

25 Διαφορές μεταξύ μηχανικής και ανθρώπινης μετάφρασης αρκτικόλεξων και ακρωνύμων από ελάσσονες προς μείζονες γλώσσες: Καθορισμός μεταφραστικών στρατηγικών

Δημήτρης Μπουμπάρης, Χρήστος Γιαννούτσος

ΠΕΡΙΛΗΨΗ

Σκοπός της παρούσας εργασίας είναι η εξέταση των διαφορών μεταξύ μηχανικής και ανθρώπινης μετάφρασης των αρκτικόλεξων και ακρωνύμων, με έμφαση στα γλωσσικά ζεύγη Ελληνικά-Αγγλικά, Ελληνικά-Γαλλικά και Ελληνικά-Γερμανικά. Αφού αρχικά μελετώνται οι θέσεις των γλωσσών αυτών στο παγκόσμιο γλωσσικό σύστημα, αξιολογούνται οι μεταφράσεις των αρκτικόλεξων και των ακρωνύμων στα γλωσσικά ζεύγη Ελληνικά-Αγγλικά, Ελληνικά-Γαλλικά και Ελληνικά-Γερμανικά. Κατά την αξιολόγησή τους παρατηρούμε εσφαλμένη απόδοση βάσει της τυπολογίας λαθών DQF-MQM. Για αυτόν τον λόγο, ανατρέξαμε στη μέθοδο αναγνώρισης ονοματισμένων οντοτήτων (Named Entity Recognition – NER) για να εντοπίσουμε πταίσματα στη μεταφραστική στρατηγική των συστημάτων νευρωνικής μηχανικής μετάφρασης. Τέλος, καθορίζουμε τις μεταφραστικές στρατηγικές για όλα τα γλωσσικά ζεύγη με γλώσσα αφετηρίας την ελληνική.

Differences between machine and human translation of initialisms and acronyms from minor into major languages: Defining translation strategies

Dimitris Boumparis, Christos Yannoutsos

ABSTRACT

The purpose of this paper is to examine the differences between machine and human translation of initialisms and acronyms, emphasizing on the Greek-English, Greek-French and Greek-German language pairs. First of all, we study the positions of these languages in the world language system, evaluating at the same time the translations of initialisms and acronyms in the language pairs Greek-English, Greek-French and Greek-German. While evaluating them, we observe incorrect translations based on the DQF-MQM error typology. For this reason, we referred to the Named Entity Recognition (NER) method in order to identify the errors in the translation strategy of neural machine translation (NMT) systems. Finally, we define the translation strategies for all language pairs with Greek as the source language.

0 Εισαγωγή

Στην ελληνόφωνη βιβλιογραφία υπάρχουν δημοσιεύσεις για τη μετάφραση ακρωνύμων και αρκτικόλεξων (Παριανού, 2004, Κεραμίδης, 2015), με έμφαση στις μεταφραστικές στρατηγικές που ακολουθούνται στο γλωσσικό ζεύγος Αγγλικά-Ελληνικά (Κελάνδριας, 2007), καθώς και το Ελληνικό Πρότυπο για τον σχηματισμό ελληνικών αρκτικόλεξων και ακρωνύμων ΕΛΟΤ 1455: 2017 (Βαλεοντής, 2019). Ωστόσο, παρατηρείται ότι δεν έχουν ερευνηθεί οι μεταφραστικές στρατηγικές από μια ελάσσονα γλώσσα, όπως η ελληνική, προς μείζονες γλώσσες, όπως η αγγλική, η γαλλική και η γερμανική. Στο πλαίσιο της παρούσας εργασίας, αρχικά, μελετάται η θέση της αγγλικής, της γαλλικής, της γερμανικής και της ελληνικής γλώσσας στο παγκόσμιο γλωσσικό σύστημα (Radó, 1987, Song, 1991, Cronin 2003/2007, Παριανού, 2009) και αξιολογούνται οι μεταφράσεις των αρκτικόλεξων και τα ακρωνύμων στα γλωσσικά ζεύγη Ελληνικά-Αγγλικά, Ελληνικά-Γαλλικά και Ελληνικά-Γερμανικά. Κατά την αξιολόγησή τους παρατηρούμε εσφαλμένη απόδοση βάσει της τυπολογίας λαθών DQF-MQM (Klubička et al., 2017/2018, Carl and Báez, 2019, Stasimioti & Sosoni, 2019). Για αυτόν τον λόγο, ανατρέξαμε στη μέθοδο αναγνώρισης ονοματισμένων οντοτήτων (Named Entity Recognition – NER) για να εντοπίσουμε πταίσματα στη μεταφραστική στρατηγική των συστημάτων νευρωνικής μηχανικής μετάφρασης (NMT). Καθορίζουμε, τέλος, τις μεταφραστικές στρατηγικές για όλα τα γλωσσικά ζεύγη με γλώσσα αφετηρίας την ελληνική. Το σώμα κειμένων που χρησιμοποιήθηκε για την παρούσα εργασία αντλήθηκε από την ιστοσελίδα του Τμήματος Ξένων Γλωσσών, Μετάφρασης και Διερμηνείας (ΤΞΓΜΔ) στο Ιόνιο Πανεπιστήμιο, στην Κέρκυρα.

