

Η Ερευνητική Υποδομή CLARIN

Στέλιος Πιπερίδης, Πένυ Λαμπροπούλου, Μαρία Γαβριλίδου

ΠΕΡΙΛΗΨΗ

Η Ερευνητική Υποδομή CLARIN είναι μια πανευρωπαϊκή προσπάθεια μεγάλης εμβέλειας με σκοπό τη δημιουργία, τον συντονισμό και τη διάθεση γλωσσικών πόρων και γλωσσικής τεχνολογίας στους ερευνητές των Κοινωνικών και Ανθρωπιστικών Επιστημών. Στοχεύει να παράσχει στους ερευνητές ένα σύνολο ισχυρών εργαλείων ενσωματωμένο στον υπολογιστή τους που θα επιτρέπουν την άσκηση της ερευνητικής εργασίας με καινοτόμους τρόπους βασισμένους στην επεξεργασία γλώσσας. Το άρθρο παρουσιάζει τους στόχους, τις βασικές αρχές και τα χαρακτηριστικά της υποδομής, καθώς και τα μέχρι τώρα επιτεύγματα σε ευρωπαϊκό και ελληνικό επίπεδο και τις μελλοντικές ενέργειες.

The CLARIN Research Infrastructure

Stelios Piperidis, Penny Labropoulou, Maria Gavrilidou

ABSTRACT

The CLARIN Research Infrastructure constitutes a pan-European large-scale initiative aiming to create, coordinate and disseminate Language Resources and Technologies to the research communities of Social Sciences and Humanities. It intends to provide researchers with a set of powerful tools integrated in their computer working environment, enabling them to conduct their research in innovative ways based on language processing technologies. The paper presents the objectives, principles and main features of the infrastructure, as well as the steps that have been taken until now on the European and Greek level towards the implementation of the infrastructure, and the forecoming actions.

0 Εισαγωγή

Το **CLARIN** (Common LAnguage Resources and Technology INfrastructure, www.clarin.eu) είναι ένα έργο Ερευνητικών Υποδομών (ΕΥ) που εμπίπτει στον θεματικό τομέα των Κοινωνικών και Ανθρωπιστικών Επιστημών (ΚΑΕ) του Οδικού Χάρτη ESFRI (European Strategy Forum on Research Infrastructures). Πρόκειται για πανευρωπαϊκό συνεργατικό έργο μεγάλης εμβέλειας με σκοπό την κατασκευή μιας **υποδομής που θα ενισχύσει και θα προωθήσει την έρευνα στις ΚΑΕ αξιοποιώντας τη Γλωσσική Τεχνολογία (ΓΤ) και τους Γλωσσικούς Πόρους (ΓΠ)**.

Πιο συγκεκριμένα, η υποδομή CLARIN θα προσφέρει στους ερευνητές τη δυνατότητα πρόσβασης μέσω *μίας και μοναδικής εγγραφής (single sign-on)* σε μια πλατφόρμα που ενσωματώνει γλωσσικούς πόρους και τεχνολογίες αιχμής σε ευρωπαϊκό επίπεδο. Η δυνατότητα αυτή θα διατεθεί μέσω της κατασκευής και λειτουργίας μιας κοινής καταμεμημένης υποδομής που θα ενσωματώσει *ψηφιακό περιεχόμενο* κάθε είδους (κείμενο, ήχο, εικόνα, βίντεο) και τους *σχετικούς Γλωσσικούς Πόρους* (επισημειωμένες ηχογραφήσεις, μαγνητοσκοπήσεις ή κείμενα, λεξικά, οντολογίες κτλ.) και *εργαλεία Γλωσσικής Τεχνολογίας* (εργαλεία αναγνώρισης φωνής, λημματοποιητές, συντακτικούς αναλυτές, εργαλεία εξαγωγής περίληψης, εργαλεία εξαγωγής πληροφορίας κτλ.) σε μια *ομοσπονδία αποθετηρίων*, η οποία θα είναι διαθέσιμη μέσω *διαδικτυακών υπηρεσιών* σε ερευνητές όλων των επιστημών.

Το **CLARIN-EL** (www.clarin.gr) αποτελεί το ελληνικό τμήμα της ευρωπαϊκής ΕΥ. Ως εκ τούτου, σκοπεύει να υλοποιήσει τους στόχους και τα οράματα του CLARIN, εστιάζοντας στην ελληνική γλώσσα και στις ανάγκες της ελληνικής ερευνητικής κοινότητας. Το CLARIN-EL υιοθετεί τις οδηγίες και τις προδιαγραφές της αντίστοιχης ευρωπαϊκής ΕΥ, προσαρμόζοντάς τις στις ελληνικές ιδιαιτερότητες, αλλά και προωθεί διεθνώς τις εθνικές προτεραιότητες και επηρεάζει το ευρωπαϊκό CLARIN προς όφελος του ελληνικού τμήματος.

