

ΣΥΛΛΟΓΗ ΗΛΕΚΤΡΟΝΙΚΩΝ ΟΡΟΛΟΓΙΚΩΝ ΠΟΡΩΝ: ΜΕΘΟΔΟΛΟΓΙΑ ΣΥΓΚΕΝΤΡΩΣΗΣ ΚΑΙ ΑΠΟΤΕΛΕΣΜΑΤΑ

Έ. Μάντζαρη, Μ. Γαβριηλίδου, Π. Λαμπροπούλου, Γ. Καραγιάννης

ΠΕΡΙΛΗΨΗ

Στόχος της ανακοίνωσης είναι η παρουσίαση μιας πρώτης απόπειρας για την συγκέντρωση ηλεκτρονικών ορολογικών πόρων, δηλαδή ψηφιοποιημένων επιστημονικών και τεχνικών κειμένων και των περιλαμβανόμενων σε αυτά όρων. Η προσπάθεια αυτή εντάχθηκε στα έργα ΟΡΟΣΗΜΟ που υλοποιήθηκε στο ΙΕΛ. Παρουσιάζεται η μεθοδολογία που ακολουθήθηκε για την συγκέντρωση του υλικού. Οι μεθοδολογικές πρακτικές αποτελούν μεταφορά στον χώρο της ορολογίας της προσέγγισης που χρησιμοποιεί η γλωσσολογική ανάλυση που βασίζεται στην επεξεργασία σωμάτων κειμένων. Ακολούθως παρουσιάζονται και αναλύονται τα αποτελέσματα της διαδικασίας συγκέντρωσης και επεξεργασίας κειμένων και όρων.

ARCHIVE OF ELECTRONIC TERMINOLOGICAL RESOURCES: COLLECTION METHODOLOGY AND RESULTS

E. Mantzari, M. Gavrilidou, P. Labropoulou, G. Carayannis

SUMMARY

This paper aims at presenting a first attempt at collecting electronic terminological resources, i.e. digitised scientific and technical texts and their terms, in the framework of the OROSSIMO project, which was undertaken at ILSP. First, we present the methodology used for the collection of the resources. The approach adopted utilises methodological practices originating in the field of corpus-based linguistics. Finally, the results of the collection process, classification and organisation of the textual and terminological data are presented and analysed.

1 ΕΙΣΑΓΩΓΗ

Η συγκέντρωση και η κωδικοποίηση των ηλεκτρονικών ορολογικών πόρων (terminological resources), δηλαδή κειμένων και όρων, που αποτελούν και το αντικείμενο εργασίας της παρούσας ανακοίνωσης πραγματοποιήθηκε από το Τμήμα Ηλεκτρονικής Λεξικογραφίας του Ινστιτούτου Επεξεργασίας Λόγου (ΙΕΛ) στο πλαίσιο του έργου ΟΡΟΣΗΜΟ, το οποίο είχε ως στόχο την συγκέντρωση και καταγραφή επιστημονικής και τεχνικής ορολογίας της νέας ελληνικής. Το αποτέλεσμα της δράσης αυτής ήταν η δημιουργία

- μιας συλλογής κειμένων της σύγχρονης επιστημονικής και τεχνικής γλώσσας, και
- μιας συλλογής των περιλαμβανόμενων στα κείμενα όρων στην ελληνική και αγγλική γλώσσα.

2 ΜΕΘΟΔΟΛΟΓΙΑ ΣΥΓΚΕΝΤΡΩΣΗΣ ΚΑΙ ΚΩΔΙΚΟΠΟΙΗΣΗΣ ΤΟΥ ΥΛΙΚΟΥ

2.1 ΣΥΓΚΕΝΤΡΩΣΗ ΥΛΙΚΟΥ

Οι μεθοδολογικές πρακτικές που εφαρμόστηκαν για την συγκέντρωση του υλικού αποτελούν μεταφορά στον χώρο της ορολογίας της προσέγγισης που χρησιμοποιεί η γλωσσολογική ανάλυση που βασίζεται στην επεξεργασία σωμάτων κειμένων. Τα πλεονεκτήματα της συγκεκριμένης μεθοδολογικής προσέγγισης έγκεινται στην δυνατότητα επεξεργασίας των όρων μέσα σε πραγματωμένο λόγο, διά του οποίου αναδεικνύονται η χρήση και η σημασία τους. Συνεπώς, στόχος ήταν όχι απλά η συλλογή όρων αλλά η συλλογή όρων συνδεδεμένων με τα κείμενα στα οποία απαντούν.