1 Ορισμός μειζόνων και ελασσόνων γλωσσών: Η περίπτωση της αγγλικής, γαλλικής, γερμανικής και της ελληνικής γλώσσας

Σύμφωνα με τον Radó (1987 στο Song, 1991: 63), «μείζων»¹ είναι μια «γλώσσα απεριόριστης διάχυσης» (language of unlimited diffusion – LUD) ή «παγκόσμια γλώσσα» και «ελάσσων» μια «γλώσσα περιορισμένης διάχυσης» (language of limited diffusion – LLD). Για τον Radó (1987 στο Παριανού, 2009: 25), οι LUD περιλαμβάνουν τα Αγγλικά, τα Γαλλικά, τα Γερμανικά, τα Ισπανικά, τα Ιταλικά και τα Ρωσικά, ενώ οι LLD περιλαμβάνουν τα Αραβικά, τα Κινέζικα, τα Ιαπωνικά, τα Κορεάτικα, τα Σουαχίλι, τα Ουγγρικά, τα Τσέχικα κ.λπ. Ο Radó προτείνει μια ακόμη κατηγορία γλωσσών εντός των LLD, τις λεγόμενες «γλώσσες ενδιάμεσης διάχυσης» (languages of intermediate diffusion – LID). Αυτές είναι γλώσσες με περισσότερους από 50 εκατομμύρια εγγενείς ομιλητές, συμπεριλαμβανομένων των

¹ Η απόδοση των όρων “major language” και “minor language” ως «μείζων γλώσσα» και «ελάσσων γλώσσα» αντίστοιχα προτείνεται από τον Παναγιώτη Ι. Κελάνδρια (στο Cronin 2007).

Αραβικών, των Κινέζικων, των Ιαπωνικών, των Κορεάτικων, όπως και των Σουαχίλι, Χίντι, Ουρντού, Μπενγκάλι κ.λπ.

Ο Song (1991: 66) δηλώνει ότι ο αριθμός των ομιλητών της μεγαλύτερης ιστορικής παράδοσης ενός πολιτισμού δεν εξασφαλίζει αντίστοιχο βαθμό διάδοσης της γλώσσας του. Σύμφωνα με τον Song (1991: 67), τα Γερμανικά, τα Γαλλικά, τα Ισπανικά, τα Ιταλικά και τα Ρωσικά είναι LUD, αλλά ο λόγος για την κυριαρχία των γλωσσών αυτός δεν είναι ο αριθμός των ομιλητών τους. Στην πραγματικότητα, για τον Song (στο Parianou, 2009: 26), «η κατάσταση μιας γλώσσας σε παγκόσμια κλίμακα χρησιμεύει ως σημείο καμπής για τον ορισμό της». Σύμφωνα με τον Comrie (1990: vii), ένα από τα πιο προφανή αντικειμενικά κριτήρια για μεγάλες ή μικρές γλώσσες είναι «ο αριθμός των ομιλητών μεμονωμένων γλωσσών, είτε πρόκειται για επίσημες γλώσσες ανεξάρτητων κρατών, είτε χρησιμοποιούνται ευρέως σε περισσότερες από μία χώρες, είτε είναι φορείς μακρόχρονης λογοτεχνικής παράδοσης».

Σε αντίθεση με την πραγματιστική / διαχρονική προσέγγιση του Comrie, η φόρμουλα του Stolz (2001 στο Sherzer & Stolz, 2003: vii) για τον ορισμό των μικρών γλωσσών βασίζεται στον αριθμό των ομιλητών σε σχέση με τον παγκόσμιο πληθυσμό (Matras, 2003: 1 στο Parianou, 2009: 33). Σύμφωνα με την Parianou, (2009: 33), «το κριτήριο διάκρισης του Stolz για μια ελάσσονα γλώσσα φαίνεται να είναι δημογραφικό / στατιστικό»:

[Γ]λώσσες με πληθυσμό ομιλητών πάνω από τον μέσο όρο είναι μείζονες γλώσσες· γλώσσες με πληθυσμό ομιλητών κάτω από τον μέσο όρο είναι ελάσσονες γλώσσες. Παραδείγματα μείζονων γλωσσών είναι τα Κινέζικα, τα Αγγλικά, τα Ισπανικά, τα Χίντι, τα Αραβικά και τα Ρωσικά. Όλες αυτές είναι παγκόσμιες γλώσσες, που χρησιμοποιούνται σε περισσότερες από μία χώρες, και από σημαντικές ομάδες ανθρώπων ως δεύτερη γλώσσα ή *lingua franca*. Το μεγαλύτερο μέρος των σημερινών γλωσσών ανήκει στην κατηγορία των ελασσόνων γλωσσών.

(Stolz, 2001 στο Sherzer & Stolz, 2003: vii)

Ο Cronin, όπως αναφέρει η Parianou (2009: 42–43), «ορίζει τις μείζονες και ελάσσονες γλώσσες πιο ρεαλιστικά και εντοπίζει τρία σημαντικά ζητήματα που αφορούν τη σχέση μεταξύ των δύο κατηγοριών γλωσσών»:

[...] πρώτον, οι πολιτικές, οικονομικές και πολιτισμικές δυνάμεις καθορίζουν τη μείζονα γλώσσα. Αυτές οι δυνάμεις είναι σπάνια στατικές, οπότε όλες οι γλώσσες είναι δυνητικά ελάσσονες γλώσσες. Δεύτερον, οι μεταφραστικές σχέσεις εξαρτώνται από το αν μια γλώσσα είναι *εντατική γλώσσα υποδοχής* (*target-language intensive*) ή *εντατική γλώσσα αφητηρίας* (*source-language intensive*). Τα Αγγλικά είναι εντατική γλώσσα υποδοχής, επειδή υπάρχει έντονη δραστηριότητα μετάφρασης που ξεκινά από τα Αγγλικά προς άλλες γλώσσες (γλώσσες υποδοχής), ενώ οι

ελάσσονες γλώσσες είναι εντατικές γλώσσες αφετηρίας επειδή έχουν μεταφραστική δραστηριότητα από άλλες γλώσσες, κυρίως μείζονες γλώσσες. Τρίτον, οι ελάσσονες γλώσσες εξαρτώνται σε μεγάλο βαθμό από τη μεταφραστική δραστηριότητα, επειδή αναγκάζονται να γίνονται μεταφράσεις στη γλώσσες τους.

(Cronin, 2003: 145 κ.ε.)

