



ΙΟΝΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ
ΣΧΟΛΗ ΙΣΤΟΡΙΑΣ ΚΑΙ ΜΕΤΑΦΡΑΣΗΣ – ΔΙΕΡΜΗΝΕΙΑΣ
ΤΜΗΜΑ ΞΕΝΩΝ ΓΛΩΣΣΩΝ, ΜΕΤΑΦΡΑΣΗΣ ΚΑΙ ΔΙΕΡΜΗΝΕΙΑΣ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ
ΠΑΡΑΣΚΕΥΗ ΠΑΠΑΚΩΣΤΑ
ΑΡ. ΜΗΤΡΩΟΥ: fti.st2117

Αυτόματη και ανθρώπινη αξιολόγηση ενός μεταφρασμένου κειμένου που έχει παραχθεί από δύο διαφορετικά συστήματα νευρωνικής μηχανικής μετάφρασης για τα ζεύγη γλωσσών Αγγλικά-Ελληνικά και Αγγλικά-Ισπανικά

ΜΕΛΗ ΕΠΙΤΡΟΠΗΣ ΕΠΟΠΤΕΙΑΣ:
ΒΙΛΕΛΜΙΝΗ ΣΩΣΩΝΗ – Α' ΕΠΙΒΛΕΠΟΥΣΑ
ΚΑΤΙΑ ΛΗΔΑ ΚΕΡΜΑΝΙΔΟΥ – Β' ΕΠΙΒΛΕΠΟΥΣΑ
ΜΑΡΙΑ ΤΣΙΓΚΟΥ – Γ' ΕΠΙΒΛΕΠΟΥΣΑ

ΚΕΡΚΥΡΑ
ΦΕΒΡΟΥΑΡΙΟΣ 2024

Αυτόματη και ανθρώπινη αξιολόγηση ενός μεταφρασμένου κειμένου που έχει παραχθεί από δύο διαφορετικά συστήματα νευρωνικής μηχανικής μετάφρασης για τα ζεύγη γλωσσών Αγγλικά-Ελληνικά και Αγγλικά-Ισπανικά

Περιεχόμενα

Περίληψη	6
Abstract	6
1. Εισαγωγή	6
2. Μηχανική Μετάφραση	7
3. Σύντομη ιστορική αναδρομή	7
4. Συστήματα Μηχανικής Μετάφρασης	9
i. Συστήματα Μηχανικής Μετάφρασης Βάσει Κανόνων (Rule-Based Machine Translation, RBMT)	10
ii. Συστήματα Μηχανικής Μετάφρασης Βάσει Παραδειγμάτων (Example-Based Machine Translation, EBMT)	10
iii. Στατιστικά Συστήματα Μηχανικής Μετάφρασης (Statistical Machine Translation, SMT)	10
iv. Υβριδικά Συστήματα Μηχανικής Μετάφρασης (Hybrid Machine Translation)	10
v. Νευρωνικά Συστήματα Μηχανικής Μετάφρασης (Neural Machine Translation, NMT)	11
o Google Translate	12
o DeepL	12
5. Ποιότητα της Μετάφρασης	13
6. Προσεγγίσεις αξιολόγησης της ποιότητας της μετάφρασης	15
o Εμπειρική προσέγγιση (Experience-based approach)	15
o Θεωρητική και ερευνητική προσέγγιση (Theoretical and research-based approach)	15
o Προσέγγιση βάσει της αντίδρασης του αναγνώστη (Reader-response approach)	15
o Κειμενική και πραγματολογική προσέγγιση (Textual and pragmatic approach)	16
o Λειτουργική προσέγγιση (Functional – Componential approach)	16
7. Ποιότητα Μηχανικής Μετάφρασης	16
8. Μέθοδοι αξιολόγησης της ποιότητας της Μηχανικής Μετάφρασης	17
i. Ακρίβεια και Ευχέρεια	17
ii. Κατάταξη	18
iii. Κατηγοριοποίηση των λαθών	18
o Πρότυπο MQM	18
o Πρότυπο DQF	20
o Εναρμονισμένο πρότυπο MQM-DQF	20
9. Inter-annotator agreement	21

10. Αυτόματη Αξιολόγηση της ποιότητας της Μηχανικής Μετάφρασης	21
11. Πλατφόρμα αυτόματης αξιολόγησης MutNMT	26
12. Μεθοδολογία της έρευνας	26
13. Αποτελέσματα της έρευνας	28
i. Ανθρώπινη αξιολόγηση	28
ii. Αυτόματη αξιολόγηση	32
14. Συζήτηση αποτελεσμάτων	34
15. Επίλογος	36
16. Παράρτημα	38
17. Βιβλιογραφικές Αναφορές	44

Ευχαριστίες

Με την ολοκλήρωση της μεταπτυχιακής διπλωματικής μου εργασίας, θα ήθελα να εκφράσω τις θερμές μου ευχαριστίες σε όλους όσους συνέβαλλαν στην εκπόνησή της.

Ευχαριστώ θερμά την επιβλέπουσα καθηγήτριά μου κ. Βιλελμίνη Σωσώνη (Αναπληρώτρια Καθηγήτρια στο Τμήμα Ξένων Γλωσσών, Μετάφρασης και Διερμηνείας στο Ιόνιο Πανεπιστήμιο) για την καθοδήγηση και συνεχή υποστήριξη.

Επίσης, θα ήθελα να ευχαριστήσω την κ. Δέσποινα Μουρατίδη (Μεταδιδάκτορας Ερευνήτρια του Εργαστηρίου Πληροφορικής στις Ανθρωπιστικές και Κοινωνικές Επιστήμες στο Τμήμα Πληροφορικής στο Ιόνιο Πανεπιστήμιο) για τη συμβολή της και τον πολύτιμο χρόνο που διέθεσε στον υπολογισμό της συμφωνίας των επισημειωτών κατά την καταγραφή των αποτελεσμάτων της έρευνας.

Οφείλω επίσης ένα μεγάλο ευχαριστώ σε όλους εκείνους που συνέβαλαν πρακτικά στην ολοκλήρωση της εργασίας μου. Συγκεκριμένα, ευχαριστώ τις μεταφράστριες Κωνσταντίνα Κοντογιάννη Loayza και Μαρία Δέσποινα Παπαγεωργίου για τη μετάφραση του κειμένου-πηγή στα Ισπανικά και Ελληνικά αντίστοιχα, την καλή μου φίλη και συμφοιτήτρια Χριστίνα Δεληπαλτίδου για τη συμβολή της στην εξεύρεση επισημειωτών και τη συνεχή υποστήριξή της, καθώς και όλους όσους συμμετείχαν στην έρευνα συμπληρώνοντας τα ερωτηματολόγια.

Τέλος, θα ήθελα εκφράσω την ευγνωμοσύνη μου στην οικογένεια και τους φίλους μου για την ηθική υποστήριξη σε όλο το διάστημα των σπουδών μου.

Περίληψη

Η παρούσα διπλωματική εργασία παρουσιάζει μια έρευνα σχετικά με την αυτόματη και ανθρώπινη αξιολόγηση ενός μεταφρασμένου κειμένου που παράγεται από δύο διαφορετικά συστήματα νευρωνικής μηχανικής μετάφρασης, το DeepL και το Google Translate, για τα ζεύγη γλωσσών Αγγλικά-Ελληνικά και Αγγλικά-Ισπανικά. Ξεκινώντας με μια περιεκτική επισκόπηση των συστημάτων αυτόματης μετάφρασης, παρακολουθώντας την εξέλιξη τους από τις προσεγγίσεις που βασίζονται σε κανόνες και τις στατιστικές προσεγγίσεις έως την εμφάνιση της νευρωνικής μηχανικής μετάφρασης, η εστίαση επεκτείνεται στις μεθοδολογίες αξιολόγησης της ποιότητας της μετάφρασης, που περιλαμβάνουν την ανθρώπινη αξιολόγηση και τις αυτόματες μετρικές, συμπεριλαμβανομένης μιας αναφοράς στο MutNMT, μια πλατφόρμα αυτόματης αξιολόγησης προσαρμοσμένη για αξιολογήσεις μηχανικής μετάφρασης. Καθ' όλο το δεύτερο μέρος της εργασίας, παρουσιάζονται λεπτομερώς η μεθοδολογία και τα ευρήματα της έρευνας που διεξήχθη, τα οποία περιλαμβάνουν αξιολογήσεις των νευρωνικών συστημάτων μηχανικής μετάφρασης DeepL και Google Translate και παρέχεται μια αποχρωματισμένη κατανόηση της ποιότητας της μετάφρασης σε διάφορα γλωσσικά πλαίσια με την αντιπαραβολή των ανθρώπινων αξιολογήσεων με τις αυτοματοποιημένες μετρικές. Μέσω της εκτεταμένης ανάλυσης και συζήτησης των αποτελεσμάτων, η παρούσα διπλωματική εργασία συνεισφέρει αξιόλογες πληροφορίες στην τρέχουσα κατάσταση και τις μελλοντικές κατευθύνσεις της έρευνας για τη μηχανική μετάφραση.

