

Ηλεκτρονικά Λεξικά που βασίζονται σε Παράλληλα Κείμενα. Δυνατότητες και ο ρόλος του χρήστη

Ελίνα Συμσερίδου
Αριστοτέλειο Πανεπιστήμιο Θεσσαλονίκης

Abstract

This paper presents a study on the shift from print to electronic dictionaries (ED), the potentials that the electronic medium offers to users and the crucial role of the user in the ED content development. By the term *electronic*, I refer to internet dictionaries functioning as portals that give access to dictionary and corpus data at the same time. The motivation behind this study lies in the necessity for translation students to improve their information-mining and terminological documentation competences. The evaluation of these new, user-oriented information tools draws upon the examination of the design of the following online dictionaries: Glosbe, Linguee and TAUS. Compiling parallel texts, either corpora or translation memories, from the Internet or uploaded by users, these e-dictionaries provide a wealth of up to date evidence to show how terms or words are used in context. It goes without saying that the results given must be evaluated, but the same applies for terms deprived of context found in print dictionaries. The market for the print dictionary is shrinking and an increasing number of users access information via the internet. Research on the use of ED is the new topic in the rapidly evolving field of electronic lexicography.

Keywords: electronic dictionaries, electronic lexicography, parallel texts, corpora, translation memories, user's role, Glosbe, Linguee, TAUS

1. Εισαγωγή

Κατά τις τελευταίες δεκαετίες, με την έλευση της ψηφιακής εποχής, σημειώθηκε μια σημαντική αλλαγή στη μορφοποίηση των λεξικών τα οποία, εκτός από την έντυπη, απέκτησαν και ηλεκτρονική μορφή. Η παροχή επιγραμμικών λεξικών υπήρξε ένα τεράστιο άλμα στην εξέλιξη των λεξικών, αλλά η διασύνδεσή τους με περιεχόμενο που εντοπίζεται στο διαδίκτυο είναι αυτή που προσφέρει περισσότερες δυνατότητες δημιουργίας εργαλείων και χρήσεων.

Ειδικότερα, στόχος της παρούσας εργασίας είναι η παρουσίαση:

1. των διαφόρων ειδών ηλεκτρονικών λεξικών, με έμφαση στα ηλεκτρονικά λεξικά που βασίζονται σε παράλληλα κείμενα,
2. της δομής, των χρήσεων και των δυνατοτήτων τους, ως εργαλείων τεκμηρίωσης στη μεταφραστική διαδικασία,
3. του ρόλου του χρήστη στη διαμόρφωση του περιεχομένου τους.

Για την επίτευξη των στόχων της έρευνας θα βασιστούμε στη βιβλιογραφία και την ανάλυση των χαρακτηριστικών των διαδικτυακών λεξικών Glosbe, Linguee και TAUS.

2. Από την έντυπη στην ηλεκτρονική μορφή και οι δυνατότητες του ηλεκτρονικού μέσου

Όπως επισημαίνει η Vintar (2008: 153), η ψηφιακή εποχή έφερε στον κλάδο της μετάφρασης μια πιο σημαντική επανάσταση από αυτήν της μετάβασης από το χαρτί στην οθόνη του υπολογιστή. Η πληθώρα των ηλεκτρονικών – συχνά πολύγλωσσων – κειμένων στο διαδίκτυο, που διατίθενται στις περισσότερες γλώσσες του κόσμου είναι ένας από τους λόγους που τα έντυπα λεξικά δεν μπορούν πλέον να θεωρούνται η

πρωταρχική πηγή πληροφοριών, που σχετίζονται με τη μετάφραση. Το ηλεκτρονικό μέσο επαναπροσδιορίζει το είδος του περιεχομένου των λεξικών και ανοίγει νέες προοπτικές σχετικά με τον τρόπο με τον οποίο αυτό μπορεί να οργανωθεί, να παρουσιαστεί και να χρησιμοποιηθεί.

Ο όρος *ηλεκτρονικό λεξικό* χρησιμοποιείται γενικά για τον χαρακτηρισμό ενός λεξικού σε ηλεκτρονική μορφή. Σήμερα τα ηλεκτρονικά λεξικά αντιμετωπίζονται ως καινοτόμα γλωσσικά εργαλεία που προσφέρουν γρήγορη και εύκολη πρόσβαση σε συλλογές δομημένων ηλεκτρονικών δεδομένων, η αξιολόγηση των οποίων γίνεται από τον ίδιο τον χρήστη. Οι κατασκευαστές των νέων αυτών εργαλείων εκμεταλλεύονται στο έπακρο τα πλεονεκτήματα και τις δυνατότητες που προσφέρει ο υπολογιστής και αντλούν το περιεχόμενο των λεξικών τους – το οποίο διαμορφώνεται με βάση τις ανάγκες των χρηστών – από σύγχρονες πηγές όπως: τα ηλεκτρονικά σώματα κειμένων και το Διαδίκτυο.

Σύμφωνα με την Varantola (2002: 35), το ηλεκτρονικό λεξικό του μέλλοντος θα είναι *«ένα εργαλείο ή ένας αριθμός εργαλείων, στην εργαλειοθήκη, ενός επαγγελματία χρήστη, που θα συνυπάρχει/ουν με λοιπά εργαλεία γλωσσικής τεχνολογίας»*

Πολλοί ειδικοί εκφράζουν την άποψη ότι τα επιγραμμικά λεξικά (online dictionaries) είναι τα λεξικά του μέλλοντος και προβλέπουν την κατάργηση των έντυπων λεξικών στο εγγύς μέλλον. Για πολλούς εκδοτικούς οίκους και ακαδημαϊκά πρότζεκτ κατασκευής λεξικών, το διαδίκτυο αποτελεί ήδη την κύρια πλατφόρμα. Ο De Schryver (2003: 163-187) μόνο, απαριθμεί 118 πλεονεκτήματα των ηλεκτρονικών λεξικών έναντι των έντυπων.

Ωστόσο οι Fuertes-Olivera και Bergholtz (2011: 15) θέτουν το εξής ερώτημα:

«Εξακολουθούμε να αποκαλούμε λεξικά τα ηλεκτρονικά εργαλεία του μέλλοντος, ειδικά όταν ορισμένα από αυτά περιλαμβάνουν επίσης συστηματικές εισαγωγές, βιβλία γραμματικής και λοιπά στοιχεία που απέχουν μακράν από τα παραδοσιακά λεξικά;».

Οι ίδιοι προτείνουν ως λύση την απομάκρυνση από τον όρο ‘λεξικό’ και εισάγουν τον όρο ‘εργαλείο πληροφόρησης’ (information tool) ως ένα είδος όρου-ομπρέλα στο πλαίσιο του οποίου οι ερευνητές δύνανται να σχεδιάσουν οποιοδήποτε εργαλείο, όπως και αν αυτό αποκαλείται, με στόχο την ικανοποίηση των αναγκών του χρήστη σε τέσσερις περιστάσεις χρήσης:

- την επικοινωνιακή,
- τη γνωστική,
- τη λειτουργική και
- την ερμηνευτική.