1. Σκοπός και όραμα της υποδομής CLARIN

Η ΕΥ CLARIN σκοπεύει να παράσχει στους ερευνητές ένα σύνολο ισχυρών εργαλείων ενσωματωμένο στον υπολογιστή τους, καθώς η εποχή της έρευνας με χαρτί και μολύβι πλησιάζει στο τέλος της.

Η διαρκώς αυξανόμενη διαθεσιμότητα μεγάλων όγκων ψηφιακού περιεχομένου σε διάφορες μορφές καθιστά τις παραδοσιακές μεθόδους ξεπερασμένες και ανεπαρκείς. Η επεξεργασία της γλώσσας μέσω υπολογιστή προσφέρει νέους τρόπους πρόσβασης και ανάλυσης του περιεχομένου. Έχει μάλιστα ήδη δοκιμαστεί με ιδιαίτερη επιτυχία σε διάφορους τομείς των ΚΑΕ, στους οποίους η γλώσσα εμφανίζεται με έναν ή περισσότερους από τους ρόλους που επιτελεί, ως φορέας πολιτισμικού περιεχομένου και γνώσης, εργαλείο επικοινωνίας, συστατικό ταυτότητας και αντικείμενο μελέτης. Οι ΓΠ (προφορικοί, γραπτοί και πολυτροπικοί) είναι κομβικής σημασίας για την έρευνα και τον πολιτισμό, είτε χρησιμοποιούνται ως υλικό για γλωσσική έρευνα είτε αποτελούν φορείς πληροφορίας ή γνώσης για άλλα επιστημονικά αντικείμενα.

Πιο συγκεκριμένα, για τους αρχαιολόγους, ψυχολόγους, ιστορικούς, κοινωνιολόγους κτλ. μπορεί η γλώσσα να μην αποτελεί αντικείμενο μελέτης, ωστόσο, τα κείμενα (είτε ως πρωτογενείς είτε ως δευτερογενείς πηγές) αποτελούν βασικό υλικό της επιστήμης τους. Οι

επιστήμες αυτές μελετούν τον άνθρωπο, ερευνώντας στοιχεία του πολιτισμού που περιγράφουν τι θεωρείται σημαντικό σε κάθε ιστορική στιγμή. Οι αποδείξεις γι' αυτό είναι αναπόφευκτα συνδεδεμένες με τη γλώσσα. Πιο συγκεκριμένα, η έρευνα συχνά αφορά τη μελέτη πρωτογενών πηγών λογοτεχνικού, ιστορικού ή γλωσσικού ενδιαφέροντος, οι οποίες μπορεί να έχουν τη μορφή εντύπων, χειρογράφων, επιγραφών, παπύρων σε διάφορες γλώσσες. Μεταξύ των πρωτογενών πηγών συμπεριλαμβάνονται και άλλες που δεν είναι κειμενικές, όπως αντικείμενα τέχνης, μουσική κτλ., παρόλο που η περιγραφή τους γίνεται και πάλι με τη χρήση κειμένων. Οι πρωτογενείς πηγές, που συνήθως βρίσκονται σε βιβλιοθήκες ή/και αρχαιακούς και πολιτιστικούς οργανισμούς, επισημειώνονται και ερμηνεύονται με τη χρήση γλώσσας. Τον τελευταίο καιρό, έργα ψηφιοποίησης και τεκμηρίωσης των πρωτογενών πηγών έχουν δημιουργήσει έναν μεγάλο όγκο ψηφιακού περιεχομένου στο οποίο έχουν πρόσβαση όλοι οι επιστήμονες. Ο συνεχώς αυξανόμενος όγκος των παραγόμενων κειμένων καθιστά την ψηφιακή οργάνωσή και την επεξεργασία τους μέσω υπολογιστικών μεθόδων επιτακτική ανάγκη.