Abstract

This dissertation presents an investigation into the automatic and human evaluation of a translated text produced by two distinct neural machine translation systems, DeepL and Google Translate, for the English-Greek and English-Spanish language pairs. Commencing with a comprehensive overview of machine translation systems, tracing their evolution from rule-based and statistical approaches to the emergence of neural machine translation, the focus extends to methodologies for assessing translation quality, encompassing human evaluation and automatic metrics, including a mention of MutNMT, an automated evaluation platform tailored for machine translation assessments. Throughout the second part of the dissertation, the research methodology and findings are detailed, which include evaluations of the MT neural systems DeepL and Google Translate and a nuanced understanding of translation quality in diverse linguistic contexts is provided by juxtaposing human assessments with automated metrics. Through extensive analysis and discussion of results, this dissertation contributes valuable insights into the current state and future directions of machine translation research.

1.Εισαγωγή

Η παρούσα εργασία εκπονείται στα πλαίσια του προγράμματος μεταπτυχιακών σπουδών «Επιστήμη της Μετάφρασης» και μελετά την αυτόματη και ανθρώπινη αξιολόγηση ενός μεταφρασμένου κειμένου που έχει παραχθεί από δύο διαφορετικά συστήματα νευρωνικής μηχανικής μετάφρασης, το DeepL και το Google Translate, για τα ζεύγη γλωσσών Αγγλικά-Ελληνικά και Αγγλικά-Ισπανικά. Στόχος είναι η σύγκριση των δύο διαφορετικών συστημάτων νευρωνικής μηχανικής μετάφρασης για τα προαναφερθέντα ζεύγη γλωσσών αντιπαραβάλλοντας τις μεταφράσεις που παράγονται από αυτά με μια ανθρώπινη μετάφραση και αξιολογώντας την ακρίβεια και την ευχέρεια των μεταφράσεων που παράγονται και από τα δύο. Η σύγκριση γίνεται τόσο με αυτόματες όσο και με ανθρώπινες μετρικές, ώστε να διαπιστωθεί ποιο από τα δυο συστήματα παράγει τα καλύτερα αποτελέσματα για τους συγκεκριμένους γλωσσικούς συνδυασμούς. Οι ανάγκες για μετάφραση στη σημερινή εποχή αυξάνονται όλο και περισσότερο, πράγμα που απαιτεί την άμεση και αποτελεσματική εξέλιξη του τομέα. Πράγματι, όντας στενά συνδεδεμένος με την τεχνολογία, ο τομέας της μετάφρασης είναι ένας από τους ταχύτερα αναπτυσσόμενους τομείς. Έτσι, εκείνες οι πρώτες προσπάθειες μιας μετάφρασης λέξη προς λέξη εξελίχθηκαν σε μια αξιοσημείωτη ανάπτυξη των τεχνολογιών της μετάφρασης με ήδη αξιόλογα αποτελέσματα, αλλά και πολλά υποσχόμενες πρακτικές γύρω από αυτό το κομμάτι για το μέλλον. Ο σημερινός επαγγελματίας μεταφραστής δεν διαθέτει απλά έναν υπολογιστή για τη δουλειά του, αλλά και μια μεγάλη ποικιλία εργαλείων που βοηθούν σε ένα ταχύτερο και ποιοτικότερο αποτέλεσμα. Ανάμεσα σ' αυτά τα επιτεύγματα που η τεχνολογία έχει προσφέρει στη μετάφραση, βρίσκεται και η μηχανική μετάφραση. Η μηχανική μετάφραση (MM) είναι αποτέλεσμα της ραγδαία εξελισσόμενης υπολογιστικής γλωσσολογίας και ήρθε για να μείνει. Η χρήση της αυξάνεται συνεχώς, καθώς δεν αποτελεί πλέον ένα τομέα άγνωστο που απευθύνεται μόνο σε λίγους εξειδικευμένους επαγγελματίες, αλλά πρόκειται πλέον για μια τεχνολογική εξέλιξη στην οποία όλοι έχουν πρόσβαση. Οι πελάτες των γλωσσικών υπηρεσιών έχουν πλέον βασικές γνώσεις γύρω από τη μηχανική μετάφραση, αναγνωρίζουν τις δυνατότητες των νέων συστημάτων και τη ζητούν, με αποτέλεσμα να αποτελεί πια μία

από τις υπηρεσίες LSP. Αυτομάτως, οι επαγγελματίες αυτού του κλάδου πρέπει να είναι ενημερωμένοι και κατάλληλα εκπαιδευμένοι, ώστε να μπορούν να προσφέρουν αυτή την υπηρεσία και ό,τι χρειάζεται γύρω από αυτή, όπως για παράδειγμα συλλογή και επεξεργασία δεδομένων για τη χρήση τους σε συστήματα MM, προετοιμασία αρχείων, μετεπιμέλεια, αξιολόγηση συστημάτων MM κ.ά. Όπως είναι σαφές, δεν πρόκειται για μια απλή διαδικασία, αλλά μια νέα υπηρεσία που χρειάζεται συστηματική μελέτη και κατάρτιση, καθώς περιλαμβάνει ορισμένα στάδια τόσο πριν τη χρήση MM αλλά και μετά, μιας και ακόμα τα αποτελέσματα της MM χρειάζονται μερική ή ολική ανθρώπινη παρέμβαση για να χρησιμοποιηθούν. Ένα μεγάλο ποσοστό εταιριών, πλέον και στην Ελλάδα, προσφέρουν υπηρεσίες MM και συνεπώς μετεπιμέλειας (PE) κειμένου που έχει προκύψει από MM. Για όλους αυτούς τους λόγους, αλλά και λόγω της αυξανόμενης χρήσης των νευρωνικών συστημάτων μηχανικής μετάφρασης (Neural Machine Translation, NMT), που αποτελούν την πιο πρόσφατη εξέλιξη σ' αυτόν τον τομέα, πραγματοποιούνται πολλές έρευνες με σκοπό να αξιολογηθούν τα αποτελέσματα που παράγουν αυτά τα συστήματα και να βελτιωθούν οι αστοχίες τους, είτε συγκρίνοντας δύο συστήματα ίδιου τύπου μεταξύ τους, είτε συγκρίνοντάς τα αντιπαραβολικά με παλαιότερα συστήματα, έτσι ώστε να αποδειχθεί, ποιο σύστημα είναι πιο λειτουργικό σε ένα ζεύγος γλωσσών υπό συγκεκριμένες συνθήκες. Τέτοιες έρευνες, συμπεριλαμβανομένων των ελληνικών ως γλώσσα υπό εξέταση, έχουν διεξαχθεί από τις Sosoni και Stasimioti (2019), Stasimioti et al. (2020) και Castilho et al. (2017), με θετικά αποτελέσματα τόσο για τη χρήση MM κατά τη μεταφραστική διαδικασία, όσο και για την ποιότητα του μεταφράσματος που προέρχεται από νευρωνικά συστήματα MM.