Κατά τον Gouws (2011: 17) αναφορικά με την εμφάνιση πραγματικών ηλεκτρονικών λεξικών, δηλαδή των διαδικτυακών λεξικών και όχι αυτών σε μορφή CD ή DVD ROM, μπορεί να θεωρηθεί ότι έχουμε απλώς μια αλλαγή στο μέσο και ότι ένα λεξικό παραμένει λεξικό.

Άλλωστε, όπως εκτιμά ο Lew (2011: 246) για τον μέσο χρήστη η διαφορά μεταξύ των διαφόρων χαρακτηρισμών είναι πολύ λεπτή.

Όροι όπως:

- Επιγραμμικά λεξικά (Online Dictionaries, Karagol-Ayan 2007)
- Επιγραμμικά εργαλεία αναφοράς (Online Reference Tools, Bergholtz κ.ά. 2009)
- Ψηφιακά λεξικά, (Digital Dictionaries, Wiegand 2010)

- Διαδικτυακά λεξικά, (Internet Dictionaries, Gouws 2011)
- Διαδικτυακά ηλεκτρονικά λεξικά (Web-based electronic dictionaries, Spohr 2011) και
- Υπολογιστικά λεξικά (Computational Dictionaries, Vaquero κ.ά. 2013)

έχουν χρησιμοποιηθεί από ερευνητές, ως συνώνυμοι του όρου ηλεκτρονικό λεξικό, για να περιγράψουν αυτό το νέο είδος γλωσσικών εργαλείων η δομή των οποίων περιλαμβάνει μια μηχανή αναζήτησης για την εύρεση όρων και λεξικών συμπλεγμάτων, παράλληλα κείμενα (ΣΚ ή μεταφραστικές μνήμες) και μεταγλωσσικά κείμενα όπως λεξιλόγια, γλωσσάρια και γραμματικές.

Το 2003, ο De Schryver διακρίνει, με βάση το ηλεκτρονικό μέσο, δύο κατηγορίες ηλεκτρονικών λεξικών. Τα ηλεκτρονικά λεξικά σε αυτόνομους υπολογιστές (stand-alone computers) και τα ηλεκτρονικά λεξικά σε δικτυωμένους υπολογιστές (networked computers).

Στους αυτόνομους υπολογιστές ανήκουν τα:

- Λεξικά σε συσκευές παλάμης (Pocket Electronic Dictionaries), τα οποία χρησιμοποιούνται στις Ασιατικές χώρες και σε χώρες όπου οι μικροηλεκτρονικές συσκευές είναι οικονομικές και δημοφιλείς και τα
- Λεξικά σε οπτικούς δίσκους οι οποίοι θεωρούνται παρωχημένη τεχνολογία, που τείνει να καταργηθεί όπως και οι δισκέτες. Πολλά λεξικά, όπως το Merriam-Webster Learner's Dictionary (έκδοση 2008), δεν περιλαμβάνουν DVDROM και διατίθενται σε έντυπη και επιγραμμική έκδοση μόνο. Αξιοσημείωτο είναι το γεγονός ότι, το 2012 οι εκδόσεις MacMillan και η Encyclopedia Britannica ανακοίνωσαν τη διακοπή της έντυπης έκδοσης των λεξικών και βιβλίων τους αντίστοιχα.

Στους δικτυωμένους υπολογιστές ανήκουν τα:

- Διαδικτυακά λεξικά (Internet dictionaries)

Το 2009 ο Fuertes-Olivera αναγνωρίζει δύο είδη διαδικτυακών λεξικών:

- τα Θεσμικά διαδικτυακά έργα αναφοράς (Institutional internet reference works) τα οποία είναι λεξικά που δημιουργούνται από αναγνωρίσιμους οργανισμούς και μπορεί να παρέχονται δωρεάν ή με συνδρομή και
- τα Συλλογικά πολύγλωσσα διαδικτυακά έργα αναφοράς (Collective multiple-language internet reference works) τα οποία προκύπτουν από τη συλλογική εργασία μιας κοινότητας χρηστών και το περιεχόμενό τους διατίθεται συνήθως δωρεάν (Wikipedia, Wictionary, Omega Wiki).

Παρατηρούμε ότι, ο όρος *ηλεκτρονικό λεξικό*, αλλά και οι όροι που έχουν χρησιμοποιηθεί ως συνώνυμοί του, δεν προσδιορίζουν τη μορφή του περιεχομένου των γλωσσικών εργαλείων που περιγράφουν. Ηλεκτρονικό ή ψηφιακό μπορεί να χαρακτηριστεί και το λεξικό σε μορφή DVD ROM, όπως επίσης ηλεκτρονικά, ψηφιακά, διαδικτυακά ή επιγραμμικά είναι και τα λεξικά με τα χαρακτηριστικά και τις λειτουργίες του Wiktionary. Τα Ηλεκτρονικά Λεξικά που βασίζονται σε Παράλληλα Κείμενα – παραλλαγή του αγγλικού όρου Corpus-based Electronic Dictionaries – αν και μακροσκελής και δύσχρηστος όρος καθιστά σαφές το είδος και τις λειτουργίες του νέου εργαλείου πληροφόρησης, όπως το χαρακτηρίζουν οι Fuertes-Olivera και Bergenholtz.

3. Λεξικά που βασίζονται σε παράλληλα κείμενα

Στην πρόσφατη ιστορία της κατασκευής λεξικών, παρατηρείται ο εντεινόμενος ρόλος της τεχνολογίας των πληροφοριών και των δομημένων δεδομένων: Σώματα Κειμένων (ΣΚ), βάσεις δεδομένων, χρήση της Επεκτάσιμης Γλώσσας Σήμανσης (XML). Σταδιακά, το λεξικό, όπως το αντιμετωπίζει ο χρήστης, είναι πιθανό να αποτελέσει

απλά ένα επιφαινόμενο σε μια δομημένη λεξιλογική βάση ή ένα ΣΚ. Πράγματι, καθώς οι διεπαφές των ΣΚ γίνονται όλο και πιο εξειδικευμένες, μπορούν να χρησιμοποιηθούν με τρόπους όμοιους με αυτούς των λεξικών. Για τον χρήστη άλλωστε δεν έχει σημασία τι βρίσκεται ‘πίσω από την προσωπίδα’ εφόσον η διεπαφή μπορεί να χρησιμοποιηθεί ως ένα είδος λεξικού (Lew 2011: 246).

Η παραγωγή πληροφοριών που βασίζεται στην ανάλυση αυθεντικών δεδομένων λόγου είναι μια παλιά καθιερωμένη λεξικογραφική πρακτική. Πριν από την εμφάνιση των Ηλεκτρονικών Σωμάτων Κειμένων (ΗΣΚ), οι λεξικογράφοι χρησιμοποιούσαν δεδομένα από έντυπες πηγές, τα οποία εξήγαγαν αποκλειστικά για τη σύνταξη ενός συγκεκριμένου λεξικού.