Είναι, λοιπόν, ζωτικής σημασίας να κατανοήσουμε τη θέση των ομιλητών ελάσσονων γλωσσών στις σχέσεις γλώσσας και εξουσίας (Pariianou & Kelandrias, 2005: 293 κ.ε.). Η Παριανού (2009: 43) μιλά για την κατάσταση της ελληνικής γλώσσας, ως εξής:

Μιλώντας για την ελληνική γλώσσα, σήμερα είναι μια LLD ή μια λιγότερο ομιλούμενη γλώσσα, για να χρησιμοποιήσουμε τον όρο του Compie που στερείται ιδεολογικών επιπτώσεων. Δεν είναι επίσημη γλώσσα διεθνών οργανισμών όπως του ΟΗΕ (Αραβικά, Κινέζικα, Αγγλικά, Γαλλικά, Ρωσικά και Ισπανικά), του ΝΑΤΟ (Αγγλικά, Γαλλικά), και του ΠΟΕ (Αγγλικά, Γαλλικά, Ισπανικά), αντίθετα με τις LUD όπως είναι τα Αγγλικά ή τα Γαλλικά, και δεν συμβάλλει σε μεγάλο βαθμό στους επιστημονικούς και τεχνικούς τομείς. Ως εκ τούτου, τα Ελληνικά δεν είναι ούτε γλώσσα ευρύτερης επικοινωνίας (όπως τα Αγγλικά) ούτε κύρια μείζονα γλώσσα σε τοπικό επίπεδο (π.χ. όπως τα Γερμανικά ή τα Γαλλικά).

2 Οι συντομορφές στην ελληνική γλώσσα: Η περίπτωση των αρκτικόλεξων και των ακρωνύμων (Ελληνικό Πρότυπο ΕΛΟΤ 1455: 2017)

Στο 4ο Συνέδριο «Ελληνική Γλώσσα και Ορολογία» (Οκτώβριος 2003) παρουσιάστηκε η εργασία «Ελληνικά αρκτικόλεξα και ακρώνυμα υπό το πρίσμα των σύγχρονων επικοινωνιακών αναγκών» (Βαλεοντής, 2003), όπου εκτέθηκε ένα «ΣΧΕΔΙΟ ΚΩΔΙΚΑ ΔΗΜΙΟΥΡΓΙΑΣ ΑΡΚΤΙΚΟΛΕΞΩΝ ΚΑΙ ΑΚΡΩΝΥΜΩΝ (ΚΔΑΑ)». 14 χρόνια αργότερα, το 2017, με εισήγηση της ΕΛΕΤΟ μέσω του Προέδρου της, Κώστα Βαλεοντή, περιλήφθηκε στο πρόγραμμα της τεχνικής επιτροπής ΕΛΟΤ/ΤΕ21 «Ορολογία – Γλωσσικοί πόροι» η σύνταξη – με βάση την εργασία του 2003 και τα βασικά πρότυπα της Ορολογίας που έχει εκπονήσει η επιτροπή αυτή – Σχεδίου Ελληνικού Προτύπου, η υποβολή του σε δημόσιο σχολιασμό με τις διαδικασίες της τυποποίησης και η τελική έγκριση και έκδοσή του.

Στο 12ο Συνέδριο «Ελληνική Γλώσσα και Ορολογία», τον Νοέμβριο του 2019, παρουσιάστηκε και επίσημα το Ελληνικό Πρότυπο ΕΛΟΤ 1455:2017 με τίτλο «Σχηματισμός ελληνικών αρκτικόλεξων και ακρωνύμων» από τον Βαλεοντή (2019). Τα αρκτικόλεξα και τα ακρώνυμα αποτελούν σύμφωνα με τον Βαλεοντή (2019, σ. 60) μέρη της πολυλεκτικής συντομολογίας, η οποία και γράφεται και προφέρεται συντομότερα από την πλήρη μορφή.

Η συντομολογία μπορεί να είναι μονολεκτική, π.χ. προκάτ (προκατασκευασμένος), προχώ (προχωρημένος), ή πολυλεκτική συντομολογία. Τα τρία είδη της τελευταίας είναι:

1. Η **συνοπτική μορφή**, που αποτελείται από λιγότερες λέξεις.

Παράδειγμα 1: Μπορείτε να περιηγηθείτε στην ιστοσελίδα για να αντλήσετε περισσότερες πληροφορίες για το **Τμήμα**, του τρόπους εισαγωγής σε αυτό, τις σπουδές, την έρευνα, τις εκδόσεις καθώς και τα Συνέδρια – Ημερίδες που οργανώνει.

2. Το **ακρώνυμο**, που σχηματίζεται από τα αρχικά γράμματα των συστατικών στοιχείων της πλήρους μορφής ή από συλλαβές της πλήρους μορφής και προφέρεται συλλαβικά.

Παράδειγμα 2: ΕΤΕΠ/ΕΕΔΙΠ/ΕΕΠ

3. Το **αρκτικόλεξο**, που σχηματίζεται από τα αρχικά γράμματα των συστατικών στοιχείων της πλήρους μορφής ή από συλλαβές της πλήρους μορφής και προφέρεται γράμμα–γράμμα με τα ονόματα των γραμμάτων. Τέλος, το αρκτικόλεξο μπορεί να είναι **πληρωνυμικό αρκτικόλεξο**, που προφέρεται με τα πλήρη ονόματα των γραμμάτων του, π.χ. ΑΕ [άλφα-έpsilon] (Ανώνυμη Εταιρεία), ή **λιτωνυμικό αρκτικόλεξο**, που προφέρεται με τα λιτά ονόματα των γραμμάτων του.