Ειδικότερα, η ιδιομορφία και οι δυσκολίες του συγκεκριμένου εγχειρήματος συνίσταντο στις εξής δεσμεύσεις, τις οποίες έθετε από την σύλληψή του το έργο ΟΡΟΣΗΜΟ:

- Στόχος της συγκεκριμένης δράσης ήταν η συγκέντρωση όσο το δυνατόν περισσότερων ειδικών γλωσσικών πόρων, που θα ικανοποιούσαν τις απαιτήσεις ενός ευρέος φάσματος εφαρμογών. Η ικανοποίηση του στόχου αυτού επέβαλε μία από κάτω προς τα επάνω προσέγγιση, η οποία δεν έθετε περιορισμούς ως προς τους γνωστικούς τομείς στους οποίους θα ανήκαν τα κείμενα και οι όροι.
- Απαιτείτο η εξεύρεση γλωσσικά ποιοτικών ειδικών κειμένων σε ηλεκτρονική μορφή, τα οποία επιπλέον να μπορούν να παραχωρηθούν από τους συγγραφείς ή/και τους εκδότες τους για περαιτέρω επεξεργασία και χρήση από το ΙΕΛ.
- Επίσης, απαιτείτο η εξασφάλιση μαζί με τα κείμενα και της ορολογίας τους στα ελληνικά και στα αγγλικά, εργασία που θα απαιτούσε την συνδρομή ειδικών του κάθε γνωστικού τομέα, για τον εντοπισμό των όρων αλλά και του μεταφραστικού τους αντιστοίχου στην άλλη γλώσσα.

Εκτιμώντας τις παραπάνω δυσκολίες, θεωρήθηκε ως αποτελεσματικότερη η λύση που προέβλεπε την εξής διαδικασία συγκέντρωσης του ορολογικού υλικού:

- Προκήρυξη διαγωνισμού, ο οποίος απευθυνόταν κυρίως σε μέλη ΔΕΠ των Ανωτάτων και Ανωτέρων Ιδρυμάτων και σε ερευνητές, αλλά και σε ιδιώτες επιστήμονες και συγγραφείς επιστημονικών εγχειριδίων. Αντικείμενο του διαγωνισμού ήταν η υποβολή προτάσεων σχετικά με την παραχώρηση επ' αμοιβή τμημάτων βιβλίων τους, σημειώσεων και επιστημονικών πονημάτων σε οποιονδήποτε γνωστικό τομέα, καθώς και καταλόγου με τους περιεχόμενους σε

αυτά όρους (στα ελληνικά και αγγλικά) σε ηλεκτρονική μορφή¹. Για να θεωρηθεί έγκυρη η υποβολή των προτάσεων, οι συμμετέχοντες έπρεπε να δεσμευτούν για την παραχώρηση κειμένων 200-300 σελίδων σε ηλεκτρονική μορφή (200 λέξεις ανά σελίδα), τα οποία να έχουν γραφτεί μετά το 1976 σε μονοτονικό σύστημα.

➤ Αξιολόγηση (από επιτροπή ειδικών) των προτάσεων με βάση τα εξής κριτήρια:

- Γλωσσική ποιότητα κειμένων (με βάση δείγματά τους), ως προς την ορθογραφία, την σύνταξη και το ύφος.
- Γλωσσική ποιότητα όρων (με βάση δείγματά τους), ως προς την απόδοση και αντιστοίχιση.
- Συνολική παρουσίαση της πρότασης: αξιολογήθηκαν οι επιπλέον πληροφορίες που υποβλήθηκαν από τους διαγωνιζόμενους (π.χ. πληροφορίες γνωστικού τομέα, παρατηρήσεις για δυσκολίες στην απόδοση ενός όρου) και η γενικότερη συγκρότηση της πρότασης.

➤ Πρόκριση 35 προτάσεων και εκτέλεση του τελικού έργου παράδοσης των κειμένων και όρων από τους προκριθέντες.