Με τη συγκρότηση και ανάπτυξη των ψηφιακών βιβλιοθηκών, οι δυνατότητες πρόσβασης και αναζήτησης σε σώματα κειμένων και πολυμεσικές πηγές έχουν αυξηθεί σημαντικά. Ο χώρος των ΓΠΤ τα τελευταία χρόνια παρουσιάζει ιδιαίτερη άνθηση: σε πολλές χώρες αλλά και συνεργατικά σε ευρωπαϊκό επίπεδο, έχουν γίνει και συνεχίζουν να γίνονται προσπάθειες ανάπτυξης, αφενός, ψηφιακών συλλογών δεδομένων βασισμένων στη γλώσσα (με τη μορφή κειμενικών, ηχητικών, οπτικών ή πολυμεσικών και πολυτροπικών δεδομένων) και, αφετέρου, τεχνολογιών και εργαλείων που εξερευνούν, ανακτούν, επεξεργάζονται, αξιοποιούν, μελετούν ή ενισχύουν τα δεδομένα αυτά. Αποτέλεσμα αυτών των προσπαθειών είναι σήμερα να διατίθεται μεγάλος όγκος πόρων και εργαλείων, που ωστόσο παραμένουν γεωγραφικά και τεχνικά απομονωμένοι, με αποτέλεσμα να μην μπορούν εύκολα να εντοπιστούν από τους επιστήμονες των ΚΑΕ. Καθώς, μάλιστα, η χρήση τους απαιτεί σημαντικό βαθμό εξοικείωσης με την τεχνολογία από την πλευρά του χρήστη, κάτι που οι περισσότεροι επιστήμονες των ΚΑΕ δεν διαθέτουν, η αξιοποίησή τους γίνεται ακόμα δυσχερέστερη.

Ως αποτέλεσμα, σήμερα υφίσταται ένα κενό μεταξύ της έρευνας που διεξάγεται στο πλαίσιο των ανθρωπιστικών επιστημών και των δυνατοτήτων που προσφέρουν οι ΓΤ. Η υπόθεση, λοιπόν, είναι ότι είναι δυνατό να ξεπεραστούν αυτά τα προβλήματα με την συγκρότηση ενός εύχρηστου δικτύου που θα περιλαμβάνει τους διαθέσιμους πόρους και τις υπάρχουσες ΓΤ κατά τρόπο φιλικό προς τους χρήστες και που θα συμβάλει πραγματικά στην έρευνα στις ΚΑΕ. Στόχος του CLARIN, και σε ό,τι αφορά την ελληνική γλώσσα του CLARIN-EL, είναι να παράσχει αυτή την υποδομή.

Η υλοποίηση της ΕΥ προβλέπει τη δημιουργία μιας ομοσπονδίας αρχειακών συλλογών και αποθετηρίων, με στόχο να συγκεντρωθεί η κρίσιμη μάζα πόρων, η πρόσβαση στους οποίους θα γίνεται με τη χρήση κατάλληλων εργαλείων. Με την ΕΥ CLARIN θα γίνει εφικτή η διάθεση των ΓΠΤ μέσω διαδικτυακών υπηρεσιών που χαρακτηρίζονται από διαλειτουργικότητα, προσβασιμότητα, σταθερότητα, επεκτασιμότητα και φιλικότητα προς τον χρήστη. Ο απώτερος στόχος είναι να κατασκευαστούν ροές εργασιών ΓΤ (workflows) για συγκεκριμένες εργασίες, οι οποίες θα προσφέρονται έτοιμες στους χρήστες. Εν τω μεταξύ, οι διαθέσιμοι ΓΠΤ θα συγκεντρωθούν σε έναν εικονικό χώρο, από τον οποίο οι χρήστες θα μπορούν να έχουν πρόσβαση σε αυτά, αλλά και σε πληροφορίες για τον τρόπο λειτουργίας τους: οι ΓΠΤ θα προτυποποιηθούν και με τον τρόπο αυτό θα επιλυθούν προβλήματα διαλειτουργικότητας, έτσι ώστε η έξοδος του ενός εργαλείου να αποτελεί την είσοδο σε άλλο, σε μια αλυσίδα διαδικασιών που θα οδηγήσουν στο επιθυμητό από τον χρήστη αποτέλεσμα. Μέσω της ΕΥ θα προσφέρονται επίσης συμβουλευτικές και εκπαιδευτικές υπηρεσίες σε όσους χρήστες το χρειάζονται.