Το θέμα, λοιπόν, της παρούσας εργασίας επιλέχθηκε τόσο για το ενδιαφέρον της γράφουσας, ως μέρος της κοινότητας LSP, για τις εξελίξεις στο συγκεκριμένο τομέα, όσο και για την αφορμή που μπορεί να αποτελέσει η συγκεκριμένη εργασία για περαιτέρω εκτενή μελέτη και σε περισσότερους γλωσσικούς συνδυασμούς, αλλά και διαφορετικά είδη κειμένων. Η παρούσα εργασία ξεκινά με ένα πρώτο θεωρητικό μέρος που περιλαμβάνει αρχικά μια σύντομη ιστορική αναδρομή και τα είδη των συστημάτων MM. Στη συνέχεια, ακολουθεί ο ορισμός της ποιότητας της μετάφρασης γενικότερα, και παρατίθενται ορισμένες θεωρίες, οι οποίες αναπτύχθηκαν σχετικά με την αξιολόγησή της, καθώς και κάποιες προσεγγίσεις αξιολόγησης. Αργότερα, αναλύεται η αξιολόγηση της ποιότητας της MM και οι μέθοδοι διεξαγωγής της. Ακολουθεί το δεύτερο ερευνητικό μέρος της εργασίας, όπου παρουσιάζεται η μεθοδολογία της έρευνας, τα αποτελέσματα που προκύπτουν και τα τελικά συμπεράσματα.

2. Μηχανική Μετάφραση

Η μηχανική μετάφραση είναι ένα υποπεδίο της υπολογιστικής γλωσσολογίας που εξετάζει τη χρήση λογισμικού για τη μετάφραση κειμένου ή ομιλίας από μια γλώσσα σε μια άλλη. Ουσιαστικά, η μηχανική μετάφραση αφορά την αντικατάσταση λέξεων σε μια γλώσσα με λέξεις σε μια άλλη, χωρίς να έχει ακόμη την ικανότητα να αντιμετωπίζει διαφορές στη γλωσσική τυπολογία, να μεταφράζει ιδιωτισμούς, να απομονώνει ανωμαλίες και αυτός είναι ο λόγος που δεν μπορεί μέχρι στιγμής να παράγει μια καλή μετάφραση. Το αποτέλεσμα, ωστόσο, εξαρτάται τόσο από τον τομέα στον οποίο ανήκει το κείμενο ή η ομιλία και το σκοπό που επιτελεί, όσο και από τα δεδομένα με τα οποία εκπαιδεύεται το κάθε σύστημα μηχανικής μετάφρασης. Η ανθρώπινη παρέμβαση παίζει επίσης σημαντικό ρόλο στην ενίσχυση της μηχανικής μετάφρασης και είναι απαραίτητη, καθώς ο τομέας αυτός εξακολουθεί να απαιτεί συνδυασμό των ικανοτήτων της μηχανής και του ανθρώπου. Η τρέχουσα τεχνολογία δεν είναι ακόμη σε θέση να δημιουργήσει έναν υπολογιστή που να μπορεί να αντικαταστήσει έναν μεταφραστή και την ικανότητά του να αναλύει ένα κείμενο σε συνδυασμό με τη βαθιά γνώση της γραμματικής, της σημασιολογίας, της σύνταξης, των ιδιωτισμών κ.λπ. της γλώσσας-πηγής και της γλώσσας-στόχου, καθώς και της κουλτούρας των ομιλητών της. Επομένως, κάθε κείμενο που παράγεται από τη μηχανική μετάφραση χρειάζεται στη συνέχεια κάποια επεξεργασία (μετεπιμέλεια) προκειμένου να χρησιμοποιηθεί.

3. Σύντομη ιστορική αναδρομή

Η εισαγωγή της μηχανής στον τομέα της μετάφρασης χρονολογείται κατά το 17ο αιώνα με τη χρήση μηχανικών λεξικών. Ωστόσο, συγκεκριμένες προτάσεις για τη χρήση της μηχανικής μετάφρασης εμφανίστηκαν τον 20ό αιώνα και συγκεκριμένα το 1933 από τον George Artsrouni, έναν Γαλλοαρμένιο, και από έναν Ρώσο, τον Petr Smirnov-Troyanskii (Hutchins, 1995). Ο Artsrouni επινόησε μια συσκευή

αποθήκευσης σε χάρτινη ταινία, η οποία θα μπορούσε να χρησιμοποιηθεί για την εύρεση του ισοδύναμου οποιασδήποτε λέξης σε άλλη γλώσσα, παρουσιάζοντας μάλιστα ένα πρωτότυπο το 1937. Παρόλο που η σκέψη αυτή ήταν αρκετά καινοτόμος εκείνη την εποχή, δεν ήταν τόσο επιτυχής όσο αναμενόταν. Αργότερα, τον Ιούλιο του 1949, ένα υπόμνημα του Warren Weaver έφερε την ιδέα της μηχανικής μετάφρασης στη γενική αντίληψη. Περιέγραφε τις προσδοκίες και τις μεθόδους που μπορούν να χρησιμοποιηθούν για την υλοποίηση αυτού του εγχειρήματος και μέσα σε λίγα χρόνια άρχισε η έρευνα σε πολλά πανεπιστήμια των Ηνωμένων Πολιτειών (Hutchins, 1995). Μετά από χρόνια μελέτης και μετά από πολλές προτάσεις, τον Ιανουάριο του 1954 στο Πανεπιστήμιο Georgetown, ο Leon Dostert συνεργάστηκε με το/ την εταιρεία IBM σε ένα έργο που είχε ως αποτέλεσμα την πρώτη δημόσια επίδειξη ενός συστήματος μηχανικής μετάφρασης (Hutchins, 1995). Ένα προσεκτικά επιλεγμένο δείγμα 49 ρωσικών προτάσεων μεταφράστηκε στα αγγλικά, χρησιμοποιώντας ένα πολύ περιορισμένο λεξιλόγιο 250 λέξεων και μόλις 6 γραμματικούς κανόνες. Αυτό το εγχείρημα δεν ήταν ακριβώς ένα θέμα υψηλής επιστημονικής σημασίας, αλλά κατάφερε να εξασφαλίσει την απαραίτητη χρηματοδότηση που χρειαζόταν η μηχανική μετάφραση εκείνη την εποχή (Hutchins, 2001).

Τα πρώτα συστήματα μηχανικής μετάφρασης χρησιμοποιούσαν μεγάλα δίγλωσσα λεξικά και χειροκίνητα κωδικοποιημένους κανόνες για την προσαρμογή της σειράς των λέξεων στο τελικό αποτέλεσμα. Αυτό θεωρήθηκε τελικά πολύ περιοριστικό για τις όποιες γλωσσικές εξελίξεις της εποχής. Κατά τη διάρκεια αυτής της περιόδου εισήχθησαν τα λειτουργικά συστήματα. Η Πολεμική Αεροπορία των Ηνωμένων Πολιτειών χρησιμοποίησε ένα σύστημα που παρήχθη από την IBM και το Πανεπιστήμιο της Ουάσιγκτον, ενώ η Επιτροπή Ατομικής Ενέργειας και η Ευρατόμ, στην Ιταλία, χρησιμοποίησαν ένα σύστημα που αναπτύχθηκε στο Πανεπιστήμιο Georgetown (Hutchins, 2005). Παρά το γεγονός ότι η ποιότητα του αποτελέσματος ήταν κακή, ήταν αρκετή για να καλύψει πολλές από τις ανάγκες των πελατών, ιδίως όσον αφορά την ταχύτητα. Στα τέλη της δεκαετίας του 1950, εμφανίστηκε το ζήτημα της σημασιολογικής ασάφειας στο πεδίο της μηχανικής μετάφρασης. Προκειμένου να επιλυθεί αυτό το πρόβλημα και να έχουμε το επιθυμητό αποτέλεσμα, οι ειδικοί κατέληξαν στο συμπέρασμα ότι τα κείμενα-πηγή για τη μηχανική μετάφραση έπρεπε να είναι γραμμένα σε μια ελεγχόμενη γλώσσα που χρησιμοποιεί ένα λεξιλόγιο, όπου κάθε λέξη έχει ακριβώς μία σημασία (Horwood, 2020).