2.2 ΚΑΤΗΓΟΡΙΟΠΟΙΗΣΗ ΚΑΙ ΚΩΔΙΚΟΠΟΙΗΣΗ ΥΛΙΚΟΥ

2.2.1 ΚΕΙΜΕΝΑ

Το κειμενικό υλικό που συγκεντρώθηκε κατηγοριοποιήθηκε και κωδικοποιήθηκε σύμφωνα με τις διεθνείς προδιαγραφές του Text Encoding Initiative (TEI)², όπως αυτές διαμορφώθηκαν στο πρότυπο Parole³. Το πρότυπο αυτό ορίζει τα στοιχεία που καθορίζουν την ταυτότητα κάθε κειμένου, σε δύο επίπεδα περιγραφής: σε επίπεδο εξωγλωσσικής περιγραφής του κειμένου (κυρίως πληροφορίες βιβλιογραφικού τύπου καθώς και πληροφορίες για την δομική οργάνωση του κειμένου) και σε επίπεδο μεταγλωσσικού χαρακτηρισμού των γλωσσικών οντοτήτων του κειμένου (μορφοσυντακτικός χαρακτηρισμός των λέξεων του κειμένου).

Από το σύνολο αυτών των χαρακτηριστικών, για τον χαρακτηρισμό των κειμένων του ΟΡΟΣΗΜΟΥ επιλέχθηκε ένα υποσύνολο χαρακτηριστικών που αποτέλεσε τις προδιαγραφές κωδικοποίησης οι οποίες δόθηκαν στους συμμετέχοντες στο πρόγραμμα για την κατηγοριοποίηση και κωδικοποίηση του υλικού τους. Οι προδιαγραφές αυτές περιελάμβαναν τα κάτωθι χαρακτηριστικά:

➤ βιβλιογραφική πληροφορία

¹ Η αμοιβή αφορούσε την επεξεργασία των κειμένων και των όρων.

² Text Encoding Initiative, version P2.

³ Parole Common Encoding Schema, PAROLE project (LE2 4017-10379).

- ◆ τίτλος κειμένου
- ◆ όνομα συγγραφέα
- ◆ εκδότης
- ◆ χρονολογία συγγραφής ή πρώτης δημοσίευσης του κειμένου
- ◆ όνομα μεταφραστή (αν πρόκειται για μεταφρασμένο κείμενο)
- ◆ αριθμός λέξεων κειμένου
- ◆ γνωστικός τομέας στον οποίο ανήκει το κείμενο
- δομική οργάνωση κειμένου
- ◆ κεφάλαια του κειμένου
- ◆ κατάλογοι (αν περιλαμβάνονται στο κείμενο)
- ◆ υποσημειώσεις (αν περιλαμβάνονται στο κείμενο)
- μορφή κειμένου: απλό κείμενο χωρίς εντολές μορφοποίησης (Text Only)

Ενδεικτικό παράδειγμα

Τίτλος: Περιβαλλοντική Αρχαιολογία

Όνομα συγγραφέα: Α. Καραλή-Γιαννακοπούλου

Χρονολογία συγγραφής: 1990

Αριθμός λέξεων: 85.306

Γνωστικός τομέας: Περιβαλλοντική Αρχαιολογία

<div>

...

Γενικά οι προϊστορικοί οικισμοί στην πλειοψηφία τους βρίσκονται χτισμένοι σε ελαφρά, γόνιμα και υγρά εδάφη, που προέρχονται από την επί τόπου διάβρωση της πλειοκαινικής μάργας¹, τον φλύσχη² και τον σερπεντίνη³. Φαίνεται σύμφωνα με την κλασική θεώρηση, ότι η παρουσία νερού και αμυντικών δυνατοτήτων δεν επηρέασαν σχεδόν καθόλου την εγκατάσταση των οικισμών παρά μόνον εκείνη των φρουρίων-ανακτόρων...

<note1>

Μάργα: μίγμα ασβεστολιθικών ή δολομιτικών πετρωμάτων με αργιλικές ουσίες.

</note1>

<note2>

Σερπεντίνης (ή οφίτης): προϊόν αποσάθρωσης, περιέχον συχνά μεταλλεύματα χρωμίου και μαγνησίου.

</note2>

</div>

<list>

1. Εδάφη και ιζηματογενή πετρώματα
 - 1α. Βασικές έννοιες
 - 1β. Φυσική διάβρωση και ιζηματογένεση
 - 1γ. Χημική διάβρωση
- </list>

2.2.2 ΟΡΟΙ

Επισημαίνεται ότι οι προδιαγραφές που δόθηκαν περιελάμβαναν οδηγίες για την κωδικοποίηση των όρων και όχι για την επιλογή τους. Η ευθύνη για την επιλογή των όρων αποτελούσε μέρος του έργου των συμμετεχόντων στο ΟΡΟΣΗΜΟ.