Με τη διάδοση του ηλεκτρονικού μέσου έγιναν διαθέσιμα όλο και περισσότερα ΣΚ τα οποία διέπυρναν επίσης τις δυνατότητες της λεξικογραφικής εργασίας. Αναμενόμενα, οι λεξικογράφοι άδραξαν την ευκαιρία για την κατασκευή λεξικών που βασίζονται σε ΣΚ. Αυτό πρακτικά σημαίνει ότι οι όροι των επιγραμμικών λεξικών μπορούν να συνδεθούν με τη σχετική συλλογή κειμένων, δίνοντας στο χρήστη τη δυνατότητα άμεσης πρόσβασης σε αυτήν.

Η βάση για ένα επιγραμμικό λεξικό περιλαμβάνει πρωτογενείς πηγές, όπως ΗΣΚ και δευτερογενείς πηγές, δηλαδή, λοιπές έντυπες και ηλεκτρονικές πηγές, όπως δίγλωσσα λεξικά και βιβλία γραμματικής.

Ένα επιγραμμικό λεξικό δεν κατασκευάζεται μόνο από λεξικογράφους, αλλά και από γλωσσολόγους ΣΚ, υπολογιστικούς γλωσσολόγους, επιστήμονες υπολογιστών, τεχνολόγους κειμένων, γραφίστες και πιθανά επαγγελματίες άλλων ειδικοτήτων.

Ορισμένα ηλεκτρονικά λεξικά παρουσιάζονται ως πύλες (portals) που παρέχουν, με μία μόνο αναζήτηση, πρόσβαση σε πολλαπλές γλωσσολογικές πηγές, μεταξύ αυτών σε λεξικά και ΣΚ. Εισάγοντας δηλαδή έναν όρο αναζήτησης, ο χρήστης λαμβάνει πληροφορίες και από τις δύο πηγές δεδομένων ταυτόχρονα. Η συνύπαρξη αυτή λεξικού και ΣΚ ή η διασύνδεση του λεξικού με διαδικτυακό περιεχόμενο παρέχει στο χρήστη πληροφορίες για τη γλώσσα στη χρήση της, οι οποίες μπορούν να χρησιμεύσουν ως πρότυπο για την παραγωγή ίδιου κειμένου από τον χρήστη.

Ωστόσο, η συμπληρωματική σχέση είναι έμμεση και πρέπει να αξιολογείται από τον χρήστη (Heid κ.ά. 2012: 276-277).

3.1. Το πολύγλωσσο επιγραμμικό λεξικό *Glosbe*

Το *Glosbe* είναι ένα «πολύγλωσσο επιγραμμικό λεξικό» σύμφωνα με τον χαρακτηρισμό που του αποδίδουν οι δημιουργοί του. Κατασκευάστηκε από δύο προγραμματιστές, με έδρα την Πολωνία, και διατέθηκε στο διαδίκτυο το 2011.¹ Υποστηρίζει οποιαδήποτε γλώσσα έχει κωδικό ISO693-3, δηλαδή περίπου 7.000 γλώσσες επί του παρόντος, ενώ γίνονται δεκτές αιτήσεις, από χρήστες, για τη δημιουργία νέων λεξικών σε προτεινόμενα γλωσσικά ζεύγη. Πρόκειται για μία διαδραστική βάση δεδομένων που εξελίσσεται συνεχώς, με την προσθήκη νέων στοιχείων, και προσφέρει τη δυνατότητα δημιουργίας οποιουδήποτε δίγλωσσου λεξικού, πρακτική που αντιστοιχεί στην ύπαρξη 50.000.000 δίγλωσσων λεξικών στον ιστότοπο του *Glosbe*.

Τα στοιχεία που παρέχονται στο *Glosbe* προέρχονται από παράλληλα κείμενα ανοιχτού κώδικα που εντοπίζονται στο διαδίκτυο και τη συμμετοχή των χρηστών οι οποίοι δύνανται να προσθέσουν παραδείγματα μεταφράσεων. Δίπλα σε κάθε απόδοση

¹ Πηγή: www.glosbe.com

μετάφρασης, εμφανίζεται ένα εικονίδιο που δηλώνει την ταυτότητα του συντάκτη ή της πηγής. Τέτοιες πηγές είναι οι ιστοσελίδες της Ευρωπαϊκής Κεντρικής Τράπεζας (ECB) και της Επίσημης Εφημερίδας της Ε.Ε. (ΟJ), το Europarl (Παράλληλο Σώμα Κειμένων για Στατιστική Μηχανική Μετάφραση), το Open Subtitles, η βάση δεδομένων Eurlex, το Setimes (Παράλληλο Σώμα Κειμένου που περιέχει άρθρα εφημερίδων στις Βαλκανικές Γλώσσες) και το επιγραμματικό λεξικό Wictionary.

Στο Glosbe έχουν προστεθεί και μεταφραστικές μνήμες, ώστε κάθε όρος αναζήτησης να εμφανίζεται μαζί με το συγκείμενό του, να παρέχονται δηλαδή παραδείγματα μεταφρασμένων προτάσεων. Οι μεταφραστικές μνήμες προέρχονται κυρίως από δημοσιευμένα παράλληλα ΣΚ, στα οποία η πρόσβαση είναι ελεύθερη. Το Glosbe λειτουργεί κυρίως βάσει «ανάκλησης» παρά ακρίβειας και συχνά επισημαίνονται λανθασμένες ή κακώς στοιχισμένες λέξεις. Επίσης, η διεπαφή χρήστη δεν επιτρέπει την ταξινόμηση των αποτελεσμάτων ανάλογα με τις μεταφραστικές παραλλαγές ή το κείμενο-πηγή.

Με την είσοδο του χρήστη στην ιστοσελίδα, εμφανίζεται αριστερά στην οθόνη μία λίστα ορισμένων διαθέσιμων όρων και εκφράσεων σε αλφαβητική σειρά. Το Glosbe χρησιμοποιεί τεχνολογία αποκατάληξης (stemming) δηλαδή την αναγωγή στο θέμα της λέξης. Δεν λειτουργεί όμως ως τυπικό λεξικό και περιέχει πολύ περισσότερους όρους από αυτούς που εμφανίζονται στη λίστα, χάρη στις μεταφραστικές μνήμες. Η αναζήτηση, δηλαδή, δεν περιορίζεται στο περιεχόμενο της λίστας, αλλά ο χρήστης μπορεί να αναζητήσει οποιονδήποτε όρο ανεξάρτητα εάν εμφανίζεται στη λίστα ή όχι. Τα αποτελέσματα αναζήτησης εμφανίζονται σε δύο στήλες, στην αριστερή τα παραδείγματα προτάσεων ή παραγράφων στη γλώσσα-πηγή (ΓΠ) και στη δεξιά τα παραδείγματα στη γλώσσα-στόχο (ΓΣ). Δεδομένου ότι, για κάθε λέξη αναζήτησης δεν εμφανίζεται ένα αλλά περισσότερα παραδείγματα, η μετάφρασή της διαμορφώνεται ανάλογα με το συγκείμενο και ο χρήστης έχει τη δυνατότητα να επιλέξει αναλύοντας τα αποτελέσματα, την απόδοση που ταιριάζει στη δική του παραγωγή κειμένου. Επιπλέον, πάνω από τις δύο στήλες εμφανίζεται η μετάφραση του όρου ή της έκφρασης, επεξεργασμένη με Google translate και η απόδοση, όπως αυτή εμφανίζεται στο εκάστοτε δίγλωσσο λεξικό.