Στη δεκαετία του 1960 η έρευνα τόσο στη Σοβιετική Ένωση όσο και στις Ηνωμένες Πολιτείες επικεντρώθηκε κυρίως στο γλωσσικό ζεύγος Ρωσικά-Αγγλικά. Το προς μετάφραση υλικό ήταν κυρίως επιστημονικά και τεχνικά έγγραφα και οι μεταφράσεις που παρήχθησαν ήταν επαρκείς για μια βασική κατανόηση των άρθρων. Σε περίπτωση που ένα άρθρο πραγματευόταν ένα θέμα που θεωρούνταν εμπιστευτικό, αποστέλλονταν σε άνθρωπο μεταφραστή. Το 1966 η έκθεση ALPAC έφερε πολλές αλλαγές στην έρευνα για τη μηχανική μετάφραση. Η έκθεση ανατέθηκε από την κυβέρνηση των ΗΠΑ και παραδόθηκε από την ALPAC, τη Συμβουλευτική Επιτροπή για την Αυτόματη Επεξεργασία της Γλώσσας (Hutchins, 2005). Η κυβέρνηση των ΗΠΑ ανησυχούσε ότι δεν είχε σημειωθεί πρόοδος παρά τις σημαντικές δαπάνες. Σύμφωνα με την έκθεση, η μηχανική μετάφραση ήταν πιο ακριβή, λιγότερο ακριβής και πιο αργή από την ανθρώπινη μετάφραση. Η έκθεση κατέληξε στο συμπέρασμα ότι παρά τις δαπάνες, η μηχανική μετάφραση δεν ήταν πιθανό να φτάσει την ποιότητα ενός ανθρώπινου μεταφραστή στο εγγύς μέλλον. Παρ' όλα αυτά, η έκθεση ανέφερε ότι η ανάπτυξη εργαλείων που κατασκευάζονται για να βοηθήσουν τους μεταφραστές καθώς και η έρευνα στην υπολογιστική γλωσσολογία θα εξακολουθούσαν να ενθαρρύνονται. Η έκθεση είχε πολύ πιο αρνητικό αντίκτυπο στην έρευνα για τη μηχανική μετάφραση στις Ηνωμένες Πολιτείες απ' ό,τι στη Σοβιετική Ένωση και το Ηνωμένο Βασίλειο. Η έρευνα, τουλάχιστον στις ΗΠΑ, σχεδόν εγκαταλείφθηκε για πάνω από μια δεκαετία. Στον Καναδά, τη Γαλλία και τη Γερμανία, ωστόσο, η έρευνα συνεχίστηκε. Στις ΗΠΑ οι κύριες εξαιρέσεις ήταν οι ιδρυτές της Systran (Peter Toma) και της Logos (Bernard Scott), οι οποίοι ίδρυσαν τις εταιρείες τους το 1968 και το 1970 αντίστοιχα και δούλεψαν για το αμερικανικό υπουργείο Άμυνας (Horwood, 2020). Το 1970, το σύστημα Systran εγκαταστάθηκε για την Πολεμική Αεροπορία των Ηνωμένων Πολιτειών και στη συνέχεια για την Επιτροπή των Ευρωπαϊκών Κοινοτήτων το 1976. Το σύστημα METEO, που αναπτύχθηκε στο Université de Montréal, χρησιμοποιήθηκε στον Καναδά το 1977 για τη μετάφραση των μετεωρολογικών προγνώσεων από τα αγγλικά στα γαλλικά, και μετέφραζε σχεδόν 80.000 λέξεις την ημέρα ή 30 εκατομμύρια λέξεις ετησίως μέχρι που αντικαταστάθηκε από ένα ανταγωνιστικό σύστημα στις 30 Σεπτεμβρίου 2001 (Hutchins, 2005).

Μέχρι τη δεκαετία του 1980, υπήρχαν διάφορα και πολυάριθμα εγκατεστημένα συστήματα μηχανικής μετάφρασης. Η τεχνολογία κεντρικών υπολογιστών εισήχθη επίσης και χρησιμοποιήθηκε από συστήματα όπως τα Systran, Logos, Ariane-G5 και Metal. Κατά τη διάρκεια της δεκαετίας του 1980 υπήρξε μεγάλη δραστηριότητα στον τομέα της μηχανικής μετάφρασης στην Ιαπωνία ειδικά (Hutchins, 2001). Η έρευνα εκείνη την εποχή επικεντρώθηκε στη μετάφραση με την ενσωμάτωση της μορφολογικής, συντακτικής και σημασιολογικής ανάλυσης. Στα τέλη της δεκαετίας του 1980, εμφανίστηκαν τόσο οι στατιστικές μέθοδοι όσο και η μηχανική μετάφραση βάσει παραδειγμάτων. Κοινό χαρακτηριστικό τους ήταν η αδιαφορία για τους συντακτικούς και σημασιολογικούς κανόνες και η στήριξη στη διαχείριση μεγάλων σωμάτων κειμένων (Hutchins, 2001).

Η δημιουργία υπολογιστών χαμηλού κόστους και ισχυρότερων υπολογιστών ενθάρρυνε την αύξηση της χρήσης της μηχανικής μετάφρασης. Στις αρχές της δεκαετίας του 1990 η μηχανική μετάφραση άρχισε να κάνει τη μετάβαση από τους μεγάλους κεντρικούς υπολογιστές προς τους προσωπικούς υπολογιστές και τους σταθμούς εργασίας. Επίσης, έγιναν διαθέσιμες ιστοσελίδες στο διαδίκτυο, όπως το Babel Fish της AltaVista και η Google. Στη δεκαετία του 1990, σύμφωνα με την τεχνολογική πρόοδο που σημειώθηκε εκείνη την εποχή άρχισε η έρευνα και για τη μετάφραση ομιλίας (Hogwood, 2020).

Στη δεκαετία του 2000, ο τομέας της μηχανικής μετάφρασης γνώρισε σημαντικές αλλαγές. Η έρευνα επικεντρώθηκε κυρίως στη στατιστική μηχανική μετάφραση (SMT) και στη μηχανική μετάφραση βάσει παραδειγμάτων, οι οποίες χρησιμοποιούνταν από εταιρείες για χρόνια. Η στατιστική μηχανική μετάφραση προσπαθεί να παράγει μεταφράσεις χρησιμοποιώντας στατιστικές μεθόδους που βασίζονται σε δίγλωσσα σώματα κειμένων. Όπου υπάρχουν τέτοια σώματα κειμένων, μπορούν να επιτευχθούν καλά αποτελέσματα κατά τη μετάφραση παρόμοιων κειμένων, αλλά τέτοια σώματα κειμένων εξακολουθούν να είναι σπάνια για πολλά ζεύγη γλωσσών (Sfetcu, 2009). Με ελαφρώς παρόμοιο τρόπο, η μηχανική μετάφραση βάσει παραδειγμάτων βασίζεται στην ιδέα της αναλογίας. Σε αυτή την προσέγγιση, το σώμα που χρησιμοποιείται είναι αυτό που περιέχει κείμενα που έχουν ήδη μεταφραστεί. Δεδομένου ενός πρότασης που πρόκειται να μεταφραστεί, επιλέγονται προτάσεις από το σώμα αυτό που περιέχουν παρόμοιες υποδηλώσεις (Hutchins, and L. Somers, 1992). Οι παρόμοιες προτάσεις χρησιμοποιούνται στη συνέχεια για τη μετάφραση. τα υπο-εννοιολογικά στοιχεία της αρχικής πρότασης στη γλώσσα-στόχο, και οι φράσεις αυτές συνδυάζονται για να σχηματίσουν μια πλήρη μετάφραση. Και οι δύο αυτές προσεγγίσεις ήταν ατελείς από γλωσσολογική άποψη, καθώς εντοπίστηκαν λάθη στη σύνταξη, τη μορφολογία, την ευχέρεια και πολλά άλλα.

Ωστόσο, οι τελευταίες εξελίξεις στον τομέα της μηχανικής μετάφρασης φαίνονται περισσότερο από ελπιδοφόρες για το μέλλον. Η νευρωνική μηχανική μετάφραση (NMT) έγινε πραγματικότητα τα τελευταία χρόνια και παρουσιάζει βελτιωμένα αποτελέσματα σε σύγκριση με προηγούμενες προσεγγίσεις μηχανικής μετάφρασης. Χρησιμοποιεί ένα τεχνητό νευρωνικό δίκτυο για να προβλέψει την πιθανότητα μιας ακολουθίας λέξεων, μοντελοποιώντας συνήθως ολόκληρες προτάσεις σε ένα ενιαίο ολοκληρωμένο μοντέλο. Τα συστήματα νευρωνικής μηχανικής μετάφρασης απαιτούν μόνο ένα κλάσμα της μνήμης που χρειάζονται τα παραδοσιακά μοντέλα στατιστικής μηχανικής μετάφρασης. Επιπλέον, σε αντίθεση με τα συμβατικά συστήματα μετάφρασης, όλα τα μέρη του μοντέλου νευρωνικής μετάφρασης εκπαιδεύονται από κοινού (end-to-end) για τη μεγιστοποίηση της απόδοσης της μετάφρασης (Chen, et al, 2019). Από το 2015 που εμφανίστηκαν αυτά τα συστήματα έχουν γίνει ιδιαίτερα γνωστά και χρησιμοποιούνται ευρέως.