Συγκεκριμένα, οι προδιαγραφές προέβλεπαν: αναφορά του ελληνικού όρου που απαντά στο κείμενο, παράθεση του αντίστοιχου αγγλικού όρου (ή όρων, σε περίπτωση που στον ελληνικό όρο αντιστοιχούν περισσότεροι από έναν όροι), και, τέλος, σχόλια ή παρατηρήσεις του κωδικοποιητή (για παράδειγμα, πληροφορίες για περιορισμούς στην χρήση του όρου, ανυπαρξία όρου στην μία από τις δύο γλώσσες, κλπ.), σύμφωνα με το εξής πρότυπο:

ελληνικός όρος; αγγλικός όρος; παρατηρήσεις

Ενδεικτικά παραδείγματα

αγοραστική δύναμη; purchasing power;

ΑΕΠ (Ακαθάριστο Εθνικό Προϊόν); GNP (General National Product);

άλλοθι εκσυγχρονισμού; modernization alibi;

αληθοφάνεια; vraisemblance, verisimilitude, lifelikeness; ο πρώτος ξένος όρος είναι από τα γαλλικά, υιοθετημένος και από τα αγγλικά, ενώ οι άλλοι δύο είναι και αυτοί σε χρήση σε σχέση με τον κινηματογράφο;

αμερικανικός δομισμός; american structuralism; ο "δομισμός" αποδίδεται και ως "στρουκτουραλισμός" επίσης, σε παλαιότερα συνήθως κείμενα ή πλέον με μία δόση υποτίμησης

3 ΑΠΟΤΕΛΕΣΜΑΤΑ

Υποβλήθηκαν 85 συνολικά προτάσεις, από τις οποίες προκρίθηκαν μετά από αξιολόγηση 35, οι οποίες καλύπτουν 19 γνωστικούς τομείς.

Στον πίνακα παρουσιάζεται η κατανομή του κειμενικού και του ορολογικού υλικού ανά γνωστικό τομέα. Συνολικά, το κειμενικό υλικό που συγκεντρώθηκε περιλαμβάνει **2.891.885** λέξεις, ενώ το ορολογικό υλικό αποτελείται από **24.214** όρους. Τα διαγράμματα 1 και 2 παρουσιάζουν σχηματικά την κατανομή αυτή.

Πίνακας

Γνωστικός τομέας	Αριθμός λέξεων κειμένων	Αριθμός όρων
Ανθρωπολογία	105.163	1.248
Ανοσοχημεία	68.400	1.915
Αρχαιολογία	187.610	783
Αστρονομία	85.947	439
Βιολογία	193.326	859
Γενετική	86.800	361
Γεωλογία	173.284	1.825
Γεωπονική	34.097	181
Γλωσσολογία	61.871	772
Δίκαιο	139.495	3.774
Ιατρική	161.719	1.434
Ιστορία	397.191	533
Μέσα Μαζικής Επικοινωνίας	62.146	1.683
Μηχανολογία	242.769	2.747
Οικονομία	342.072	2.634
Περιβαλλοντική Αρχαιολογία	13.030	475
Πληροφορική	382.671	1.263
Ραδιοϊσοτοπική	54.810	316
Φυσιολογία Φυτών	99.484	972
Σύνολο	2.891.885	24.214