Ορισμένες φορές ενδέχεται το ισοδύναμο του όρου αναζήτησης από μία ΓΠ Α (π.χ. Αγγλικά) σε ΓΣ Β (π.χ. Ελληνικά) να μην εμφανίζεται στο αντίστοιχο δίγλωσσο λεξικό. Υπάρχει όμως περίπτωση ο συγκεκριμένος όρος να εντοπιστεί σε δίγλωσσο λεξικό από τη ζητούμενη ΓΠ Α σε ΓΣ Γ (π.χ. Γαλλικά) και ο γαλλικός όρος να μεταφράζεται στη συνέχεια στα ελληνικά. Εφόσον ο όρος Α μπορεί να μεταφραστεί σε Γ, εμφανίζεται στη σελίδα και ο όρος Γ. Επειδή όμως, η μετάφραση αυτή δεν μπορεί να επιβεβαιωθεί, οφείλει ο χρήστης να επαληθεύει το αποτέλεσμα. Για να αποφασίσει εάν πρόκειται για ορθή μετάφραση, εμφανίζεται επίσης και ο όρος Β ως ενδιάμεση απόδοση.

Σε πολλές γλώσσες χρησιμοποιούνται ειδικοί χαρακτήρες όπως για παράδειγμα à, é, ç, ù, κλπ. στα Γαλλικά. Για την εισαγωγή αυτών των χαρακτήρων υπάρχει το εικονίδιο ενός πληκτρολογίου στο οποίο περιλαμβάνονται οι ειδικοί αυτοί χαρακτήρες για κάθε γλώσσα. Το Glosbe χρησιμοποιεί επίσης το μη-απωλεστικό σχήμα κωδικοποίησης χαρακτήρων UTF-8. Αυτό πρακτικά σημαίνει ότι, ακόμα και εάν ένας όρος πληκτρολογηθεί λανθασμένα ορθογραφικά, το Glosbe θα προτείνει τη σωστή ορθογραφία του.

3.2. Το μεταφραστικό εργαλείο *Linguee*

Παρόμοιο, αλλά πιο προσεκτικά δομημένο σύστημα, είναι το *Linguee*.² Κατασκευάστηκε από τους Dr. Gereon Frahling και Leonard Fink και ολοκληρώθηκε το Μάιο του 2010, περιλαμβάνοντας αρχικά το ζεύγος γλωσσών Γερμανικά-Αγγλικά. Σήμερα, βρίσκεται μεταξύ των 100 δημοφιλέστερων ιστοσελίδων στη Γερμανία και εμπλουτίζεται συχνά με νέα ζεύγη γλωσσών, θέτοντας την υπηρεσία διαθέσιμη σε ένα ολοένα αυξανόμενο αριθμό χρηστών. Πρόκειται για ένα μεταφραστικό εργαλείο – σύμφωνα με τους κατασκευαστές – που συνδυάζει ένα δίγλωσσο επιγραμμικό (online) λεξικό και μια μηχανή αναζήτησης και το οποίο μπορεί να χρησιμοποιηθεί είτε ως λεξικό είτε ως μηχανή αναζήτησης στο διαδίκτυο. Τα αποτελέσματα της αναζήτησης του *Linguee* παρέχουν παραδείγματα προτάσεων από διάφορες πηγές, δίνοντας στο χρήστη τη δυνατότητα να μελετήσει το πώς έχει μεταφραστεί ο όρος αναζήτησης μέσα σε διάφορα συγκείμενα. Και το *Linguee*, όπως και το *Glosbe*, παρέχει επίσης την απόδοση των όρων από δίγλωσσα λεξικά, η οποία εμφανίζεται πάνω από τις δύο στήλες αποτελεσμάτων.

Η πλειοψηφία των ενδεικτικών προτάσεων που εμφανίζονται στη δεξιά πλευρά προέρχεται από δίγλωσσες ιστοσελίδες, κυρίως από επαγγελματικά μεταφρασμένες ιστοσελίδες εταιρειών, οργανισμών και πανεπιστημίων. Άλλες πηγές περιλαμβάνουν πολύτιμα έγγραφα της ΕΕ και προδιαγραφές διπλωμάτων ευρεσιτεχνίας. Ένα εξειδικευμένο πρόγραμμα ηλεκτρονικού υπολογιστή, ένας ανιχνευτής ιστού (web crawler) αναζητά συνεχώς στο διαδίκτυο νέες πολύγλωσσες ιστοσελίδες. Αυτές οι σελίδες ανιχνεύονται αυτόματα και εξάγονται οι μεταφρασμένες προτάσεις και λέξεις, δίνοντας στο χρήστη τη δυνατότητα να εντοπίσει εξαιρετικά σπάνια τεχνική ορολογία. Τα κείμενα στη συνέχεια αξιολογούνται από έναν αλγόριθμο μηχανικής μάθησης, ο οποίος φιλτράρει τις υψηλής ποιότητας μεταφράσεις προς προβολή. Αυτό το σύστημα διαθέτει επίσης τη δυνατότητα αυτόματης εκμάθησης νέων κριτηρίων ποιότητας από τα σχόλια των χρηστών, προκειμένου να διαχωρίζει τις ορθές μεταφράσεις από τις προβληματικές. Έχει διαπιστωθεί, για παράδειγμα, ότι μια σελίδα είναι συνήθως μεταφρασμένη από μηχανή μετάφρασης αν περιέχει τη λέξη "Wordpress" και πολλές λέξεις έχουν μεταφραστεί κυριολεκτικά. Μέσω αυτής της διαδικασίας εκπαίδευσης, ο αλγόριθμος μαθαίνει συνεχώς να εντοπίζει χιλιάδες τέτοιες συσχετίσεις και να εξάγει αξιόπιστα τις ορθότερες μεταφράσεις. Μέσω του προγράμματος αυτού έχουν ήδη συγκριθεί περισσότερες από ένα τρισεκατομμύριο προτάσεις. Στο τέλος, διατηρείται μόνον το κορυφαίο 0,01%, δηλαδή 100 εκατομμύρια από τις μεταφρασμένες προτάσεις. Το *Linguee* αποτελείται από δύο μέρη, το ένα είναι το λεξικό, τα περιεχόμενα του οποίου θεωρούνται πιο αξιόπιστα, και το δεύτερο μέρος είναι η μηχανή αναζήτησης. Η μηχανή αναζήτησης παρουσιάζει παραδείγματα δίγλωσσων προτάσεων από το διαδίκτυο. Οι καταχωρήσεις αυτές δεν έχουν υποβληθεί σε επεξεργασία και έλεγχο και το *Linguee* δεν μπορεί να εγγυηθεί για την ποιότητά και την ορθότητά τους. Πολλές από αυτές, ωστόσο, προέρχονται από αξιόπιστες πηγές, όπως το Ευρωπαϊκό Κοινοβούλιο. Η πηγή μιας ενδεικτικής πρότασης εμφανίζεται ως επιλέξιμος σύνδεσμος κάτω από την πρόταση. Με τη βοήθεια αυτών των ενδεικτικών προτάσεων μπορεί κανείς να παρατηρήσει πώς χρησιμοποιείται η λέξη ή η έκφραση αναζήτησης και στη συνέχεια να αποφασίσει, με βάση και την πηγή του παραδείγματος, αν θα εμπιστευθεί

² Πηγή: www.linguee.com

τη μετάφραση ή όχι. Η ένδειξη προειδοποίησης υποδεικνύει μια πιθανώς εσφαλμένη μετάφραση που θα πρέπει πρώτα να αξιολογείται.