Παράδειγμα 3: ΤΞΓΜΔ = Τμήμα Ξένων Γλωσσών, Μετάφρασης και Διερμηνείας [tu-ksu-yu-mu-ðý] (Βαλεοντής, 2019, σ. 66)

Παράδειγμα 4: Στις 3 Δεκεμβρίου 2016 το Τμήμα Ξένων Γλωσσών, Μετάφρασης και Διερμηνείας διοργάνωσε σε συνεργασία με την **Ελληνική Εταιρεία Μεταφρασεολογίας (ΕΕμ)** την 1η Ημερίδα Νέων Ερευνητών με θέμα «Μετάφραση και Μεταφρασεολογία στον ελληνόφωνο χώρο».

Στη συνέχεια, θα μελετήσουμε τις αποδόσεις των παραπάνω παραδειγμάτων στην αγγλική, γαλλική και γερμανική γλώσσα με τη βοήθεια των συστημάτων νευρωνικής μετάφρασης και με τη μέθοδο αναγνώρισης ονοματισμένων οντοτήτων (Named Entity Recognition – NER).

3 Μετάφραση ακρωνύμων και αρκτικόλεξων με συστήματα νευρωνικής μετάφρασης

Σήμερα, η μηχανική μετάφραση (machine translation – MT) χρησιμοποιείται περισσότερο από ποτέ, σε πληθώρα εφαρμογών: αύξηση παραγωγικότητας των μεταφραστών, διεθνοποίηση ιστοτόπων και υπηρεσιών, βελτίωση μοντέλων φυσικής γλώσσας κ.ά.

Παρ' όλη την αλματώδη ανάπτυξη που έχει δει τόσο ο κλάδος των μεταφραστικών υπηρεσιών χάρη στην παγκοσμιοποίηση (Cronin, 2003/2007) όσο και η γλωσσική τεχνολογία χάρη, κυρίως, στα νευρωνικά δίκτυα κατά τα τελευταία 5–6 χρόνια, οι αδυναμίες των συστημάτων αυτών είναι εμφανείς ακόμα και στον μη ειδικό, επαγγελματία μεταφραστή.

Θα εστιάσουμε στις αποδόσεις των ακρωνύμων, αναλύοντας συγκεκριμένα παραδείγματα και αξιολογώντας τες με βάση την τυπολογία λαθών DQF-MQM. Χρησιμοποιήσαμε τις διεπαφές (API) της Google, της Microsoft και της Ευρωπαϊκής Ένωσης (eTranslation)² που παρέχονται για μετάφραση, ώστε να μεταφράσουμε αποσπάσματα από τον ελληνικό ιστότοπο του ΤΞΓΜΔ στα Αγγλικά, τα Γαλλικά και τα Γερμανικά. Τα αποτελέσματα, που θα αναλύσουμε παρακάτω, εμφανίζουν ιδιαίτερο ενδιαφέρον:

Ελληνικά (EL)	Ειδικό Τεχνικό Εργαστηριακό Προσωπικό (ΕΤΕΠ)	Ειδικό Εργαστηριακό Διδακτικό Προσωπικό (ΕΕΔΙΠ)	Ειδικό Εκπαιδευτικό Προσωπικό (ΕΕΠ)
Google (EN)	Special Technical Laboratory Staff (EIB)	Laboratory Teaching Staff (EEDIP)	Special Teaching Staff (SPT)
Microsoft (EN)	Special Technical Laboratory Personnel (EEP)	Special Laboratory Teaching Staff (EDIP)	Special Educational Staff (SEP)
eTranslation (EN)	Special Technical Laboratory Staff (ERD)	Special Laboratory Teaching Personnel (SPD)	Special Educational Personnel (SPD)
Google (DE)	Spezielles Technisches Laborpersonal (EIB)	Lehrkräfte für Sonderlaboratorien (EEDIP)	Sonderlehrkräfte (SPT)
Microsoft (DE)	Spezielles technisches Laborpersonal (EΤΕΠ)	Spezielles Lehrpersonal im Labor (EPPD)	Sonderschulpersonal (DPP)
eTranslation (DE)	Spezielles technisches Laborpersonal (ERD)	Speziallaborpädagogik (SPD)	Sonderpädagogik (SPD)
Google (FR)	Personnel de laboratoire technique spécial (BEI)	Enseignants Spécialisés de Laboratoire (EEDIP)	Personnel enseignant spécial (SPT)
Microsoft (FR)	Personnel de laboratoire technique spécial (EΤΕΠ)	Personnel enseignant spécial en laboratoire (DOCUP)	Personnel enseignant spécial (DOCUP)
eTranslation (FR)	Personnel de laboratoire technique spécial (RED)	Personnel enseignant spécial de laboratoire (DSP)	Personnel de l'enseignement spécial (DSP)

Ας δούμε και δύο περιπτώσεις με συγκείμενο³:

² Διαθέσιμα στο <https://cloud.google.com/natural-language>, <https://azure.microsoft.com/en-us/services/cognitive-services/text-analytics/> και <https://webgate.ec.europa.eu/etranslation>, αντίστοιχα [τελευταία πρόσβαση 29 Ιουλίου 2021].

³ Για οικονομία χώρου, δεν παραθέτουμε την απόδοση ολόκληρης της περιόδου αλλά μόνο των εξεταζόμενων μεταφραστικών μονάδων.