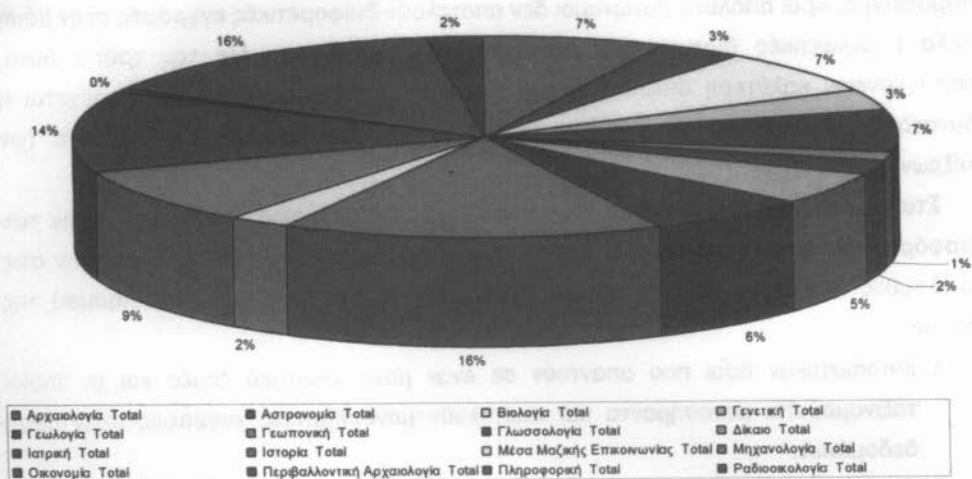
4 ΕΠΕΞΕΡΓΑΣΙΑ ΥΛΙΚΟΥ

Την συγκέντρωση του υλικού ακολούθησε φάση επεξεργασίας από το ΙΕΛ με στόχο την καλύτερη ταξινόμηση και αποθήκευσή του ώστε να είναι εύκολα προσβάσιμο και αξιοποιήσιμο.

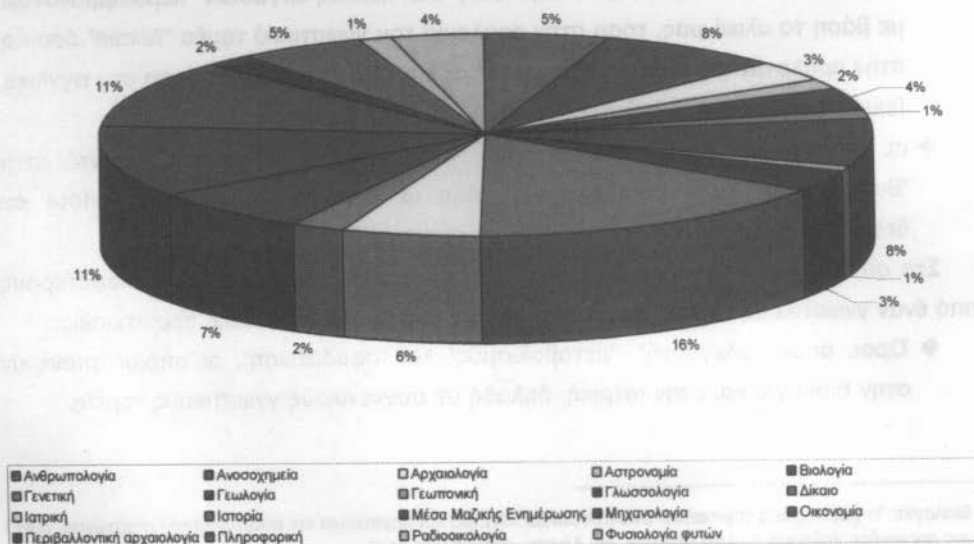
Συγκεκριμένα, τα κείμενα κατηγοριοποιήθηκαν επιπλέον ως προς τα χαρακτηριστικά "κειμενικό είδος" (genre) και "θέμα" (topic), όπως αυτά ορίζονται στο πρότυπο Parole, και στην συνέχεια ενσωματώθηκαν στην βάση κειμένων του ΙΕΛ.

Όσον αφορά τους όρους, η επεξεργασία βρίσκεται στο στάδιο της καλύτερης οργάνωσης του υλικού και σχεδιασμού της βάσης στην οποία θα αποθηκευτεί. Έχει προβλεφθεί σύνδεση των κειμένων και των όρων, ώστε να είναι δυνατή και αποτελεσματική η πρόσβαση από τα κείμενα στους όρους και αντίστροφα.

Διάγραμμα 1:
Κατανομή κειμενικού υλικού ανά γνωστικό τομέα



Διάγραμμα 2:
Κατανομή ορολογικού υλικού ανά γνωστικό τομέα



Η προσέγγιση ανάλυσης των όρων είναι εννοιοκεντρική (concept-based). Αυτό σημαίνει ότι οι όροι δεν αντιμετωπίζονται ως λέξεις αλλά ως πραγματώσεις εννοιών. Για παράδειγμα, όροι απόλυτα συνώνυμοι δεν αποτελούν διαφορετικές εγγραφές στην βάση αλλά εναλλακτικές διατυπώσεις στην ίδια εγγραφή-έννοια. Με τον τρόπο αυτό, επιτυγχάνεται καλύτερη απεικόνιση των σχέσεων μεταξύ των όρων και παρέχεται η δυνατότητα πρόσβασης στα κείμενα μέσω και των όρων-εννοιών και όχι μόνο των λέξεων.