Η λειτουργία αναζήτησης του Linguee επιτρέπει τη χρήση τριών χαρακτήρων μπαλαντέρ (wildcards) “”, - , + προκειμένου τα αποτελέσματα που θα εμφανιστούν να είναι πιο συγκεκριμένα. Πιο αναλυτικά, τα εισαγωγικά επιτρέπουν την αναζήτηση προτάσεων, όπου οι λέξεις απαντώνται με την ακριβή μορφή και σειρά που έχουν πληκτρολογηθεί.

Το σύμβολο μείον χρησιμοποιείται για την εξαίρεση λέξεων ενώ το σύμβολο συν επιβάλλει τη συμπερίληψη λέξεων σε μία συγκεκριμένη μορφή. Αυτό είναι χρήσιμο με τις αντωνυμίες που διαφορετικά αγνοούνται από το Linguee.

Παράδειγμα: Η αναζήτηση *επικοινωνία+μας* βρίσκει το «*Μείνετε σε επικοινωνία μαζί μας*» αλλά όχι το «*Θα μείνουμε σε επικοινωνία*».

Το Linguee περιλαμβάνει 100 εκατομμύρια ενδεικτικές προτάσεις, όγκος εξαιρετικά μεγάλος για να διερευνηθεί και να δοθούν πληροφορίες σχετικά με τη γλώσσα-πηγή και τη γλώσσα-στόχος. Οι διαδικτυακές διευθύνσεις (URL) των ενδεικτικών προτάσεων επίσης δεν βοηθούν. Πολλές εταιρείες διαθέτουν ιστοσελίδες με διευθύνσεις .com, οι οποίες δεν παρέχουν πληροφορίες για την εθνικότητα της εταιρείας ή, αν και εδρεύουν σε αγγλόφωνη χώρα, έχουν τμήματα στις ιστοσελίδες τους μεταφρασμένα στα αγγλικά και κατά συνέπεια η ΓΠ δεν είναι απαραίτητα η αγγλική.

Στη δεξιά πλευρά των αποτελεσμάτων μετάφρασης υπάρχει ένα σύμβολο-αντίχειρας, με το οποίο ο χρήστης μπορεί να βαθμολογήσει τις ενδεικτικές προτάσεις, για παράδειγμα, εάν η μετάφραση είναι λανθασμένη ή κακής ποιότητας, ακόμη και αν αυτό ισχύει μόνο για τμήματα της πρότασης. Αν μια ενδεικτική πρόταση δεν αντιστοιχεί στην αναζήτηση του χρήστη, δίνεται η δυνατότητα επιλογής «*Η μετάφραση δεν ταιριάζει με την αναζήτησή μου*». Αν η λέξη ή η φράση αναζήτησης πρέπει να απαλειφτεί σε μια ενδεικτική πρόταση, μπορεί ο χρήστης επίσης να επισημάνει την ενδεικτική πρόταση, κάνοντας κλικ στο σύμβολο με τον αντίχειρα προς τα κάτω.

Η επισήμανση των όρων αναζήτησης στις ενδεικτικές προτάσεις-αποτελέσματα δημιουργείται αυτόματα και κατά συνέπεια μπορεί να είναι λάθος. Η επισήμανση έχει σε γενικές γραμμές στόχο να βοηθήσει τους χρήστες να εντοπίζουν τη λέξη αναζήτησης ταχύτερα. Μπορεί κανείς να αναφέρει μία λανθασμένη επισήμανση σε μία ενδεικτική πρόταση, κάνοντας κλικ στο σύμβολο με τον αντίχειρα προς τα κάτω και επιλέγοντας «*Έχουν επισημανθεί λάθος λέξεις*».

3.3. Το αποθετήριο δεδομένων TAUS

Στις αρχές του 2007, η Εταιρεία Χρηστών για την Αυτοματοποιημένη Μετάφραση (Translation Automation User Society / TAUS),³ μια παγκόσμια κοινότητα που απαριθμεί στους κόλπους της 100 οργανώσεις, ανέλαβε τον καθορισμό των εργαλείων που θεωρούνται απαραίτητα για την ανταπόκριση στις ολοένα αυξανόμενες εμπορικές και κοινωνικές απαιτήσεις στον κλάδο της μετάφρασης. Σύντομα, έγινε αντιληπτή η ανάγκη δημιουργίας μίας κοινής βάσης δεδομένων μεταφραστικών μνημών και ορολογίας που θα υποστήριζε τη μελλοντική ανάπτυξη και καινοτομία. Το 2008, με την υποστήριξη 45 διεθνών εταιριών, ιδρύθηκε η Ένωση Δεδομένων TAUS η οποία παρέχει, έκτοτε, μια τεράστια πολύγλωσση βάση δεδομένων. Το αποθετήριο δεδομένων TAUS (TAUS Data Repository) περιλαμβάνει δεδομένα μετάφρασης με τη μορφή

³ Πηγές: www.tausdata.org,

http://www.europarl.europa.eu/stoa/webdav/site/cms/shared/2_events/workshops/2013/20131203/Jaap%20van%20der%20Meer.pdf

παραγράφων, προτάσεων ή τμημάτων. Πρόκειται για μία πλατφόρμα υπολογιστικού νέφους (cloud platform) με περισσότερες από 55 δισεκατομμύρια λέξεις σε 2.200 ζεύγη γλωσσών. Η βασική του χρήση είναι η βελτίωση των συστημάτων μηχανικής μετάφρασης και αποτελεί πολύτιμο εργαλείο για τους αγοραστές και παρόχους μεταφραστικής τεχνολογίας και υπηρεσιών μετάφρασης για τη βελτίωση της ποιότητας των μεταφρασμάτων και της αυτοματοποίησης. Επίσης, αποδεικνύεται σημαντική πηγή πληροφοριών και ορολογικής τεκμηρίωσης τόσο για τους επαγγελματίες μεταφραστές, όσο και για τους φοιτητές μετάφρασης και μεταφραστικής τεχνολογίας. Τα δεδομένα, δηλαδή οι μεταφραστικές μνήμες, παρέχονται από οποιοδήποτε μέλος ή χρήστη του TAUS οι οποίοι συμβάλλουν στη διεύρυνσή του, ανεβάζοντας τα δικά τους δεδομένα και κατεβάζοντας άλλα για δική τους χρήση. Επομένως, το αποθετήριο δεδομένων TAUS περιλαμβάνει όλα τα δημόσια διαθέσιμα δεδομένα μεταφραστικών μνημών (κάτοχοι δεδομένων: Ε.Ε., Κυβέρνηση και Κοινοβούλιο του Καναδά, Η.Ε., κ.ά.) και μεταφραστικές μνήμες που παρέχονται από τα μέλη, δηλαδή από μεταφραστικά γραφεία, παρόχους γλωσσικών υπηρεσιών (LSP's) μεγάλες εταιρίες (Microsoft, Oracle, Intel, eBay, SDL, κ.ά.), καθώς και από ανεξάρτητους χρήστες (μεταφραστές, κ.ά.).