4. Συστήματα Μηχανικής Μετάφρασης

Τα κυριότερα συστήματα Μηχανικής Μετάφρασης που θα μας βοηθήσουν να δούμε την εξέλιξη αυτού του τομέα φτάνοντας στα νευρωνικά συστήματα της σημερινής εποχής, που είναι και το αντικείμενο μελέτης της παρούσας εργασίας, αναφέρονται παρακάτω και είναι τα Συστήματα Μηχανικής Μετάφρασης Βάσει Κανόνων, τα Συστήματα Μηχανικής Μετάφρασης Βάσει Παραδειγμάτων, τα Στατιστικά Συστήματα Μηχανικής Μετάφρασης, τα Υβριδικά Συστήματα Μηχανικής Μετάφρασης και τα Νευρωνικά Συστήματα Μηχανικής Μετάφρασης.

i. Συστήματα Μηχανικής Μετάφρασης Βάσει Κανόνων (Rule-Based Machine Translation, RBMT)

Τα συστήματα μηχανικής μετάφρασης βάσει κανόνων (Rule-Based Machine Translation, RBMT) ανήκουν στα οντολογικά συστήματα (Knowledge-Based Machine Translation), τα οποία χρησιμοποιούν βάσεις γνώσεων και τεχνικές τεχνητής νοημοσύνης για την μετάφραση ενός κειμένου από μια γλώσσα σε μια άλλη. Υπάρχουν τρεις τύποι συστημάτων μηχανικής μετάφρασης βάσει κανόνων (RBMT): α) το λεξιλογικό σύστημα (Dictionary Based Machine Translation), β) το σύστημα μεταφοράς (Transfer Based Machine Translation), και γ) το διαγλωσσικό σύστημα (Interlingual RBMT Systems) (Nirenburg, 1989). Το λεξιλογικό σύστημα μεταφράζει λέξη προς λέξη από μια ΓΠ σε μια ΓΣ λαμβάνοντας υπόψιν μερικά μορφολογικά στοιχεία και κάποιους κανόνες συσχέτισης των λέξεων στις δυο γλώσσες. Αντίθετα, τα συστήματα μεταφοράς λειτουργούν βάσει των διαφορών στη δομή της κάθε γλώσσας. Σε αυτά τα συστήματα, η μετάφραση του κειμένου περνά από τρία στάδια, την ανάλυση, την μεταφορά και την παραγωγή. Αρχικά, γίνεται μια γραμματική ανάλυση στο ΚΠ και σε αυτή την ανάλυση γίνεται εφαρμογή συντακτικών και λεξιλογικών κανόνων, ώστε να μεταφερθούν στην άλλη γλώσσα, και τέλος γίνεται η παραγωγή στη ΓΣ. Τα διαγλωσσικά συστήματα περιλαμβάνουν μια ουδέτερη διαγλώσσα (interlingua), εκτός των γλωσσών που περιλαμβάνονται στη μεταφραστική διαδικασία, η οποία περιέχει και αναλύει τα χαρακτηριστικά όλων των προς μετάφραση γλωσσών, έτσι ώστε το σύστημα να συνθέσει πλέον σωστά το ΚΣ σε μορφολογικό, γραμματικό και συντακτικό επίπεδο (Hutchins, 1986).

ii. Συστήματα Μηχανικής Μετάφρασης Βάσει Παραδειγμάτων (Example-Based Machine Translation, EBMT)

Τα συστήματα μηχανικής μετάφρασης βάσει παραδειγμάτων (Example-Based Machine Translation) λειτουργούν με παράλληλα σώματα κειμένου. Το σύστημα αυτού του είδους εκπαιδεύεται να χωρίζει τις προτάσεις σε μικρότερα τμήματα (φράσεις) και βάσει παλαιότερων μεταφράσεων να αποδίδει αυτά τα τμήματα. Πιο συγκεκριμένα, το σύστημα αυτό αποτελεί ένα σύστημα που μεταφράζει λέξη προς λέξη, με τη χρήση δίγλωσσων γλωσσαρίων και με την σύνδεση φράσεων της ΓΠ με αντίστοιχες φράσεις της ΓΣ (Nagao, 1984).

iii. Στατιστικά Συστήματα Μηχανικής Μετάφρασης (Statistical Machine Translation, SMT)

Τα στατιστικά συστήματα μηχανικής μετάφρασης (Statistical Machine Translation, SMT) βασίζονται στις πιθανότητες (Brown, et.al, 1993). Η λειτουργία τους βασίζεται στη θεωρία της πληροφορίας του Claude Shannon και τη θεωρία των πιθανοτήτων και της στατιστικής, κι εδώ το σύστημα παρέχει ως μετάφραση την πιο πιθανή απόδοση από μια γλώσσα σε μια άλλη. Τα πρώτα στατιστικά συστήματα μηχανικής μετάφρασης (Word-Based Machine Translation) λειτουργούσαν σε επίπεδο λέξης, δηλαδή κάθε λέξη της ΓΠ αντιστοιχούσε σε κάποια λέξη της ΓΣ. Η αντιστοιχία αυτή δεν περιοριζόταν στη θέση των λέξεων στην εκάστοτε πρόταση, αλλά η κάθε λέξη της ΓΠ μπορεί να αντιστοιχούσε σε μια ή παραπάνω ή και σε ορισμένες περιπτώσεις σε καμία λέξεις της ΓΣ, λόγω των μορφολογικών διαφοροποιήσεων κάθε γλώσσας. Με την εξέλιξη των στατιστικών συστημάτων ξεκίνησε η εσωτερική επεξεργασία ενός κειμένου σε επίπεδο φράσης (Phrase-Based Machine Translation, PBMT), η οποία φράση μπορεί να περιλαμβάνει μια ή παραπάνω λέξεις. Το σύστημα, σ' αυτήν την περίπτωση, επεξεργάζεται την πιθανότητα συνεμφάνισης μιας φράσης σε μια ΓΠ με μια φράση στη ΓΣ (Kalchbrenner, Blunsom, 2013). Σύμφωνα με έρευνα των Koehn et. Al (2003), σε επίπεδο φράσης, όταν το περιεχόμενο του κειμένου που εκπαιδεύεται το σύστημα είναι απλό, τα αποτελέσματα είναι καλύτερα και υπάρχει μεγαλύτερη ακρίβεια, όταν η φράση περιλαμβάνει έως και τρεις λέξεις. Συγκρίνοντας πολλά συστήματα που λειτουργούν σε επίπεδο φράσης με παραδοσιακά στατιστικά συστήματα εκπαιδευμένα σε επίπεδο λέξης, η ίδια έρευνα απέδειξε ότι η εκπαίδευση σε επίπεδο φράσης δίνει καλύτερα αποτελέσματα. Στα συγκεκριμένα συστήματα αρκετά σημαντική είναι και η κατανομή βάρους σε συμφράσεις, ώστε να αυξηθούν οι πιθανότητες επανεμφάνισης τους.

iv. Υβριδικά Συστήματα Μηχανικής Μετάφρασης (Hybrid Machine Translation)

Πριν την εμφάνιση του πιο σημαντικού επιτεύγματος στη Μηχανική Μετάφραση σήμερα, των νευρωνικών συστημάτων (Neural Machine Translation, NMT), η τελευταία εξέλιξη του τομέα ήταν τα υβριδικά συστήματα