Στο πλαίσιο της ταξινόμησης του υλικού έγιναν συγκρίσεις μεταξύ των όρων των διαφόρων γνωστικών τομέων. Τα αποτελέσματα των συγκρίσεων αυτών οδήγησαν στις ακόλουθες παρατηρήσεις, οι οποίες καθορίζουν τις προδιαγραφές σχεδιασμού της βάσης.

1. Εντοπίστηκαν όροι που απαντούν σε έναν μόνο γνωστικό τομέα και οι οποίοι ταξινομούνται μονοσήμαντα και αποτελούν μονοσήμαντες εγγραφές στην βάση δεδομένων.
2. Εντοπίστηκαν όροι που ανήκουν σε περισσότερους από έναν τομείς. Ως τέτοιοι θεωρήθηκαν οι ακόλουθοι:
 - 2.1. Όροι που απαντούν με την ίδια ακριβώς μορφή στα ελληνικά και με την ίδια μετάφραση στα αγγλικά, σε διαφορετικούς γνωστικούς τομείς.

Σε αυτή την περίπτωση, επιβάλλεται ο έλεγχος των συμφραζομένων στα αντίστοιχα κείμενα προκειμένου να εξακριβωθεί αν ο ίδιος όρος αναφέρεται:

- ◆ στην ίδια έννοια: π.χ. οι όροι "κυρώσεις" και "κύκλος εργασιών" περιλαμβάνονται, με βάση το υλικό μας, τόσο στην ορολογία του γνωστικού τομέα "Δίκαιο" όσο και στην ορολογία του τομέα "Οικονομικά" με την ίδια ακριβώς απόδοση στα αγγλικά, [sanctions] και [turnover] αντίστοιχα, ή
- ◆ σε διαφορετική έννοια: π.χ. ο όρος "περιβάλλον" [environment] απαντά στην "Βιολογία" και στην "Πληροφορική" αλλά με διαφορετική σημασία⁴, οπότε και θεωρείται διαφορετικός όρος.

Στο σύνολο των όρων που θα μπορούσαν να θεωρηθούν κοινοί σε περισσότερους από έναν γνωστικούς τομείς, θα μπορούσαν να διερευνηθούν οι εξής περιπτώσεις:

- ◆ Όροι, όπως "φλεγμονή", "μεταβολισμός" και "αφυδάτωση", οι οποίοι απαντούν στην Βιολογία και στην Ιατρική, δηλαδή σε συγγενικούς γνωστικούς τομείς.

⁴ Βιολογία: "ο χώρος μέσα στον οποίο αναπτύσσονται κύτταρα και οργανισμοί και από τον οποίο εξαρτώνται, αφού τους προσφέρει ενέργεια ή υλικά στοιχεία, και δέχεται τα προϊόντα του".

Πληροφορική: "το πλαίσιο επιλογών που παρέχεται στην χρήση υπολογιστικού προγράμματος με την αξιοποίηση του λογισμικού και του μηχανολογικού εξοπλισμού τους" [Μπαμπινιώτης, Γ. (1998) Λεξικό της Νέας Ελληνικής Γλώσσας].

- ◆ Όροι, όπως "δομή", "λειτουργία", "μεθοδολογία", που εμφανίζονται σε αρκετούς γνωστικούς τομείς μη συγγενικούς, και οι οποίοι θα μπορούσαν να θεωρηθούν ως όροι γενικότερα της θεωρίας των επιστημών.
- ◆ Όροι, των οποίων η εμφάνιση σε γνωστικούς τομείς μη συγγενικούς οφείλεται στην χρήση, από τον έναν τομέα, διαδικασιών έρευνας ή μεθοδολογικών εργαλείων ανάλυσης που χρησιμοποιήθηκαν αρχικά από τον άλλο τομέα: ενδεικτικά, ο τομέας της Θεωρίας του Κινηματογράφου δανείζεται από την Γλωσσολογία τον όρο "διπλή άρθρωση" για την ανάλυση ταινιών.

2.2. Όροι που απαντούν με την ίδια ακριβώς μορφή στα ελληνικά, αλλά με διαφορετική μετάφραση στα αγγλικά, σε διάφορους γνωστικούς τομείς.