Το TAUS υποστηρίζει πάνω από 50 γλώσσες και όλα τα κείμενα είναι στοιχισμένα σε επίπεδο λέξεων. Στην αναζήτηση παρέχεται μια σειρά ρυθμίσεων προκειμένου τα αποτελέσματα να φιλτράρονται ώστε να ανταποκρίνονται όσο το δυνατό περισσότερο στις ανάγκες του χρήστη. Για παράδειγμα, δίνεται η δυνατότητα επιλογής τομέα (7 επιλογές: Υπολογιστικό Λογισμικό, Νομικές Υπηρεσίες, Φαρμακευτική και Βιοτεχνολογία, κ.ά.) και κειμενικού είδους (6 επιλογές: Οδηγίες χρήσεως, Διαφημιστικό υλικό, Πρότυπα, Νόμοι και Κανονισμοί, κ.ά.).

Επιπλέον, κάθε λέξη του κειμένου-πηγή είναι γραμματικά δεικτοδοτημένη έτσι ώστε, τοποθετώντας τον κέρσορα πάνω σε κάθε λέξη να εμφανίζεται η γραμματική του κατηγορία και να επιτρέπονται (απλές) αναζητήσεις ανάλογα με το μέρος του λόγου (επίθετο, ουσιαστικό, ρήμα) στο οποίο ανήκει. Η ακρίβεια της αναζήτησης PoS εξαρτάται από την ακρίβεια της δεικτοδότησης (PoS tagger). Ιδιαίτερο χαρακτηριστικό του TAUS είναι η παροχή πληροφοριών για την κατεύθυνση της μετάφρασης, που έχει γίνει από επαγγελματία μεταφραστή, για κάθε ζεύγος προτάσεων, ώστε ο χρήστης να γνωρίζει εάν πρόκειται για πρωτότυπο ή μεταφρασμένο κείμενο. Και στο TAUS, όπως και στο Glosbe και το Linguee, δίπλα σε κάθε αποτέλεσμα υπάρχουν τα εικονίδια που παρέχουν πληροφορίες για την ιδιοκτήτρια εταιρία των δεδομένων και επιτρέπουν στον χρήστη να αναφέρει τυχόν λάθη.

Τα πλεονεκτήματα του TAUS, σύμφωνα με τους δημιουργούς του, είναι τα εξής:

- Αυξάνει την ταχύτητα ολοκλήρωσης της μετάφρασης και την ποιότητα του μεταφράσματος,
- Βοηθά στην επίλυση προβλημάτων διασφάλισης ποιότητας,
- Βοηθά στην αποσαφήνιση της ορολογίας,
- Είναι ένα πολύτιμο εργαλείο για την έρευνα και την εκπαίδευση μεταφραστών.

4. Διαφέρουν οι μεταφραστικές μνήμες από τα παράλληλα κείμενα;

Στην προηγούμενη ενότητα, χρησιμοποιήθηκε ο όρος παράλληλα κείμενα για την αναφορά τόσο στα σώματα κειμένων όσο και τις μεταφραστικές μνήμες. Εύλογα, ενδέχεται να προκύψει το ερώτημα γιατί οι τρεις αυτές έννοιες χρησιμοποιούνται εναλλακτικά και πώς σχετίζονται μεταξύ τους. Σύμφωνα με τον Gouadec (2007: 279):

«Μια Μεταφραστική Μνήμη (πολλές φορές γνωστή ως 'Δικείμενο') είναι ένα αρχείο που περιλαμβάνει ένα πίνακα με αντιστοιχίες μεταξύ των

τμημάτων δύο κειμένων, ένα από τα οποία είναι η μετάφραση του άλλου. Με άλλα λόγια, μια μεταφραστική μνήμη είναι ένας πίνακας όπου στοιχίζονται τμήματα λόγου σε δύο διαφορετικές γλώσσες»

Στην υπολογιστική γλωσσολογία, τα κείμενα-πηγή και οι μεταφράσεις τους ονομάζονται Παράλληλα Κείμενα και ο όρος χρησιμοποιείται ως συνώνυμο του όρου Δικείμενο. Έτσι, ένα δικείμενο σώμα (bibtex corpus) μπορεί να προκύψει αυτόματα από παράλληλα κείμενα.

Όπως διαβάζουμε στους Wright και Budin (2001: 867) ένα δικείμενο ή στοιχισμένο κείμενο αποτελείται από κείμενο-πηγή και κείμενο-στόχος η αντιστοίχιση των οποίων έγινε είτε μετά την ολοκλήρωση της μεταφραστικής διαδικασίας είτε το μετάφρασμα προέκυψε σε αντιστοίχιση, χρησιμοποιώντας μια μεταφραστική μνήμη,

Στα περιβάλλοντα των Μεταφραστικών Μνημών τα στοιχισμένα κείμενα ονομάζονται συχνά και Παράλληλα Κείμενα. Οι Foo και Merkel (2010: 164-166) χρησιμοποιούν τους όρους Μεταφραστικές Μνήμες και Παράλληλα Κείμενα ως συνώνυμες και ταυτόσημες έννοιες, ενώ η Safadi (2010: 25) υποστηρίζει ότι τα Παράλληλα Κείμενα μπορούν να θεωρηθούν Μεταφραστικές Μνήμες.

Θεωρώντας τα 'παράλληλα κείμενα' ή 'δικείμενα' – με τους όρους της υπολογιστικής γλωσσολογίας – ως μια συλλογή κειμένων-πηγή και των μεταφράσεών τους, καταλήγουμε στο συμπέρασμα ότι, είτε οι μεταφράσεις παράγονται εκτός μεταφραστικών μνημών και η αντιστοίχισή τους γίνεται σε δεύτερο στάδιο, είτε παράγονται με τη χρήση μεταφραστικών μνημών και η αντιστοίχιση γίνεται ταυτόχρονα, το αποτέλεσμα παραμένει το ίδιο και αυτό είναι ένα δίγλωσσο παράλληλο αντι-στοιχισμένο σώμα κειμένου. Αυτή είναι και η εικόνα που λαμβάνει ο χρήστης, δηλαδή ένας δίστηλος πίνακας με το κείμενο-πηγή και το κείμενο-στόχος.