Ελληνικά (EL)	Τμήμα	Ελληνική Εταιρεία Μεταφρασεολογίας (ΕΕμ)
Google (EN)	Department	Hellenic Society of Translation Studies (EEM)
Microsoft (EN)	Department	Hellenic Society of Translationology (EEM)
eTranslation (EN)	Department	Hellenic Society of Translations
Google (DE)	Abteilung	Hellenic Society of Translation Studies (EEM)
Microsoft (DE)	Abteilung	Griechische Übersetzungsgesellschaft (EEM)
eTranslation (DE)	Abteilung	Griechische Gesellschaft für Übersetzungen
Google (FR)	Département	Société hellénique de traductologie (EEM)
Microsoft (FR)	Département	Société hellénique de traduction (EEM)
eTranslation (FR)	Département	Société hellénique des traductions

Πρόταση 1	Μπορείτε να περιηγηθείτε στην ιστοσελίδα για να αντλήσετε περισσότερες πληροφορίες για το Τμήμα , τους τρόπους εισαγωγής σε αυτό, τις σπουδές, την έρευνα, τις εκδόσεις καθώς και τα Συνέδρια – Ημερίδες που οργανώνει.
Πρόταση 2	Στις 3 Δεκεμβρίου 2016 το Τμήμα Ξένων Γλωσσών, Μετάφρασης και Διερμηνείας διοργάνωσε σε συνεργασία με την Ελληνική Εταιρεία Μεταφρασεολογίας (ΕΕμ) την 1η Ημερίδα Νέων Ερευνητών με θέμα «Μετάφραση και Μεταφρασεολογία στον ελληνόφωνο χώρο».

Εκ πρώτης όψεως διακρίνουμε ότι τα αποτελέσματα διαφέρουν όχι μόνο ανάμεσα στις διεπαφές, αλλά και ανάμεσα στις γλώσσες της κάθε διεπαφής. Ακόμα και όταν τα συστήματα αποδίδουν σωστά τις λέξεις, παρατηρούμε δύο ειδών στρατηγικές στη μεταχείριση των ακρωνύμων:

- A. Μεταγραφή με λατινικούς χαρακτήρες
- B. «Μετάφραση» στη γλώσσα υποδοχής

Η στρατηγική A. είναι συνήθης, ειδικά σε περιπτώσεις όπου η πλήρης μορφή δεν συναντά αντιστοιχία στη γλώσσα υποδοχής (π.χ. ΕΛ.ΑΣ. → HE.PO.) ούτε «αξίζει» να αποδοθεί με νέο ακρώνυμο (π.χ. Πανελλήνια Ένωση Μεταφραστών – ΠΕΜ → ΡΕΜ → AGT). Παρ' όλα αυτά, το ενδιαφέρον μάς κεντρίζει η στρατηγική B. και τούτο διότι σε καμία απόδοση δεν συνάδει το ακρώνυμο με τα αρχικά των λέξεων που μεταφράστηκαν.

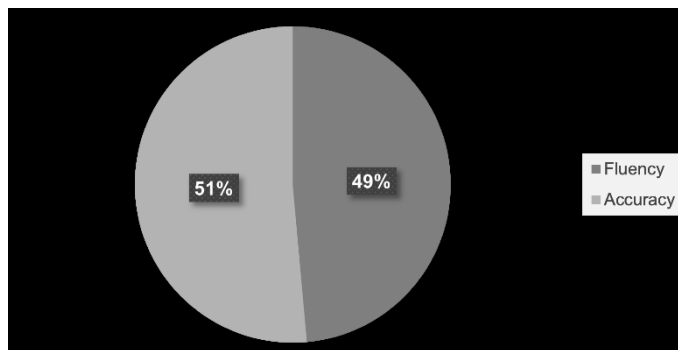
Συγκεκριμένα, η Google και το eTranslation έχουν την τάση να διατηρούν το ίδιο ακρώνυμο

ανεξαρτήτως γλώσσας υποδοχής, ενώ η Microsoft φαίνεται να μεταχειρίζεται το ακρώνυμο σε κάθε γλώσσα ξεχωριστά. Τη μόνη εξαίρεση σε αυτή την παρατήρηση αποτελεί η *ΕΕμ* η οποία αποδίδεται ως *EEM* (με λατινικά) σε όλες τις γλώσσες και από όλες τις διεπαφές.

4 Αξιολόγηση της μετάφρασης ακρωνύμων με συστήματα νευρωνικής μετάφρασης

Εύλογα θα αναρωτηθεί κανείς: Πώς θα καταλήξουμε σε ορθό ακρώνυμο, εάν πρώτα η ίδια η απόδοση της πλήρους μορφής δεν είναι ορθή; Μπορούμε να αξιολογήσουμε τα ίδια τα αποτελέσματα, με τη βοήθεια της τυπολογίας λαθών DQF-MQM (Klubička et al., 2017 και 2018, Carl and Bález, 2019, Stasimioti & Sosoni, 2019).

Όπως φαίνεται στο Γράφημα 1, τα μισά λάθη της μηχανικής μετάφρασης αφορούν σε ασυνέπεια (inconsistency) μεταξύ της ορθής μεν απόδοσης της πλήρους μορφής αλλά και της εσφαλμένης απόδοσης του ακρωνύμου της. Αντίθετα, τα άλλα μισά αφορούν σε εσφαλμένη μετάφραση (mistranslation) της πλήρους μορφής μερικώς ή εντελώς.



Γράφημα 1. Η κατανομή των λαθών βάσει της τυπολογίας DQF-MQM.

Η μόνη περίπτωση η οποία δεν εμπίπτει σε λάθος είναι αυτή της συνοπτικής μορφής (π.χ. *Τμήμα* αντί για *Τμήμα Ξένων Γλωσσών, Μετάφρασης και Διερμηνείας* ή *ΤΞΓΜΔ*). Θα αναφερθούμε στην ενότητα 7 για το πώς προτείνουμε να διαχειριζόμαστε τέτοιες περιπτώσεις.

5 Ο πιθανός ρόλος της Αναγνώρισης ονοματισμένων οντοτήτων (NER)

Δυστυχώς, δεν μπορούμε να γνωρίζουμε και, ως εκ τούτου, να αξιολογήσουμε το πώς οι διάφοροι αλγόριθμοι καταλήγουν στο αποτέλεσμα που εμφανίζεται σε εμάς ως «μετάφραση». Η Google και η Microsoft, καθώς παρέχουν και υπηρεσίες αναζήτησης ιστού (Google Search και Microsoft Bing αντίστοιχα), βασίζονται στη μέθοδο αναγνώρισης

ονοματισμένων οντοτήτων (Named Entity Recognition – NER) για καλύτερα αποτελέσματα αναζήτησης⁴. Υποψιαζόμαστε, λοιπόν, ότι χρησιμοποιούν τη μέθοδο αυτή και στις διεπαφές μετάφρασης που παρέχουν. Φυσικά, αυτό δεν ισχύει για την περίπτωση του eTranslation.