Στην κατηγορία των όρων αυτών εντοπίστηκαν οι εξής υποπεριπτώσεις:

- ◆ οι όροι να έχουν την ίδια σημασία: στην περίπτωση αυτή, συνήθως ο αγγλικός όρος αποδίδεται άλλοτε με αγγλική λέξη και άλλοτε με κάποια λέξη ελληνογενούς ή λατινογενούς προέλευσης. Για παράδειγμα, ο όρος "κράνιο" εμφανίζεται στο υλικό μας με δύο μεταφράσεις ([skull] και [cranium]), ο όρος "δέρμα" ως [skin] και [derm] και ο όρος "θεραπεία" ως [cure] και [therapy].
- ◆ οι όροι να έχουν διαφορετική σημασία: π.χ. "αξίωμα" [axiom] στην "Γλωσσολογία" ενώ [public office] στα "ΜΜΕ", "χρόνος" [time] στα "ΜΜΕ" ενώ [tense] στην "Γλωσσολογία", "σπέρμα" [sperm] στην "Γενετική" ενώ [seed] στην "Φυσιολογία Φυτών".

2.3. Όροι που απαντούν με την ίδια ακριβώς μορφή στα αγγλικά, αλλά με διαφορετική μετάφραση στα ελληνικά, σε διάφορους γνωστικούς τομείς: π.χ. ο όρος [liver] εμφανίζεται στα ελληνικά ως "συκώτι" και "ήπαρ", ο όρος [host] εμφανίζεται ως "Ξενιστής" και "δέκτης". Όπως και στην περίπτωση 2.2., διακρίνουμε δύο υποπεριπτώσεις, ανάλογα με το αν οι όροι έχουν την ίδια ή διαφορετική σημασία.

Δεδομένου ότι η προσέγγισή μας είναι εννοιοκεντρική, προτείνεται διαφορετική αντιμετώπιση για τις ανωτέρω περιπτώσεις εφόσον οι όροι αναφέρονται σε ίδια ή διαφορετική σημασία:

- Διαφορετική σημασία: ο όρος καταχωρίζεται σε διαφορετικές εγγραφές. Αν και η μετάφραση στα αγγλικά ή ελληνικά είναι κοινή, η διάκριση του όρου είναι εμφανής από το πεδίο του γνωστικού τομέα, το οποίο διαφέρει για την κάθε εγγραφή.
- Ίδια σημασία: ο όρος καταχωρίζεται ως μία εγγραφή, με εναλλακτικές αποδόσεις είτε στα ελληνικά είτε στα αγγλικά. Όσον αφορά το πεδίο του γνωστικού τομέα, προβλέπονται δύο λύσεις:
 - ◆ πολλαπλές τιμές στο πεδίο του γνωστικού τομέα
 - ◆ δημιουργία ιεραρχίας γνωστικών τομέων: στην περίπτωση αυτή, οι όροι θα συνδέονται με τον υπερώλυμο γνωστικό τομέα.

5 ΣΥΜΠΕΡΑΣΜΑΤΑ

Το έργο ΟΡΟΣΗΜΟ ήδη από το στάδιο της σύλληψής του είχε ως στόχο την συγκέντρωση ορολογικών πόρων με βάση τις μεθοδολογικές αρχές της γλωσσολογικής ανάλυσης που βασίζεται σε επεξεργασία σωμάτων κειμένων. Στις ίδιες μεθοδολογικές αρχές βασίστηκαν η κατηγοριοποίηση και η οργάνωση του υλικού σε ένα ενιαίο περιβάλλον το οποίο θα διευκολύνει την σύνδεση των όρων με το κειμενικό περιβάλλον χρήσης τους. Τέλος, η μετάβαση από το επίπεδο της λεξικής πραγμάτωσης του όρου στο επίπεδο της έννοιας επιτρέπει την συστηματικότερη οργάνωση του υλικού και την αποτελεσματικότερη πρόσβαση στο ορολογικό υλικό.

Ε. Μάντζαρη, Μ. Γαβριηλίδου, Π. Λαμπροπούλου, Γ. Καραγιάννης

Ερευνητές ΙΕΛ

Ινστιτούτο Επεξεργασίας Λόγου (ΙΕΛ),

Επιδάουρου και Αρτέμιδος 6, 15125 Μαρούσι

Helena@ilsp.gr, maria@ilsp.gr, penny@ilsp.gr, gcara@ilsp.gr