Μηχανικής Μετάφρασης (Hybrid Machine Translation), τα οποία αποτελούνται από δυο τύπους συστημάτων, το στατιστικό σύστημα (Statistical Machine Translation) και το σύστημα μηχανικής μετάφρασης βάσει κανόνων (Rule-Based Machine Translation). Υπάρχουν τρία είδη υβριδικών συστημάτων και αυτά είναι: α) η σύζευξη των συστημάτων (coupling of systems), β) οι προσαρμογές στην αρχιτεκτονική (architecture adaptations) και γ) τα γνήσια υβριδικά συστήματα (genuine hybrid systems). Όσον αφορά την πρώτη κατηγορία, τη σύζευξη, εδώ γίνεται ένας συνδυασμός των στατιστικών συστημάτων (SMT) με τα συστήματα μηχανικής μετάφρασης βάσει κανόνων (RBMT). Η σύζευξη λειτουργεί με δυο τρόπους: α) η μετεπεξεργασία (Post-Editing, PE) του κειμένου που προέρχεται από το σύστημα βάσει κανόνων (RBMT) μπορεί να γίνει με τους κανόνες που διέπουν την μετεπεξεργασία (PE) κειμένων που προέρχονται από στατιστικό σύστημα μηχανικής μετάφρασης (SMT) ή β) η δημιουργία ενός μεταφράσματος καλής ποιότητας μπορεί να γίνει με τον συνδυασμό μεταφράσεων που προκύπτουν από τα δυο συστήματα MM. Σχετικά με τις προσαρμογές στην αρχιτεκτονική, η αρχιτεκτονική των συστημάτων ακολουθεί ένα από τα δύο συστήματα, το στατιστικό σύστημα ή το σύστημα βάσει κανόνων, και προστίθενται πρακτικές του άλλου συστήματος αντίστοιχα. Αυτό επιτυγχάνεται με δυο τρόπους, με την προεπεξεργασία (PreEditing) του κειμένου πριν μεταφραστεί από το σύστημα ή με μετατροπές (συνήθως προσθήκες του άλλου) στο πυρήνα του ενός συστήματος. Τέλος, τα γνήσια υβριδικά συστήματα δεν λειτουργούν με προσθήκες σε ένα από τα δυο συστήματα, αλλά συντίθενται με ολόκληρα τμήματα των δυο συστημάτων, δημιουργώντας ένα νέο (Thurmaier, 2009).

ν. Νευρωνικά Συστήματα Μηχανικής Μετάφρασης (Neural Machine Translation, NMT)

Η τελευταία εξέλιξη στον τομέα της μηχανικής μετάφρασης είναι η δημιουργία συστημάτων βάσει νευρωνικών δικτύων. Έρευνες όπως αυτές των Kalchbrenner και Blunsom (2013), των Cho et al. (2014b), και των Sutskever et al. (2014) υπήρξαν από τις πρώτες που περιείχαν αναφορές στα συστήματα αυτού του τύπου. Τα νευρωνικά συστήματα μηχανικής μετάφρασης ουσιαστικά περνούν τη μετάφραση από δυο στάδια, τη κωδικοποίηση και την αποκωδικοποίηση (Sutskever et al. 2014 and Cho et al., 2014b). Το κείμενο εισόδου είναι αυτό στη ΓΠ, το οποίο κωδικοποιείται σε μια μορφή αναπαράστασης και αποκωδικοποιείται κατά την έξοδο ώστε το αποτέλεσμα να είναι μια σωστή μετάφραση στη ΓΣ (Cho et al., 2014b). Κατά την επεξεργασία του κειμένου στη ΓΠ, η αναπαράσταση του κειμένου περνά από κάποια κρυφά επίπεδα για την παραγωγή του μεταφράσματος. Προς το παρόν, η πλήρης λειτουργία των δικτύων αυτών παραμένει άγνωστη, ωστόσο η απόδοσή τους φαίνεται να δίνει πολύ καλά αποτελέσματα. Τα νευρωνικά συστήματα μηχανικής μετάφρασης χωρίζονται σε γενικής χρήσης (generic) και ειδικά εκπαιδευμένα (factored/custom-trained) (Vashee, 2018). Οι δυο αυτοί τύποι έχουν να κάνουν με τα δεδομένα, με τα οποία εκπαιδεύεται το κάθε σύστημα. Τα συστήματα γενικής χρήσης (generic) εκπαιδεύονται με γενικά κείμενα, μη εξειδικευμένα, χωρίς ιδιαίτερη ειδική ορολογία. Αντίθετα, τα ειδικά εκπαιδευμένα συστήματα εκπαιδεύονται με συγκεκριμένα δεδομένα ειδικών κειμενικών τύπων και ορολογίας, ώστε να επικεντρώνονται σε ένα συγκεκριμένο πεδίο και σε αυτό να δίνουν καλύτερα αποτελέσματα. Τέτοιου είδους συστήματα χρησιμοποιούνται κυρίως από εταιρίες, οι οποίες εκπαιδεύουν τα συστήματά τους βάσει του τομέα στον οποίο εξειδικεύονται, ώστε να έχουν καλύτερα αποτελέσματα. Ένα από τα πλεονεκτήματα των νευρωνικών συστημάτων μηχανικής μετάφρασης είναι ότι μαθαίνουν απευθείας από τα δεδομένα χωρίς να βασίζονται σε χειροποίητα χαρακτηριστικά ή ρητούς γλωσσικούς κανόνες. Επίσης, τα νευρωνικά δίκτυα, ιδίως με μηχανισμούς προσοχής, διακρίνονται για την καταγραφή πληροφοριών σχετικά με το συγκεκριμένο, μια λειτουργία ζωτικής σημασίας για τον χειρισμό των ασαφειών και την παραγωγή πιο εύληπτων μεταφράσεων. Οι μηχανισμοί προσοχής βοηθούν στην αντιμετώπιση της πρόκλησης του χειρισμού μεγάλων προτάσεων, επιτρέποντας στο μοντέλο να εστιάζει επιλεκτικά σε σχετικά τμήματα της ακολουθίας εισόδου. Τα νευρωνικά συστήματα μηχανικής μετάφρασης τείνουν να γενικεύουν καλύτερα σε διάφορα ζεύγη γλωσσών, τομείς και μεταφραστικές εργασίες.

Συνολικά, τα νευρωνικά συστήματα μηχανικής μετάφρασης έχουν επιδείξει ανώτερη ποιότητα μετάφρασης σε σύγκριση με τα παραδοσιακά συστήματα στατιστικής μηχανικής μετάφρασης, ιδίως όσον αφορά τη σύλληψη σύνθετων γλωσσικών προτύπων. Ωστόσο, ενώ τα νευρωνικά συστήματα μηχανικής μετάφρασης έχουν παρουσιάσει σημαντικές βελτιώσεις, είναι σημαντικό να σημειωθεί ότι διεξάγεται συνεχής έρευνα για την περαιτέρω ενίσχυση των δυνατοτήτων τους.

Για τους σκοπούς της παρούσας εργασίας, τα δύο νευρωνικά συστήματα μηχανικής μετάφρασης που θα χρησιμοποιηθούν ως προς τη συγκριτική μελέτη είναι το Google Translate και το DeepL. Ας δούμε αναλυτικά παρακάτω ορισμένα χαρακτηριστικά τους.

- Google Translate

Το σύστημα Google Neural Machine Translation (GNMT) είναι ένα σύστημα νευρωνικής μηχανικής μετάφρασης (NMT) που αναπτύχθηκε από την Google, παρουσιάστηκε τον Νοέμβριο του 2016 και χρησιμοποιεί ένα τεχνητό νευρωνικό δίκτυο για να αυξήσει την ευχέρεια και την ακρίβεια στο Google Translate. Έκτοτε, το Google Translate άρχισε να χρησιμοποιεί νευρωνική μηχανική μετάφραση (NMT) αντί των προηγούμενων στατιστικών μεθόδων (SMT) που χρησιμοποιούσε από τον Οκτώβριο του 2007, με την ιδιότητα, εσωτερική τεχνολογία SMT (Turovsky, 2016).

