Γλώσσας.

Όλα τα πιστοποιημένα μέλη των φορέων-μελών του δικτύου (μέλη ΔΕΠ, ερευνητές, φοιτητές) είναι χρήστες της υποδομής με πλήρη δικαιώματα. Η πιστοποίησή τους γίνεται μέσω της ΕΔΕΤ, με την αναγνώριση της ηλεκτρονικής ακαδημαϊκής τους διεύθυνσης. Οι Φορείς-Μέλη

- έχουν πρόσβαση στον κεντρικό κατάλογο πόρων της υποδομής,
- έχουν πρόσβαση στους πόρους της υποδομής,
- έχουν πρόσβαση σε όλες τις υπηρεσίες της υποδομής και χρησιμοποιούν όλες τις λειτουργίες της,
- έχουν πρόσβαση σε εξατομικευμένες υπηρεσίες (στατιστικές πληροφορίες για τους πόρους και τις υπηρεσίες τους),
- παρέχουν στα μέλη τους
  - τα δικαιώματα πρόσβασης στην υποδομή
  - το δικαίωμα συμμετοχής σε δράσεις του δικτύου (συνέδρια, σεμινάρια, επιμορφωτικές ημερίδες και εκπαιδευτικά προγράμματα κτλ.).
- έχουν τη δυνατότητα να συνεργάζονται με άλλα μέλη του εθνικού, αλλά και του ευρωπαϊκού δικτύου (μέσω της διασύνδεσης με την ευρωπαϊκή Υποδομή) για τη δημιουργία, διάθεση και επεξεργασία Γλωσσικών πόρων, Τεχνολογιών και Υπηρεσιών.

Όταν κάποιος φορέας ή άτομο γίνεται μέλος του clarin:el, γίνεται μέλος της κοινότητας διαμοιρασμού γλωσσικών πόρων. Έτσι, μοιράζεται τους δικούς του πόρους και έχει πρόσβαση σε πόρους άλλων, σε εθνικό και ευρωπαϊκό επίπεδο. Οι πόροι και οι υπηρεσίες μπορούν να αξιοποιηθούν τόσο στην έρευνα όσο και στην εκπαιδευτική πράξη.

Το δίκτυο είναι ανοιχτό σε όσους φορείς και ιδιώτες ενδιαφέρονται να διαθέσουν πόρους ή/και υπηρεσίες μέσω της υποδομής. Οι διαδικασίες εγγραφής βρίσκονται στη διαδικτυακή πύλη (<http://www.clarin.gr/el/content/join-clarin-el-network>).

## 8. Συνοδευτικές δράσεις

Για να επιτύχει τους στόχους του, το Δίκτυο clarin:el έχει αναλάβει ένα ευρύ φάσμα δράσεων, και συγκεκριμένα:

- τη λειτουργία της εθνικής Υποδομής,
- την υιοθέτηση και χρήση διεθνών προτύπων και βέλτιστων πρακτικών τεκμηρίωσης και αναπαράστασης γλωσσικών πόρων,
- την υποστήριξη δράσεων για τη διάθεση υπαρχόντων πόρων μέσω της Υποδομής,
- τη διοργάνωση εκπαιδευτικών δράσεων καθώς και δράσεων διάχυσης και ενημέρωσης σε θέματα γλωσσικών πόρων και αξιοποίησης της γλωσσικής τεχνολογίας,
- την προώθηση νομικού πλαισίου διάθεσης γλωσσικών πόρων, και παράλληλα πλαισίου προστασίας πνευματικών και ηθικών δικαιωμάτων παρόχων και καταναλωτών γλωσσικών πόρων,
- την προώθηση δράσεων ανοικτών δεδομένων, περιεχομένου, έρευνας και λογισμικού στον βαθμό που είναι συμβατές προς τους στόχους του δικτύου,
- τη μεταφορά τεχνογνωσίας και την προώθηση συνεργασίας μεταξύ των μελών του εθνικού δικτύου και της ευρωπαϊκής υποδομής.

## Βιβλιογραφία

Gavrilidou, Maria, Penny Labropoulou, Elina Desipri, Stelios Piperidis, Haris Papageorgiou, Monica Monachini, Francesca Frontini, Thierry Declerck, Gil Francopoulo, Victoria Arranz, and Valerie Mapelli. 2012. "The META-SHARE Metadata Schema for the Description of Language Resources." In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC 12)*. European Language Resources Association (ELRA). [http://www.lrec-conf.org/proceedings/lrec2012/pdf/998\\_Paper.pdf](http://www.lrec-conf.org/proceedings/lrec2012/pdf/998_Paper.pdf)

- Hinrichs, Erhard, and Steven Krauwer. 2014. "The CLARIN Research Infrastructure: Resources and Tools for eHumanities Scholars." In *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC 14)*. European Language Resources Association (ELRA). [http://www.lrec-conf.org/proceedings/lrec2014/pdf/415\\_Paper.pdf](http://www.lrec-conf.org/proceedings/lrec2014/pdf/415_Paper.pdf)
- Odijk, Jan. ed. 2016. "Linguistic Research in the CLARIN Infrastructure." *Lingua* 178:1–126.
- Piperidis, Stelios, Dimitris Galanis, Juli Bakagianni, and Socrates Sofianopoulos. 2015. "Combining and Extending Data Infrastructures with Linguistic Annotation Services". In *Worldwide Language Service Infrastructure*, edited by Yohei Murakami, and Donghui Lin. Second International Workshop, WLSI 2015, Kyoto, Japan.
- Van Uytvanck, Dieter, Claus Zinn, Daan Broeder, Peter Wittenburg, and Mariano Gardellini. 2010. "Virtual Language Observatory: The Portal to the Language Resources and Technology Universe." In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC 10)*. European Language Resources Association (ELRA). [http://www.lrec-conf.org/proceedings/lrec2010/pdf/273\\_Paper.pdf](http://www.lrec-conf.org/proceedings/lrec2010/pdf/273_Paper.pdf)
- Wittenburg, Peter, Nuria Bel, Lars Borin, Gerhard Budin, Nicoletta Calzolari, Eva Hajicova, Kimmo Koskenniemi, Lothar Lemnitzer, Bente Maegaard, Maciej Piasecki, Jean-Marie Pierrel, Stelios Piperidis, Inguna Skadina, Dan Tufis, Remco Van Veenendaal, Tamas Váradi, and Martin Wynne. 2010. "Resource and Service Centres as the Backbone for a Sustainable Service Infrastructure." In *Proceedings of the Seventh International Conference on Language Resources and Evaluation (LREC 10)*. European Language Resources Association (ELRA). [http://www.lrec-conf.org/proceedings/lrec2010/pdf/679\\_Paper.pdf](http://www.lrec-conf.org/proceedings/lrec2010/pdf/679_Paper.pdf)