Το όραμα του CLARIN¹ μπορεί να συνοψιστεί ως εξής:

Ένας ερευνητής από το γραφείο του στην Κέρκυρα θα μπορεί:

- με μία εγγραφή (single sign-on) με πιστοποίηση (authentication)
- να ψάξει, να βρει και να πάρει την έγκριση για να χρησιμοποιήσει κείμενα
- από την Οξφόρδη, το Μπέργκεν και το Λέιντεν
- να επιλέξει το ακριβές σύνολο δεδομένων στα οποία θέλει να δουλέψει και να αποθηκεύσει την επιλογή του
- να τρέξει πάνω στην επιλογή του εργαλεία σημασιολογικής ανάλυσης από την Αθήνα και
- στατιστικά εργαλεία από τη Βουδαπέστη
- να χρησιμοποιήσει την υπολογιστική ισχύ ενός άλλου υπολογιστικού κέντρου, όπου και όποτε απαιτείται,
- να αποθηκεύσει τη διαδικασία και τα αποτελέσματα της ανάλυσης και
- να τα μοιραστεί με συνεργάτες του στο Παρίσι, στη Βιέννη και στο Ελσίνκι.

¹ Από παρουσίαση του Martin Wynne, ECDL, Κέρκυρα, 2009

2 Βασικές αρχές της ΕΥ

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

- Το προτεινόμενο δίκτυο θα είναι ανοικτό και κατανεμημένο, με τη βέλτιστη γεωγραφική κατανομή.
- Πόροι, εργαλεία και διαδικτυακές υπηρεσίες θα βρίσκονται στην υπηρεσία των χρηστών κατόπιν εγγραφής τους στο δίκτυο.
- Η πρόσβαση θα γίνεται μέσω πιστοποίησης-εξουσιοδότησης και σύμφωνα με τα δικαιώματα χρήσης.
- Οι ΓΠΤ θα βρίσκονται μόνιμα αποθηκευμένοι σε τοπικούς εξυπηρετητές.
- Πόροι, εργαλεία και διαδικτυακές υπηρεσίες θα παραμένουν, προσφέρονται, ελέγχονται, βελτιώνονται, κτλ. από τους κόμβους-μέλη του δικτύου μέσω εξειδικευμένων αποθετηρίων.
- Θα υπάρχουν, κατόπιν διμερών συμφωνιών μεταξύ του εκάστοτε αποθετηρίου/κόμβου και του κεντρικού συσσωρευτή, εφεδρικά αντίγραφα των πόρων και εργαλείων στον κεντρικό συσσωρευτή.
- Οι ΓΠΤ θα συνοδεύονται από μεταδεδομένα, τα οποία είναι συμβατά με διεθνή πρότυπα και ανοιχτά σε διαδικασίες συγκομιδής (harvesting).
- Η αναζήτηση πόρων, εργαλείων και υπηρεσιών θα γίνεται μέσω των περιγραφών τους με τη χρήση των μεταδεδομένων.
- Ανοιχτά πρότυπα πρέπει να χρησιμοποιούνται κατά το δυνατόν για μεταδεδομένα, δεδομένα και λογισμικό.
- Όλα τα ζητήματα πνευματικών δικαιωμάτων πρέπει να έχουν ξεκαθαριστεί προκειμένου ένας ΓΠ να συμπεριληφθεί στην ΕΥ.
- Ιδανικά, οι ΓΠΤ θα διατίθενται με ανοιχτές άδειες χρήσης. Εφόσον, όμως, υπάρχουν περιορισμοί στον τύπο των χρηστών ή στο είδος της χρήσης, αυτοί πρέπει να είναι σαφείς και σεβαστοί.
- ΓΠΤ που έχουν παραχθεί με δημόσια χρηματοδότηση πρέπει να είναι ανοιχτοί ή να προσφέρονται για διαμοιρασμό.
- Προκειμένου να είναι ανοιχτοί οι πόροι, πρέπει να είναι νομικά και τεχνικά ανοιχτοί κατά την ένταξή τους στο κεντρικό αποθετήριο της ΕΥ αλλά και μελλοντικά.

3 Επιτεύγματα και δράσεις του CLARIN

Το χρονοδιάγραμμα του CLARIN προέβλεπε μια προπαρασκευαστική φάση (2008 – 11), με κύριο αντικείμενο την προετοιμασία της υποδομής ως προς τις βασικές διαστάσεις. Στη

διάρκεια της φάσης αυτής (η οποία βαίνει προς την ολοκλήρωση), το έργο έχει επιτύχει τους στόχους του, και συγκεκριμένα:

- **Χρηματοδότηση και οργάνωση:** Το έργο CLARIN έχει συγκεντρώσει τις χρηματοδοτικές αρχές και έχει συντάξει ένα αρχικό συμφωνητικό μεταξύ των χρηματοδοτικών αρχών στις συμμετέχουσες χώρες, το οποίο καθορίζει τους όρους διαχείρισης, χρηματοδότησης, κατασκευής και λειτουργίας της υποδομής.
- **Τεχνικές προδιαγραφές:** Έχουν καθοριστεί οι τεχνικές προδιαγραφές της υποδομής και έχουν συμφωνηθεί τα βασικά πρότυπα για τη διαλειτουργικότητα τεχνολογιών και δεδομένων.
- **Χρήστες:** Με σκοπό την πλήρη αξιοποίηση του δυναμικού των ΓΠΤ για την κοινότητα των ΚΑΕ: (α) έχει ολοκληρωθεί μελέτη των υφιστάμενων πρακτικών χρήσης της ΓΤ από τους ερευνητές των ΚΑΕ με σκοπό τον καθορισμό των αναγκών, (β) έχει υλοποιηθεί μέρος της υποδομής με τη μορφή πρωτοτύπου και έχει ξεκινήσει ένας μικρός αριθμός τυπικών έργων των ΚΑΕ με σκοπό την αξιολόγησή του, (γ) έχουν γίνει προωθητικές δραστηριότητες με σκοπό την ενημέρωση των κοινοτήτων των ΚΑΕ ως προς τις δυνατότητες που τους προσφέρει η ΓΤ για τις ερευνητικές τους εργασίες, και (δ) έχουν έρθει σε επικοινωνία οι κοινότητες των ΚΑΕ και της ΓΤ μέσω εργαστηρίων και ειδικών συναντήσεων ώστε να εξασφαλιστούν μακροχρόνιες συνεργασίες.

Ακολουθούν η Κατασκευαστική Φάση, με αντικείμενο την υλοποίηση της υποδομής, και η Λειτουργική Φάση, κατά την οποία η υποδομή θα είναι πλήρως λειτουργική και θα προσφέρει τις υπηρεσίες της στην ερευνητική κοινότητα.

4. Οι δράσεις CLARIN στον ελληνικό χώρο

Στην Ελλάδα, η ΓΓΕΤ έχει εγκρίνει το έργο CLARIN-EL PREP, με σκοπό να καταδείξει τη χρησιμότητα της ΕΥ CLARIN για την ελληνική ερευνητική κοινότητα. Στο πλαίσιο αυτό, έχει ολοκληρωθεί και κατατεθεί **μελέτη σκοπιμότητας**, η οποία βασίστηκε στα ευρήματα εστιασμένων ερευνών, στα συμπεράσματα των **ενημερωτικών δράσεων** που αναλήφθηκαν στη διάρκεια του έργου και σε προηγούμενες μελέτες.

Πιο συγκεκριμένα, η κοινοπραξία του έργου² διενήργησε τις εξής επιμέρους εργασίες:

- καταγραφή του ανθρώπινου δυναμικού των δύο εμπλεκόμενων ερευνητικών κοινοτήτων (των ΚΑΕ και της ΓΤ), μέσω ανοιχτής διαδικτυακής έρευνας και από προηγούμενες στατιστικές μελέτες για το ερευνητικό δυναμικό της χώρας,
- έρευνα των αναγκών και των απαιτήσεων των δυνητικών χρηστών της ΕΥ σε σχέση με τους ΓΠΤ, όπου καταγράφονται οι μέχρι τώρα πρακτικές αναζήτησης και αξιοποίησης του γλωσσικού υλικού από τους Έλληνες ερευνητές των ΚΑΕ και εν γένει της ερευνητικής διαδικασίας που ακολουθούν σε σχέση με το πρωτογενές υλικό (σε ψηφιακή ή μη μορφή), τη χρήση ΓΤ και πόρων αναφοράς,
- επισκόπηση του χώρου των ΓΠΤ στην Ελλάδα, με στόχο να διαφανεί αν η υποδομή είναι σε θέση να καλύψει τις ανάγκες των χρηστών άμεσα ή μεσο/μακροπρόθεσμα
- έρευνα για τους πόρους (αναφορικά με τον όγκο και τη μορφή), και ειδικότερα ψηφιακές αρχειακές συλλογές πολιτιστικών και επιστημονικών φορέων, οι οποίες και αποτελούν το κατεξοχήν υλικό μελέτης των ερευνητών, ώστε να διαπιστωθεί η δυνατότητα (και ο βαθμός ευκολίας) σύνδεσης αυτών των πόρων στην ΕΥ
- μελέτη για τις λειτουργικές και τεχνολογικές απαιτήσεις της ΕΥ, και, τέλος,
- κατάρτιση προτάσεων σε θέματα οργάνωσης και διαχείρισης της ΕΥ.