Ας δοκιμάσουμε, λοιπόν, να δούμε εάν και κατά πόσον οι διεπαφές τους αναγνωρίζουν τις οντότητες που μεταφράζουν:

	Τμήμα Ξένων Γλωσσών, Μετάφρασης και Διερμηνείας	Ελληνική Εταιρεία Μεταφρασεολογίας	ΕΕμ
Google (EN)	43% (ORG)	81% (ORG)	74% (ORG)
Microsoft (EN)	95% (ORG)	98% (ORG)	76% (ORG)
Google (DE)	44% (ORG)	80% (ORG)	20% (LOC)
Microsoft (DE)	66% (ORG)	93% (ORG)	85% (ORG)
Google (FR)	42% (ORG)	83% (ORG)	60% (ORG)
Microsoft (FR)	78% (ORG)	99% (ORG)	89% (ORG)

Σε όλες τις περιπτώσεις και ανεξαρτήτως γλώσσας, τα παραδείγματα αυτά αναγνωρίζονται ως οργανισμοί (ORG). Μόνη εξαίρεση αποτελεί το αρκτικόλεξο EEM στη γερμανική μετάφραση της Google όπου αναγνωρίζεται ως... τοποθεσία (LOC)! Με μια γρήγορη αναζήτηση, ωστόσο, διαπιστώνουμε ότι το μοντέλο δεν κάνει λάθος, αφού υπάρχει ο ποταμός Eem στην Ολλανδία⁵.

6 Καθορισμός μεταφραστικών στρατηγικών για τα αρκτικόλεξα

Ας εξετάσουμε τις μεταφραστικές στρατηγικές που ακολουθούμε όταν μεταφράζουμε βραχυγραφίες από την ελληνική προς την αγγλική, γαλλική και γερμανική γλώσσα. Οι στρατηγικές αυτές εφαρμόζονται στις πρωτότυπες ελληνικές βραχυγραφίες.

Μετάφραση του ελληνικού όρου χωρίς διατήρηση του ελληνικού/ξενόγλωσσου αρκτικόλεξου

Η επιλογή μη διατήρησης του αρκτικόλεξου οφείλεται, συχνά, στο ότι η μετάφρασή της

⁴ Δείτε την τεκμηρίωση (documentation) της Microsoft και της Google για το NER στα Αγγλικά: <https://docs.microsoft.com/en-us/azure/cognitive-services/text-analytics/named-entity-types?tabs=general> και <https://cloud.google.com/natural-language/docs/analyzing-entities> αντίστοιχα [τελευταία πρόσβαση 29 Ιουλίου 2021]. Η Google, επίσης, εισήγαγε το «Γράφημα γνώσης» (Knowledge Graph) το 2012 για «να καταλαβαίνει πράγματα κι όχι συμβολοσειρές» (<https://blog.google/products/search/introducing-knowledge-graph-things-not/>) [τελευταία πρόσβαση 29 Ιουλίου 2021].

⁵ <https://de.wikipedia.org/wiki/Eem> [τελευταία πρόσβαση 29 Ιουλίου 2021].

δυσχεραίνει την ανάγνωση της ως αρκτικόλεξο/ακρωνύμιο. Συγκεκριμένα, το παραγόμενο αρκτικόλεξο/ακρωνύμιο στην ξένη γλώσσα ή και ακόμα το ίδιο το πρωτότυπο ακρωνύμιο ενδέχεται να μην είναι κατανοητό από τον αναγνώστη, οπότε να μην πληροί και την αρχή της διαφάνειας στον σχηματισμό ορολογίας, καθώς «η μορφολογία του όρου θα πρέπει να αποκαλύπτει, ως ένα βαθμό, την έννοια στην οποία αντιστοιχεί» (Κεραμίδης 2015: 45). Στην περίπτωση της *Ελληνικής Εταιρείας Μεταφρασεολογίας (ΕΕμ)* παρατηρούμε μέσω του NER στην προηγούμενη υποενοότητα ότι το λιτωναμικό αρκτικόλεξο *ΕΕμ* δεν γίνεται κατανοητό μόνο ως οργανισμός αλλά και ως τοποθεσία, οπότε δεν πληροί την αρχή της διαφάνειας.

Ελληνικά	Αγγλικά	Γαλλικά	Γερμανικά
Ελληνική Εταιρεία Μεταφρασεολογίας (ΕΕμ)	Hellenic Society of Translation Studies	Société hellénique des études de traduction	Hellenische Gesellschaft für Translationswissenschaft

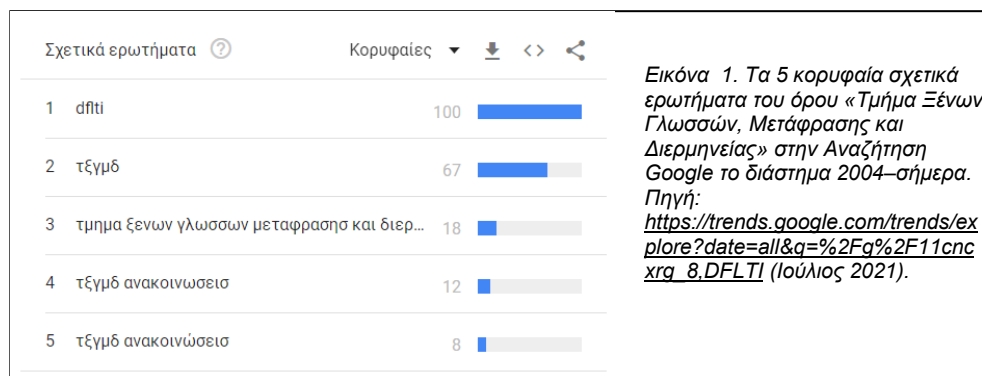
Μετάφραση του ελληνικού όρου και του ελληνικού αρκτικόλεξου