Για το λόγο αυτό, στην παρούσα εργασία επιλέχθηκε ο όρος *Ηλεκτρονικά Λεξικά που βασίζονται σε Παράλληλα Κείμενα* – αντί σε Σώματα Κειμένων, από το Corpus-based Electronic Dictionaries – ως ευρύτερος όρος, για το χαρακτηρισμό των εργαλείων Glosbe, Linguee και TAUS τα οποία αντλούν το γλωσσικό υλικό τους από παράλληλα κείμενα, σώματα κειμένων ή μεταφραστικές μνήμες, ή συνδυασμό αυτών.

5. Ο ρόλος του χρήστη

Η έρευνα για τη χρήση του λεξικού είναι το νεότερο πεδίο έρευνας στο ευρύτερο πλαίσιο της έρευνας για τα λεξικά. Οι Heid, Prinsloo και Bothma το 2012 διαπιστώνουν ότι, ο προσανατολισμός των λεξικών προς τις ανάγκες των χρηστών θεωρείται η πιο αδιαμφισβήτητη βασική προϋπόθεση. Αρκετοί είναι οι ερευνητές (Bergenholtz και Tarp 2002, Tarp 2008, Wiegand κ.ά. 2010, Müller-Spitzer 2014) που έχουν επισημάνει το γεγονός ότι τα λεξικά είναι χρηστικά εργαλεία και ότι οι λεξικογράφοι καλούνται να τα σχεδιάζουν βάσει των υποθέσεών τους για την προϋπάρχουσα γνώση του χρήστη, τις συνθήκες χρήσης και τις ιδιαίτερες ανάγκες του χρήστη (με δεδομένη προϋπάρχουσα γνώση) σε δεδομένη συνθήκη.

Ομοίως, ο Gouws (2011: 25-29) υποστηρίζει ότι ο σχεδιασμός και η κατασκευή των λεξικών καθορίζονται από τις ανάγκες του χρήστη. Η προσέγγιση αυτή, επηρεάζει τις αποφάσεις που αφορούν τις λειτουργίες, τα περιεχόμενα και τις δομές των λεξικών. Για το σκοπό αυτό προτείνει, τη διαδικασία της ταυτόχρονης (simultaneous feedback)⁴ και της συνεχούς ανατροφοδότησης/ανάδρασης (continuous feedback), με τις οποίες οι

⁴ Την θεωρητική έννοια της ταυτόχρονης ανατροφοδότησης, εισήγαγε και περιέγραψε ο De Schryver στη μεταπτυχιακή του εργασία, το 1997 (Gouws, 2014: 548).

λεξικογράφοι, στη διάρκεια κατασκευής ενός λεξικού, λαμβάνουν πληροφορίες και σχόλια από τους χρήστες. Προκειμένου οι λειτουργίες ενός λεξικού να ανταποκρίνονται στις ανάγκες του χρήστη, οι λεξικογράφοι καλούνται να κατασκευάσουν το προφίλ των χρηστών της ομάδας στην οποία απευθύνονται (Gouws 2011, Dung 2011). Η χρήση αρχείων καταγραφής δεδομένων (log files) μπορεί να καταδείξει ποια λήμματα και πληροφορίες αναζητήθηκαν από τον χρήστη και κυρίως ποια δεν εντοπίστηκαν ή δεν αναζητήθηκαν στο λεξικό (De Schryver και Joffe 2004, Bergholtz και Johnsen 2005).⁵ Το μέσο αυτό παρατήρησης της συμπεριφοράς αναζήτησης του χρήστη μπορεί να παράγει πραγματικά αξιόπιστα αποτελέσματα (Bergholtz 2011: 33).

Αξίζει να σημειωθεί ότι, το 2014 το Διεθνές Συνέδριο Euralex, που έλαβε χώρα στο Μπολτσάνο της Ιταλίας και το οποίο απευθύνεται σε λεξικογράφους, ορολόγους, εκδότες λεξικών, γλωσσολόγους, ερευνητές και εκπροσώπους εταιριών γλωσσικής τεχνολογίας, είχε τίτλο: «**Ο χρήστης στο επίκεντρο**», καθώς μια από τις προκλήσεις των λεξικογραφικών προϊόντων είναι η ανταπόκρισή τους στις συγκεκριμένες και ποικίλες ανάγκες των χρηστών.

6. Συμπερασματικά

Τα ηλεκτρονικά λεξικά που βασίζονται σε παράλληλα κείμενα διαμορφώνουν ένα καινοτόμο ηλεκτρονικό περιβάλλον που προσφέρει ταχύτατη ανάκληση εννοιών ενταγμένων στο συγκεκριμένο τους και πρόσβαση σε αυθεντικά δεδομένα γλώσσας. Πρόκειται για συγκεντρωτικές βάσεις δομημένων ηλεκτρονικών δεδομένων που παρέχουν, με τη μορφή παράλληλων προτάσεων και παραγράφων, πληροφορίες για τη γλώσσα στη χρήση της, επεκτείνοντας την έννοια της ισοδυναμίας του λεξικού. Και αυτό γιατί το εύρος των ισοδυναμιών που προκύπτουν από τη συμφραστική προσέγγιση είναι πολύ πιο εκτενές από τον αριθμό ισοδύναμων που εντοπίζονται σε ένα λεξικό, καθώς πολλοί όροι εξαρτώνται από τα συμφραζόμενα. Επιπλέον, το διαδίκτυο, από όπου αντλείται το υλικό των ηλεκτρονικών λεξικών, λόγω του μεγέθους του και της αδιάκοπης τροφοδότησής του με επικαιροποιημένα δεδομένα αποτελεί το πρώτο μέσο εμφάνισης νεολογισμών και νέων χρήσεων υπαρχόντων λέξεων. Σαφώς, τέτοιου είδους δεδομένα επιτείνουν την ανάγκη ελέγχου της υπόθεσης διατύπωσης έναντι της, υποθετικά, «συνήθους» χρήσης.

Όπως προκύπτει από τη βιβλιογραφική έρευνα και τη μελέτη της δομής και των λειτουργιών των λεξικών Glosbe, Linguee και TAUS, ο στόχος για την κατασκευή ενός ηλεκτρονικού λεξικού είναι να λειτουργεί, για τον χρήστη, ως γέφυρα μεταξύ του όρου και μιας ποικιλίας διαδικτυακών πηγών, ώστε να ενισχύει την πρόσβαση σε πιθανά σχετικές πληροφορίες.

Καθώς, ο σχεδιασμός ενός ηλεκτρονικού/διαδικτυακού λεξικού εξαρτάται από τις ανάγκες του χρήστη, η επιλογή των πληροφοριών που θα παρέχει και η διάθεση των ανάλογων συνδέσεων-πηγών, εξαρτάται από το σκοπό για τον οποίο αυτό χρησιμοποιείται και τους χρήστες στους οποίους απευθύνεται.