Είναι σημαντικό ότι οι πληροφορίες που εντάχθηκαν στη μελέτη σκοπιμότητας προήλθαν μετά από συζήτηση και συνεργασία με εκπροσώπους των δύο ερευνητικών κοινοτήτων. Για τον σκοπό αυτό, χρησιμοποιήθηκε τόσο το ίδιο το διαδίκτυο για τη διεξαγωγή ερευνών βάσει ερωτηματολογίων αλλά και ως πηγή πληροφορίας των τρόπων που ασκούνται επιμέρους ερευνητικές πρακτικές διαδικτυακά, καθώς και η προσωπική επαφή με επιστήμονες των δύο κοινοτήτων στο πλαίσιο ενημερωτικών δράσεων, (της ημερίδας παρουσίασης του έργου, εστιασμένων συνεντεύξεων με ερευνητές συγκεκριμένων ειδικοτήτων, και συζητήσεων με ερευνητές διαφόρων ειδικοτήτων στο πλαίσιο συναντήσεων εργασίας του δικτύου).

Σημαντικό στοιχείο για την κατάδειξη της δυνατότητας υλοποίησης της ΕΥ υπήρξε η χαρτογράφηση των ΓΠΤ. Από την έρευνα που διεξήχθη – τα πορίσματα της οποίας μπορεί κανείς να δει και στον ιστότοπο του έργου (<http://www.clarin.gr/clarinmaps>) – συμπεραίνει

2 Στην κοινοπραξία του έργου συμμετείχαν οι εξής εταίροι: Ε.Κ. "Αθηνά" (Ινστιτούτο Επεξεργασίας του Λόγου (ΙΕΛ) & Ινστιτούτο Πληροφοριακών Συστημάτων και Προσομοίωσης (ΙΠΣΥΠ)), ΕΚΕΦΕ "ΔΗΜΟΚΡΙΤΟΣ" / Ινστιτούτο Πληροφορικής και Επικοινωνιών, Εθνικό Δίκτυο Έρευνας και Τεχνολογίας (ΕΔΕΤ) Α.Ε., Ιόνιο Πανεπιστήμιο / Εργαστήριο Ψηφιακών Βιβλιοθηκών και Ηλεκτρονικής Δημοσίευσης, Εθνικό Ίδρυμα Ερευνών / Εθνικό Κέντρο Τεκμηρίωσης, Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών - Τμήμα Φιλολογίας / Τομέας Γλωσσολογίας, Κέντρο Ελληνικής Γλώσσας (ΚΕΓ) και Πανεπιστήμιο Αιγαίου / Τμήμα Μεσογειακών Σπουδών - Εργαστήριο Γλωσσολογίας.

κανείς ότι αρκετά από τα βασικά συστατικά ΓΠΤ έχουν αναπτυχθεί για τα ελληνικά και περιλαμβάνουν πόρους (μονογλωσσικούς, πολυγλωσσικούς, πολυτροπικούς κτλ.), λεξικά, συντακτικούς αναλυτές (parsers), ορθογραφικούς και συντακτικούς διορθωτές, αναγνωριστές ονοματικών οντοτήτων (named entity recognizers), σημασιολογικούς επισημειωτές (semantic annotators), μεταφραστικές εφαρμογές, εφαρμογές υποβοήθησης συγγραφής κειμένων (authoring tools), τεχνολογίες αναγνώρισης και σύνθεσης φωνής, εκπαιδευτικό λογισμικό με χρήση ΓΤ κ.ά.. Η χώρα διαθέτει μια σχετικά ισχυρή βάση σε εφαρμογές επεξεργασίας κειμένου και τεχνολογίας φωνής. Και στους δύο τομείς έχουν αναπτυχθεί τα πρώτα δείγματα εμπορικής ανάπτυξης των ερευνητικών αποτελεσμάτων. Οι πόροι αυτοί μπορούν να ενσωματωθούν και διατεθούν μέσω της υποδομής, μετά από τις κατάλληλες νομικές διευκρινίσεις (εξασφάλιση των κατάλληλων αδειών χρήσης ανάλογα με τους υπάρχοντες όρους χρήσης του υλικού) και τις απαραίτητες τεχνικές μετατροπές. Πέρα από τα νομικά και τεχνικά ζητήματα, ωστόσο, είναι σημαντική η συμφωνία επί της αρχής για τη διασύνδεση πόρων όλων των κατηγοριών μεταξύ τους και με τα εργαλεία ΓΤ, καθώς είναι σημαντική η αξιοποίηση των υπαρχόντων πόρων και της αντίστοιχης τεχνογνωσίας. Η συντακτική διαλειτουργικότητα μάλλον δεν θα δημιουργήσει προβλήματα με βάση τα ευρήματα της έρευνας, ενώ η σημασιολογική διαλειτουργικότητα θα αποτελέσει το επόμενο βήμα μελέτης. Είναι σημαντική, πάντως, η αξιοποίηση των πόρων ακόμα και χωρίς τα εργαλεία ΓΤ (π.χ. αξιοποίηση ορολογικών θησαυρών για την τεκμηρίωση των πόρων, αξιοποίηση ειδικών κειμένων για την κατάρτιση ορολογικών θησαυρών κτλ.).