Πρόκειται για ακόμα μια στρατηγική, η οποία εφαρμόζεται σε περιπτώσεις που η μετάφραση του όρου και του αρκτικόλεξου «επιβάλλονται από τη λειτουργία του μεταφράσματος στο κοινό-στόχος, το οποίο είναι περιορισμένο και συγκεκριμένο, άρα προσδιορισίμο, με σαφώς καθορισμένες προσδοκίες από το κείμενο» (Κελάνδριος, 2007: 107). Στην περίπτωση του λιτωναμικού αρκτικόλεξου *ΤΞΓΜΔ*, παρατηρούμε ότι η αγγλική του απόδοση είναι κατανοητή από τον ξενόγλωσσο αναγνώστη. Και επειδή η αγγλική γλώσσα κυριαρχεί στο παγκόσμιο γλωσσικό σύστημα ως *lingua franca*, θα προτιμήσουμε τη διατήρηση της αγγλικής βραχυγραφίας με τη συνοδεία του μεταφρασμένου όρου στην εκάστοτε γλώσσα.

Ελληνικά	Αγγλικά	Γαλλικά	Γερμανικά ⁶
Τμήμα Ξένων Γλωσσών, Μετάφρασης και Διερμηνείας (ΤΞΓΜΔ)	Department of Foreign Languages, Translation and Interpreting (DFLTI)	Département de langues étrangères, traduction et interprétation (DFLTI)	Abteilung für Fremdsprachen, Übersetzen und Dolmetschen (DFLTI)

Ενδιαφέρον παρουσιάζει, επίσης, πως το αρκτικόλεξο DFLTI έρχεται πρώτο στις Τάσεις της Αναζήτησης Google όταν πρόκειται για ερωτήματα (queries) σχετικά με τον όρο «Τμήμα Ξένων Γλωσσών, Μετάφρασης και Διερμηνείας» στο διάστημα 2004–σήμερα:

⁶ Για την απόδοση του Τμήματος Ξένων Γλωσσών, Μετάφρασης και Διερμηνείας στη γερμανική γλώσσα και την αιτιολόγησή του βλ. Αδάμ κ.ά. (2019).



Αντικατάσταση της ελληνικής συνοπτικής μορφής με ξενόγλωσση βραχυγραφία

Η επιλογή της αντικατάστασης της συνοπτικής μορφής με αρκτικόλεξο γίνεται πρώτα απ' όλα προς τήρηση του κανόνα της συνέπειας, όπου «κάθε νέος όρος θα πρέπει να εντάσσεται στο προϋπάρχον σύστημα όρων του θεματικού πεδίου στο οποίο ανήκει και να είναι συνεπής ως προς αυτό» (Κεραμίδης, 2015: 45). Στην περίπτωση του όρου *Τμήμα* επιλέγουμε τη χρήση της αγγλικής βραχυγραφίας *DFLTI* σε όλες τις γλώσσες εργασίας λόγω της κυριαρχίας της αγγλικής γλώσσας στο παγκόσμιο γλωσσικό σύστημα ως *lingua franca*. Επίσης, με μια πιο πρακτική ματιά, ο ισότοπός του συμβάλλει σε αυτή την «τυποποίηση» του αρκτικόλεξου: <https://dfhti.ionio.gr/>.

Τμήμα (Ξένων Γλωσσών, Μετάφρασης και Διερμηνείας) → *DFLTI*
(αγγλικά, γαλλικά, γερμανικά)

7 Συμπέρασμα

Ως προς τα εξαγόμενα συμπεράσματα, σε αντίθεση με έρευνες στο γλωσσικό ζεύγος Αγγλικά-Ελληνικά (Castilho et al., 2017a, Castilho et al., 2017b, Stasimioti & Sosoni, 2019) που πιστοποιούν την υψηλή ποιότητα της νευρωνικής μηχανικής μετάφρασης, παρατηρούμε κατά τη διαδικασία της αξιολόγησης ότι στην περίπτωση των ακρωνύμων και αρκτικόλεξων η μηχανική μετάφραση στα γλωσσικά ζεύγη Ελληνικά-Αγγλικά, Ελληνικά-Γαλλικά και Ελληνικά-Γερμανικά δεν φέρει υψηλής ποιότητας αποδόσεις. Ευελπιστούμε ότι οι προτεινόμενες μεταφραστικές στρατηγικές για την περίπτωση των αρκτικόλεξων και της συνοπτικής μορφής θα αποτελέσουν αφορμή για περαιτέρω έρευνα πάνω στο θεματικό πεδίο της μετάφρασης των βραχυγραφιών.