Το σύστημα NMT της Google Translate χρησιμοποιεί ένα μεγάλο τεχνητό νευρωνικό δίκτυο ικανό για βαθιά μάθηση. Με τη χρήση εκατομμυρίων παραδειγμάτων, το GNMT βελτιώνει την ποιότητα της μετάφρασης, χρησιμοποιώντας ευρύτερα συμφραζόμενα για την εξαγωγή της πιο σχετικής μετάφρασης. Το αποτέλεσμα στη συνέχεια αναδιατάσσεται και προσαρμόζεται ώστε να προσεγγίζει τη γραμματική βάση της ανθρώπινης γλώσσας (Turovsky, 2016). Η προτεινόμενη αρχιτεκτονική μάθησης του συστήματος GNMT δοκιμάστηκε αρχικά σε περισσότερες από εκατό γλώσσες που υποστηρίζονται από το Google Translate. Το GNMT δεν δημιούργησε τη δική του καθολική διαγλώσσα, αλλά μάλλον στόχευσε στην εύρεση των κοινών σημείων μεταξύ πολλών γλωσσών χρησιμοποιώντας γνώσεις από την ψυχολογία και τη γλωσσολογία. Η νέα μηχανή μετάφρασης ενεργοποιήθηκε αρχικά για λίγες γλώσσες, ενώ σταδιακά μέχρι το 2017 προστέθηκαν και οι υπόλοιπες που υποστηρίζονται από το σύστημα (Turovsky, 2017).

Αναλυτικά, το μοντέλο του Google Translate αποτελείται από ένα βαθύ δίκτυο LSTM με 8 επίπεδα κωδικοποιητή και 8 επίπεδα αποκωδικοποιητή που χρησιμοποιούν συνδέσεις προσοχής και υπολειμματικές συνδέσεις. Για να βελτιωθεί ο παραλληλισμός και επομένως να μειωθεί ο χρόνος εκπαίδευσης, ο μηχανισμός προσοχής του, συνδέει το κάτω στρώμα του αποκωδικοποιητή με το πάνω στρώμα του κωδικοποιητή. Για να επιταχυνθεί η τελική ταχύτητα μετάφρασης, χρησιμοποιείται αριθμητική χαμηλής ακρίβειας κατά τη διάρκεια των υπολογισμών συμπερασμού. Για να βελτιωθεί ο χειρισμός σπάνιων λέξεων, οι λέξεις διαιρούνται σε ένα περιορισμένο σύνολο κοινών υπο-λεξιλογικών μονάδων (wordpieces) τόσο για την είσοδο όσο και για την έξοδο. Αυτή η μέθοδος παρέχει μια καλή ισορροπία μεταξύ της ευελιξίας των μοντέλων που περιορίζονται σε «χαρακτήρες» και της αποτελεσματικότητας των μοντέλων που περιορίζονται σε «λέξεις», χειρίζεται φυσικά τη μετάφραση σπάνιων λέξεων και τελικά βελτιώνει τη συνολική ακρίβεια του συστήματος. Η τεχνική αναζήτησης με δέσμη χρησιμοποιεί μια διαδικασία κανονικοποίησης του μήκους και χρησιμοποιεί μια ποινή κάλυψης, η οποία ενθαρρύνει τη δημιουργία μιας πρότασης εξόδου που είναι πιθανότερο να καλύπτει όλες τις λέξεις της αρχικής πρότασης. Με αυτή τη διαδικασία, για παράδειγμα στα γλωσσικά ζεύγη Αγγλικά-Γαλλικά και Αγγλικά-Γερμανικά, το GNMT επιτυγχάνει ανταγωνιστικά αποτελέσματα σε σχέση με την τελευταία λέξη της τεχνολογίας. Χρησιμοποιώντας ανθρώπινη αξιολόγηση δίπλα-δίπλα σε ένα σύνολο απομονωμένων απλών προτάσεων, μειώνει τα μεταφραστικά σφάλματα κατά μέσο όρο κατά 60% σε σύγκριση με το σύστημα παραγωγής με βάσει φράσεων (phrase-based production system) της Google (Sommerlad, 2018).

- DeepL

Το DeepL Translator είναι μια νευρωνική υπηρεσία αυτόματης μετάφρασης που ξεκίνησε τον Αύγουστο του 2017 και ανήκει στην DeepL SE με έδρα την Κολωνία. Το μεταφραστικό σύστημα αναπτύχθηκε αρχικά στο πλαίσιο της Linguee και ξεκίνησε αργότερα ως οντότητα DeepL. Αρχικά προσέφερε μεταφράσεις μεταξύ επτά ευρωπαϊκών γλωσσών και σταδιακά επεκτάθηκε ώστε να υποστηρίζει 31 γλώσσες. Ο αλγόριθμός του χρησιμοποιεί συνελκτικά νευρωνικά δίκτυα και έναν αγγλικό άξονα. Προσφέρει συνδρομή επί πληρωμή για

πρόσθετα χαρακτηριστικά και πρόσβαση στη διεπαφή προγραμματισμού της μεταφραστικής εφαρμογής του.¹

Είναι γνωστό ότι τα περισσότερα δημόσια διαθέσιμα συστήματα μετάφρασης αποτελούν άμεσες τροποποιήσεις της αρχιτεκτονικής Transformer. Τα νευρωνικά δίκτυα του DeepL περιέχουν επίσης τμήματα αυτής της αρχιτεκτονικής, όπως οι μηχανισμοί προσοχής. Ωστόσο, υπάρχουν επίσης σημαντικές διαφορές στην τοπολογία των δικτύων, οι οποίες οδηγούν σε συνολική σημαντική βελτίωση της ποιότητας της μετάφρασης σε σχέση με την κατάσταση της δημόσιας έρευνας. Τα δίκτυα εκπαίδευσης εκπαιδεύονται συνήθως με τη μέθοδο της «μάθησης με επίβλεψη». Στο δίκτυο παρουσιάζονται ξανά και ξανά διαφορετικά παραδείγματα. Το δίκτυο συγκρίνει επανειλημμένα τις δικές του μεταφράσεις με τις μεταφράσεις από τα δεδομένα εκπαίδευσης. Εάν υπάρχουν αποκλίσεις, τα βάρη του δικτύου προσαρμόζονται ανάλογα.²

Πιο συγκεκριμένα, η υπηρεσία DeepL χρησιμοποιεί έναν ιδιόκτητο αλγόριθμο με συνεπτυγμένα νευρωνικά δίκτυα (CNN) που έχουν εκπαιδευτεί με τη βάση δεδομένων της Linguee. Σύμφωνα με τους προγραμματιστές, η υπηρεσία χρησιμοποιεί μια νεότερη, βελτιωμένη αρχιτεκτονική νευρωνικών δικτύων, η οποία έχει ως αποτέλεσμα έναν πιο φυσικό ήχο των μεταφράσεων σε σύγκριση με τις ανταγωνιστικές υπηρεσίες. Η μετάφραση λέγεται ότι παράγεται με τη χρήση ενός υπερυπολογιστή που φτάνει τα 5,1 petaflops και λειτουργεί στην Ισλανδία με υδροηλεκτρική ενέργεια. Σε γενικές γραμμές, τα CNN είναι ελαφρώς πιο κατάλληλα για μεγάλες, συνεκτικές ακολουθίες λέξεων, αλλά μέχρι στιγμής δεν έχουν χρησιμοποιηθεί από τον ανταγωνισμό λόγω των αδυναμιών τους σε σύγκριση με τα επαναλαμβανόμενα νευρωνικά δίκτυα. Οι αδυναμίες στο DeepL αντισταθμίζονται με συμπληρωματικές τεχνικές, ορισμένες από τις οποίες είναι γνωστές στο κοινό.³

Τόσο για το DeepL όσο και άλλους παρόχους νευρωνικής μηχανικής μετάφρασης εκπαιδεύονται μεταφραστικά δίκτυα με πολλά δισεκατομμύρια παραμέτρων. Αυτά τα δίκτυα είναι τόσο μεγάλα που μπορούν να εκπαιδευτούν μόνο με κατανεμημένο τρόπο σε πολύ μεγάλες ειδικές υπολογιστικές συστάδες. Ωστόσο, οι ερευνητές του DeepL δίνουν μεγάλη σημασία στο γεγονός ότι οι παράμετροι του δικτύου χρησιμοποιούνται πολύ αποτελεσματικά. Με αυτόν τον τρόπο επιτυγχάνεται παρόμοια ποιότητα μετάφρασης ακόμη και με τα μικρότερα και ταχύτερα δίκτυα. Επομένως, υπάρχει δυνατότητα παροχής πολύ υψηλής ποιότητας μετάφρασης και στους χρήστες της δωρεάν υπηρεσίας του συστήματος.⁴