Εξίσου σημαντική ήταν και η ανάλυση των αναγκών των ερευνητών των ΚΑΕ σε σχέση με τους ΓΠΤ. Όπως φάνηκε από τις μελέτες που έγιναν, οι Έλληνες ερευνητές των ΚΑΕ έχουν αρχίσει να χρησιμοποιούν ψηφιακό υλικό στις μελέτες τους και παρόλο που η χρήση της ΓΤ δεν είναι πολύ διαδεδομένη εκτός συγκεκριμένων εφαρμογών (π.χ. ορθογραφικού διορθωτή για τη σύνταξη κειμένων, εφαρμογών αυτόματης μετάφρασης για τη γρήγορη κατανόηση ξενόγλωσσων κειμένων, στατιστικών εργαλείων ανάλυσης κειμένων κυρίως στον χώρο της εφαρμοσμένης γλωσσολογίας κτλ.), η κατάσταση δείχνει εξαιρετικά υποσχόμενη. Ήδη μέσω των ενημερωτικών δράσεων του έργου, υπήρξε ανταπόκριση και ενδιαφέρον συμμετοχής ερευνητών στην υποδομή. Πέρα από την καταγραφή των αναγκών των ερευνητών γενικά και ανά ειδικότητα, όπως παρουσιάζεται στη μελέτη σκοπιμότητας, διαφάνηκε έντονα η ανάγκη για εκπαιδευτικές και συμβουλευτικές δράσεις, κυρίως σε θέματα τεχνικά (χρήσης και ανάπτυξης πόρων και εργαλείων κτλ.) αλλά και νομικά (κατανόηση δικαιωμάτων πνευματικής ιδιοκτησίας, αδειών χρήσης πόρων κτλ.).

Με βάση τα ευρήματα της χαρτογράφησης των ΓΠΤ, αφενός, και τα πορίσματα της

ανάλυσης αναγκών των χρηστών της ΕΥ, αφετέρου, μπορούμε να συμπεράνουμε ότι η ΕΥ CLARIN-EL σαφώς και είναι σε θέση να συνεισφέρει σημαντικά στην ενίσχυση της ερευνητικής δραστηριότητας στην Ελλάδα. Η κατάσταση των ΓΠΤ για την ελληνική γλώσσα κρίνεται ικανοποιητική, με τα βασικά εργαλεία επεξεργασίας ΓΠ και τους βασικούς πόρους ήδη διαθέσιμους στο ερευνητικό κοινό. Φυσικά, απαιτούνται μετατροπές των εργαλείων και των πόρων αυτών σε μορφές πιο φιλικές προς τον χρήστη, πλήρως διαλειτουργικές μεταξύ τους, ώστε να ενσωματωθούν κατάλληλα προσαρμοσμένα στην ΕΥ. Επίσης, παρατηρούνται κάποιες ελλείψεις σε ΓΠΤ, οι οποίοι έχουν αξιολογηθεί ως ιδιαίτερα σημαντικοί από τους χρήστες (π.χ. εξειδικευμένα γλωσσάρια, θησαυροί κτλ.), αλλά είναι ενθαρρυντικό το γεγονός ότι υπάρχουν εργαλεία που μπορούν να χρησιμοποιηθούν για την υποβοήθηση της ανάπτυξης των πόρων αυτών (π.χ. εργαλεία εξαγωγής ορολογίας από κείμενα) και ότι υπάρχουν οι δομές (ανθρώπινο δυναμικό και φορείς) με την κατάλληλη τεχνογνωσία που μπορούν να κατασκευάσουν τους πόρους αυτούς.