Βιβλιογραφία

Ξενόγλωσση

- Carl, M., & Toledo Báez, M. C. (2019). "Machine Translation Errors and the Translation Process: A Study across Different Languages". *Journal of Specialised Translation*, 31, pp. 107–132.
- Castilho, S., Moorkens, J., Gaspari, F., Calixto, I., Tinsley, J., & Way, A. (2017a). "Is neural machine translation the new state of the art?" *Prague Bulletin of Mathematical Linguistics* 108, pp.109–120.
- Castilho, S., Moorkens, J., Gaspari, F., Sennrich, R., Sosoni, V., Georgakopoulou, Y., Lohar, P., Way, A., Barone, A. M., & Gialama, M. (2017b). "A Comparative quality evaluation of PBSMT and NMT using professional translators". In *Proceedings of Machine Translation Summit XVI vol.1: Research Track*. pp. 116–131.
- Comrie, B. (1990). *The World's Major Languages*. New York: Oxford University Press.
- Cronin, M. (2003). *Translation and Globalization*. London/New York: Routledge.
- Klubička, F., Toral, A., & Sánchez-Cartagena, V. M. (2017). "Fine-grained human evaluation of neural versus phrase-based machine translation". *Prague Bulletin of Mathematical Linguistics*, 108, pp. 121–132.
- Klubička, F., Toral, A., & Sánchez-Cartagena, V. M. (2018). "Quantitative fine-grained human evaluation of machine translation systems: a case study on English to Croatian". *Machine Translation*, 32(3) , pp. 195–215.
- Matras, Y. (2003). "Defining typical features of minor languages: With reference to Domari, a minor language of the Near East". In: Sherzer, Joel & Stolz, Thomas (eds.) *Minor Languages. Approaches, Definitions, Controversies. Papers from the conference on 'Minor Languages: Coming to grips with a suitable definition', Bremen, June 2001*. Bochum: Univeristätsverlag Dr. N. Brockmeyer, pp. 1–14.
- Parianou, A. (2009). *Translating from major into minor languages*. Athens: Diavlos Publishing Co.
- Radó, G. (1987). "A typology of LLD problems". *Babel* 33(1) , pp. 6–13.
- Sherzer, J. & Stolz, T. (eds.) (2003). *Minor Languages. Approaches, Definitions, Controversies. Papers from the conference on 'Minor Languages: Coming to grips with a suitable definition', Bremen, June 2001*. Bochum: Univeristätsverlag Dr. N. Brockmeyer,
- Song, Y.-I. (1991). "Remarks on Cultural Transfer from an LLD". In: *TTR* 4:1. pp. 63–79.
- Stolz, T. (2001). "Minor languages and general linguistics (with special focus on Europe)". In: Stolz, T. (ed.). *Minor languages of Europe. A series of lectures at the University of Bremen, April-July 2000*. Bochum: Brockmeyer. pp. 211–242.
- Stasimioti, M. & Sosoni, V. (2019). "MT output and post-editing effort: Insights from a comparative analysis of SMT and NMT output for the English to Greek language pair and implications for the training of post-editors." In C. Szabó & R. Besznyák (eds) *Teaching Specialised Translation and Interpreting in a Digital Age – Fit-For-Market Technologies, Schemes and Initiatives*. Wilmington: Vernon Press.

Ελληνική

Αδάμ, Ε., Βαλσαμή, Β., Γιαννούτσος, Χ., Μπουμπάρης, Δ. (2019). «Σχεδιασμός και κατάρτιση ελληνογερμανικού γλωσσάριου για ακαδημαϊκές ιστοσελίδες: Η περίπτωση του ΤΞΓΜΔ». *Ελληνική Γλώσσα και Ορολογία. Ανακοινώσεις 12^{ου} Συνεδρίου*. Αθήνα: Τεχνικό Επιμελητήριο Ελλάδος. 275–288. [τελευταία πρόσβαση: 23 Ιουλίου 2021]. Διαθέσιμο στο: http://www.eleto.gr/download/Conferences/12th%20Conference/Papers-and-speakers/12th_19-05-07_ValsamiV-GiannoutsosCh-BoumparisD-AdamE_Paper2_V04.pdf.

Cronin, M. (2007). *Μετάφραση και Παγκοσμιοποίηση* (Μτφ. Παναγιώτης Ι. Κελάνδριας). Αθήνα: Δίαυλος.

Βαλεοντής, Κ. (2003). «Ελληνικά αρκτικόλεξα και ακρώνυμα υπό το πρίσμα των σύγχρονων επικοινωνιακών αναγκών». *Ελληνική Γλώσσα και Ορολογία. Ανακοινώσεις 4^{ου} Συνεδρίου*. Αθήνα: Τεχνικό Επιμελητήριο Ελλάδος. 1–19. [τελευταία πρόσβαση: 23 Ιουλίου 2021]. Διαθέσιμο στο: http://www.eleto.gr/download/Conferences/4th%20Conference/4th_12-39-ValeontisKostas.pdf

Βαλεοντής, Κ. (2019). «Οι συντομομορφές στην ελληνική γλώσσα – Σχηματισμός ελληνικών αρκτικόλεξων και ακρωνύμων (Ελληνικό Πρότυπο ΕΛΟΤ 1455: 2017)». *Ελληνική Γλώσσα και Ορολογία. Ανακοινώσεις 12^{ου} Συνεδρίου*. Αθήνα: Τεχνικό Επιμελητήριο Ελλάδος. σ. 57–70 [τελευταία πρόσβαση: 23 Ιουλίου 2021]. Διαθέσιμο στο: http://www.eleto.gr/download/Conferences/12th%20Conference/Papers-and-speakers/12th_02-26-23_ValeontisKostas_Paper_V05.pdf

Κελάνδριας, Π. Ι. (2007). *Η Μετάφραση των Οικονομικών Κειμένων: Μια λειτουργική προσέγγιση*. Αθήνα: Δίαυλος.

Κεραμίδας, Σ. Γ. (2015). *Ζητήματα Ορολογίας στην Τεχνική Μετάφραση*. Αθήνα: Δίαυλος.

Παριανού, Α. (2004). «Διαπολιτισμική ειδική επικοινωνία». *Επιστημονική Επετηρίδα 1999–2003*. Ιόνιο Πανεπιστήμιο: Κέρκυρα. σ. 299–424.

Δημήτρης Μπουμπάρης

Πανεπιστήμιο της Αμβέρσας (University of Antwerp)
Ηλ-ταχ.: dimboump@pm.me

Χρήστος Γιαννούτσος

Ιόνιο Πανεπιστήμιο
Τμήμα Ξένων Γλωσσών, Μετάφρασης και Διερμηνείας (ΤΞΓΜΔ)
Ηλ-ταχ.: giachr1997@gmail.com