Για την αποφυγή της εμφάνισης περιττών πληροφοριών και προκειμένου η πρόσβαση σε δεδομένα και η εμφάνισή τους να είναι επιλέξιμη, έχουν προταθεί από διάφορους ερευνητές προηγμένα εργαλεία αναζήτησης και διαμόρφωσης του προφίλ του χρήστη ώστε να παρέχονται ακριβής πληροφορίες ανάλογα με τις ανάγκες του.

Η λεξικογραφία βρίσκεται επί του παρόντος στο στάδιο μετάβασης από την τεχνολογία του έντυπου βιβλίου στην τεχνολογία του Διαδικτύου. Όπως επισημαίνει ο Joffe (2004)

⁵ Πηγή: <http://download2.hermes.asb.dk/archive/download/Hermes-34-6-HB-Johnsen.pdf>

η μετάβαση από τα έντυπα στα ηλεκτρονικά λεξικά δεν αποτέλεσε επιλογή των εκδοτών αλλά υπαγορεύτηκε από τους χρήστες λεξικών.

Είναι πολύ νωρίς για να περιγράψουμε τη μορφή που θα έχουν τα καινοτόμα λεξικά του μέλλοντος. Ωστόσο διαφαίνεται ότι, η μελλοντική – ηλεκτρονική – λεξικογραφία θα βασίζεται σε Σώματα Κειμένων.

Βιβλιογραφικές αναφορές

- Bergenholtz H. (2011). *Access to and Presentation of Needs-Adapted Data in Monofunctional Internet Dictionaries. E-Lexicography: the internet, digital initiatives and lexicography*, UK: A&C Black, 30-53.
- Bergenholtz H., S. Nielsen και S. Tarp (2009). *Lexicography at a crossroads: dictionaries and encyclopedias today, lexicographical tools tomorrow*, Bern: Peter Lang.
- Bergenholtz H. και S. Tarp (2002). *Die moderne lexikographische Funktionslehre. Diskussionsbeitrag zu neuen und alten Paradigmen, die Wörterbücher als Gebrauchs-gegenstände verstehen*. In *Lexicographica* 22, 145-155.
- De Schryver G.M. (2003). *Lexicographers' dreams in the electronic-dictionary age*. *International Journal of Lexicography* 16(2): 143-199.
- De Schryver G.M. και D. Joffe (2004). *On how electronic dictionaries are really used*. In Williams G., S. Vessier (eds.), *Proceedings of the 11th EURALEX. International Congress EURALEX 2004, Lorient, France, July 6th-10th 2004*, 187-196. Lorient: EURALEX.
- Dung J. (2009). *Online Dictionaries in a Web 2.0 Environment*. In Bergenholtz H., Nielsen S., Tarp S. (eds). *Lexicography at a Crossroads: Dictionaries and Encyclopedias Today, Lexicographical Tools Tomorrow*. Bern: Peter Lang, 135-164.
- Foo J. και M. Merkel (2010). "Computer aided term banks creation and standardization – Building standardized term banks through automated term extraction and advanced editing tools". In: Thelen M. και F. Steurs (eds.) *Terminology in everyday life*, John Benjamins, 163-180.
- Fuertes-Olivera P.A. (2009). "The function theory of lexicography and electronic dictionaries: Wiktionary as a prototype of collective free multiple-language internet dictionary". In Bergenholtz H., S. Nielsen, S. Tarp (eds.), *Lexicography at a Crossroads: Dictionaries and Encyclopedias Today, Lexicographical Tools Tomorrow*. Bern: Peter Lang, 99-134.
- Fuertes-Olivera P.A και H. Bergenholtz (2011). *E-Lexicography: the internet, digital initiatives and lexicography*, UK: A&C Black.
- Gouadec D. (2007). *Translation as a profession*, Amsterdam and Philadelphia: John Benjamins.
- Gouws R.H. (2011). "Learning, unlearning and innovation in the planning of electronic dictionaries". In: Fuertes-Olivera P.A. και H. Bergenholtz (eds.) *E-Lexicography: the internet, digital initiatives and lexicography*. GB: A&C Black, 17-29.
- Gouws R.H. (2014). *Dictionaries. An international encyclopedia of lexicography: Supplementary Volume: Recent developments with focus on electronic and computational lexicography*, Berlin and Boston: Walter de Gruyter GmbH.
- Heid U., D.J. Prinsloo και T.J.D. Bothma (2012). *Dictionary and corpus data in a common portal: state of the art and requirements for the future*, *Lexicographica: International Annual for Lexicography*, Volume 28, 269-289.
- Karagol-Ayan B. (2007). *Resource Generation from Structured Documents for Low-density Languages*, US: ProQuest
- Lew R. (2011). "Online Dictionaries of English". In: Fuertes-Olivera P.A. και H. Bergenholtz (eds.), *E-Lexicography: The Internet, Digital Initiatives and Lexicography*, London: Continuum, 230-250.
- Müller-Spitzer C. (2014). *Using online dictionaries*. Berlin and Boston: Walter de Gruyter GmbH.
- Safadi H. (2010). *Crosslingual Implementation of Linguistic Taggers Using Parallel Corpora*, Lulu.com
- Spohr D. (2011). *A Multi-layer Architecture for Pluri-Monofunctional Dictionaries*. In: Fuertes-Olivera P.A. και H. Bergenholtz (eds.) *E-Lexicography: the internet, digital initiatives and lexicography*. GB: A&C Black, 103-120.
- Tarp S. (2008). *Lexicography in the Borderland between Knowledge and Non-knowledge: General Lexicographical Theory with Particular Focus on Learner's Lexicography*. Tübingen: Max Niemeyer.
- Vaquero A., F. Alvarez και F. Sáenz (2013). "Representing Computational Dictionaries in Relational Databases". In: *Dictionaries. An International Encyclopedia of Lexicography: Supplementary Volume: Recent Developments with Focus on Electronic and Computational Lexicography*, Walter de Gruyter 2014, 1209-1226.

- Varantola K. (2002). "Use and Usability of Dictionaries: Common Sense and Context Sensibility". In Correard M.H. (ed.) *Lexicography and Natural Language Processing, a Festschrift in Honour of B.T.S. Atkins*. Euralex 2002, 30-44.
- Vintar Š. (2008). "Corpora in Translator Training and Practice". In: Anderman G. και M. Rogers (eds.) *Incorporating Corpora: The Linguist and the Translator*. Multilingual Matters Ltd, 153-163.
- Wiegand H.E., M. Beißwenger, R.H. Gouws, M. Kammerer, A. Storrer, W. Wolski (2010). *Wörterbuch Zur Lexikographie und Wörterbuchforschung. Dictionary of Lexicography and Dictionary Research. Vol. 1 (A-C)*. Berlin: Walter de Gruyter.
- Wright S.E. και G. Budin (2001). *Handbook of Terminology Management: Application-oriented terminology management*, Vol. 2, Amsterdam and Philadelphia: John Benjamins, 867.