Τέλος, πρέπει να σημειωθεί ότι στη διάρκεια του έργου έχουν γίνει τα πρώτα βήματα για τη σύσταση του εθνικού δικτύου. Οργανωτικά, όπως προαναφέρθηκε, η ΕΥ θα έχει τη μορφή **διευρυμένου ανοιχτού δικτύου** στο οποίο θα συμμετέχουν οργανισμοί με ενεργό δράση στον χώρο των ΓΠΤ και των ΚΑΕ και οι οποίοι θα έχουν δεσμευτεί για τους σκοπούς του έργου. Η κοινοπραξία του CLARIN-EL PREP θα αποτελέσει τον πυρήνα του δικτύου, ενώ στο πλαίσιο της παρούσας φάσης, έχει ήδη αρχίσει να καταρτίζεται το δίκτυο με συμμετοχή από οργανισμούς που έχουν εκδηλώσει ενδιαφέρον συμμετοχής. Ο κατάλογος των μελών του δικτύου (το οποίο και θα παραμείνει ανοιχτό σε όλη τη διάρκεια λειτουργίας της ΕΥ) και το συμφωνητικό συνεργασίας που θα διέπει τις μεταξύ τους σχέσεις, τις υποχρεώσεις και τα δικαιώματά τους θα ολοκληρωθούν στην επόμενη φάση.

Ειδικότερα, στις επόμενες φάσεις του έργου, το δίκτυο

- υλοποιεί τις απαραίτητες τεχνικές εργασίες για την υλοποίηση της ΕΥ, όπως κατασκευή λογισμικού για την αποθήκευση και διάθεση των ΓΠΤ, μετατροπή των ΓΠΤ με στόχο τη διαλειτουργικότητα, περιγραφή/τεκμηρίωση των ΓΠΤ με βάση το συμφωνηθέν σχήμα μεταδεδομένων, κατάρτιση καταλόγου διαθέσιμων ΓΠΤ κτλ.
- προωθεί τη χρήση διεθνών προτύπων και βέλτιστων πρακτικών (όπου υπάρχουν) με στόχο τη διαλειτουργικότητα πόρων και εργαλείων
- στηρίζεται σε και προωθεί τις αρχές των ανοιχτών δεδομένων και του ανοιχτού λογισμικού
- αναλαμβάνει δράσεις ενημέρωσης για την ΕΥ και τη χρήση της, καθώς και προώθησης των ίδιων των ΓΠΤ που περιλαμβάνει

- αναλαμβάνει συμβουλευτικές υπηρεσίες νομικής και τεχνικής φύσης για τους πάροχους και καταναλωτές ΓΠΤ όπου αυτό απαιτείται
- προγραμματίζει εκπαίδευση στη χρήση των ΓΠΤ και υποστηρίζει πάροχους και χρήστες σε όλα τα στάδια έρευνας, από τον μετασχηματισμό δεδομένων μέχρι την παρουσίασή τους
- συνεχίζει την καταγραφή των ΓΠΤ και προγραμματίζει την κάλυψη των κενών τόσο σε πόρους όσο και σε τεχνολογίες για την ελληνική σύμφωνα με τις ανάγκες των χρηστών και τις αρχές του BLARK (<http://www.blark.org/index.php>) προσαρμοσμένων στις ΚΑΕ.

5 Επίλογος

Το CLARIN-EL συνεχίζει την έρευνα και την ανάπτυξη της ΓΤ στην Ελλάδα, θέτοντάς την στην υπηρεσία των ερευνητών κυρίως των ΚΑΕ, αλλά και στην υπηρεσία της ελληνικής κοινωνίας γενικώς. Όραμα του CLARIN-EL είναι να συγκεντρώσει τους αποσπασματικούς πόρους και τα εργαλεία που έχουν κατασκευαστεί για τα ελληνικά κατά τις προηγούμενες δεκαετίες, σε ένα εικονικό αποθετήριο και να τα μετασχηματίσει σε υπηρεσίες που χαρακτηρίζονται από *διαλειτουργικότητα, μονιμότητα, προσβασιμότητα, σταθερότητα και επεκτασιμότητα*, και τα οποία οι χρήστες θα μπορούν να μοιράζονται ή να προσαρμόζουν στους δικούς τους ερευνητικούς σκοπούς. Με αυτόν τον τρόπο, το CLARIN-EL φιλοδοξεί να ανοίξει τον δρόμο στα eHumanities.

Στέλιος Πιπερίδης

Ηλεκ/γος Μηχανικός ΕΜΠ
Αρτέμιδος 6 & Επιδάουρου, 151 25 Μαρούσι
Ηλ.ταχ.: spjp@ilsp.gr

Πένυ Λαμπροπούλου

Αρτέμιδος 6 & Επιδάουρου, 151 25 Μαρούσι
Ηλ.ταχ.: penny@ilsp.athena-innovation.gr

Μαρία Γαβριηλίδου

Γλωσσολόγος
Αρτέμιδος 6 & Επιδάουρου, 151 25 Μαρούσι
Ηλ.ταχ.: maria@ilsp.gr