5. Ποιότητα της Μετάφρασης

Ο προσδιορισμός της ποιότητας της μετάφρασης έχει αποδειχθεί ένα αρκετά δύσκολο εγχείρημα για τους ερευνητές. Ωστόσο, είναι απαραίτητο να γίνει κατανοητή η έννοια και η σημασία της ποιότητας της μετάφρασης γενικότερα, πριν φτάσουμε στην αξιολόγηση της μηχανικής μετάφρασης. Ανάμεσα σε πολλές και αντικρουόμενες θεωρίες για την ποιότητα της μετάφρασης, οι Fields et al. (2014) καταλήγουν σε δυο ορισμούς, έναν στενό ορισμό (narrow definition) κι έναν ευρύ ορισμό (broad definition). Η διαφορά αυτών των δύο ορισμών είναι ότι ο στενός ορισμός εκλαμβάνει τη μετάφραση ως μια διαδικασία που επικεντρώνεται αυστηρά στο κείμενο, ενώ ο ευρύς ορισμός έχει ως δεδομένο ότι η μετάφραση αφορά διάφορες διαδικασίες που υπερβαίνουν το κείμενο (Koby et al., 2014).

Όπως έχει αναφερθεί, ο στενός ορισμός δεν αναγνωρίζει άλλες διαδικασίες (π.χ. τοπική προσαρμογή, κ.α.) στη μετάφραση. Στην περίπτωση της Μηχανικής Μετάφρασης το μετάφρασμα που προκύπτει χρειάζεται

¹ Inside DeepL: The World's Fastest-Growing, Most Secretive Machine Translation Company, 15/09/2021, Slator.com (<https://slator.com/inside-deepl-the-worlds-fastest-growing-most-secretive-machine-translation-company/>), accessed on: September 2023

² How does DeepL work?, 01/11/2021, DeepL.com (<https://www.deepl.com/en/blog/how-does-deepl-work/>), accessed on: September 2023

³ Inside DeepL: The World's Fastest-Growing, Most Secretive Machine Translation Company, 15/09/2021, Slator.com (<https://slator.com/inside-deepl-the-worlds-fastest-growing-most-secretive-machine-translation-company/>), accessed on: September 2023

⁴ Inside DeepL: The World's Fastest-Growing, Most Secretive Machine Translation Company, 15/09/2021, Slator.com (<https://slator.com/inside-deepl-the-worlds-fastest-growing-most-secretive-machine-translation-company/>), accessed on: September 2023

πλήρη μετεπεξεργασία (Full-PE) για να υπάρχει ένα ικανοποιητικό ποσοστό ακρίβειας και ευχέρειας. Σε αυτόν τον ορισμό, οι προδιαγραφές δημιουργίας του μεταφράσματος είναι απόλυτες (π.χ. μεταφραστικές οδηγίες κλπ.).

Ο στενός ορισμός της ποιότητας της μετάφρασης σύμφωνα με τους Koby et al. (2014) είναι: «Μια μετάφραση υψηλής ποιότητας είναι εκείνη κατά την οποία το μήνυμα που ενσωματώνεται στο κείμενο-πηγή μεταφέρεται πλήρως στο κείμενο-στόχο, συμπεριλαμβανομένης της δήλωσης, της συνδήλωσης, της σημασιολογικής διαφοράς και του ύφους, και το κείμενο-στόχος πρέπει να είναι γραμμένο στη γλώσσα-στόχος με χρήση σωστής γραμματικής και σειράς των λέξεων, ώστε να παραχθεί ένα πολιτισμικά κατάλληλο κείμενο που, στις περισσότερες περιπτώσεις, διαβάζεται σαν να είχε γραφτεί αρχικά από έναν φυσικό ομιλητή της γλώσσας-στόχου για αναγνώστες του πολιτισμού-στόχου».

Από την άλλη, για τον ευρύ ορισμό κύρια προϋπόθεση είναι ότι δεν υπάρχουν απόλυτες προδιαγραφές για το μετάφρασμα, αλλά λαμβάνονται υπόψη μόνο όταν αυτό είναι εφικτό και έχει σχέση με το κοινό και με τον σκοπό του μεταφράσματος.

Ο ευρύς ορισμός της ποιότητας της μετάφρασης είναι: «Μια ποιοτική μετάφραση επιδεικνύει την ακρίβεια και την ευχέρεια που απαιτείται για τον αποδέκτη και τον σκοπό και συμμορφώνεται με όλες τις άλλες προδιαγραφές που θέτει ο αιτών και ο πάροχος, λαμβάνοντας υπόψη τις ανάγκες του τελικού χρήστη.» (Koby et al., 2014).

Παρατηρείται ότι ο στενός ορισμός θα μπορούσε να αποτελεί μια συγκεκριμένη περίπτωση του ευρύ ορισμού με κάποιες προκαθορισμένες προδιαγραφές. Ωστόσο, η επιλογή ενός από τους δύο ορισμούς επηρεάζει το πλαίσιο για την ανάπτυξη μετρικών της ποιότητας της μετάφρασης και έχει πρακτικές επιπτώσεις στον υπολογισμό της (Koby et al., 2014). Η δυσκολία στη δημιουργία ενός καθολικού και κοινώς αποδεκτού ορισμού για την ποιότητα της μετάφρασης δεν επιτρέπει στους μελετητές να καθορίσουν ένα κοινό μοντέλο για την αξιολόγηση της ποιότητας της μετάφρασης.

Παρ' όλα αυτά, η εξασφάλιση της ποιότητας ενός μεταφράσματος αποτελεί αναπόσπαστο κομμάτι της μεταφραστικής διαδικασίας. Πρόκειται για ένα θέμα που έχει απασχολήσει και εξακολουθεί να απασχολεί σε σημαντικό βαθμό την ακαδημαϊκή κοινότητα, καθώς πέρα από την έλλειψη μιας κοινής θεωρίας για την αξιολόγηση της μετάφρασης, δεν υπάρχει κι ένας καθολικός τρόπος διαχωρισμού των προσεγγίσεων γύρω από αυτή. Η Lauscher (2000) εκθέτει έρευνες από διάφορες οπτικές που είτε επικεντρώνονται στον λειτουργικό χαρακτήρα του πρωτοτύπου, είτε βασίζονται στη θεωρία του σκοπού, που έχει αναλυθεί από διάφορους άλλους μελετητές (Reiß & Vermeer 1984; Holz-Mänttari 1984; Vermeer 1989, 1996; Nord 1993). Αυτές οι οπτικές βασίζονται κυρίως στη λειτουργικότητα της μετάφρασης και αποστασιοποιούνται από το πρωτότυπο. Σκοπός της μετάφρασης εδώ είναι να ανταποκρίνεται στις ανάγκες του κοινού-στόχου κι όχι να διατηρηθεί ο σκοπός του πρωτοτύπου κειμένου και στο μετάφρασμα. Πρόκειται για τις προσεγγίσεις βάσει της ισοδυναμίας (Equivalence-Based approaches) και τις λειτουργικές προσεγγίσεις (Functional approaches) αντίστοιχα. Για τις προσεγγίσεις βάσει της ισοδυναμίας η μετάφραση είναι μια προσπάθεια αναπαραγωγής του πρωτοτύπου, με το μετάφρασμα που παράγεται να είναι όσο πιο κοντά γίνεται στο κείμενο-πηγή. Αντίθετα, για τις λειτουργικές προσεγγίσεις η μετάφραση είναι η διαδικασία της παραγωγής ενός κειμένου βάσει ενός πρωτοτύπου, με το κείμενο-στόχος να αποτελεί ένα αυτοτελές κείμενο (Lauscher, 2000).

Πέρα από τις παραπάνω οπτικές που παρουσιάζει η Lauscher (2000) υπάρχει κι άλλος ένας διαχωρισμός των θεωριών για την αξιολόγηση της μετάφρασης, η Ψυχολογική-Κοινωνική προσέγγιση και η Κειμενοστρεφής προσέγγιση. Στην Ψυχολογική-Κοινωνική Προσέγγιση διακρίνονται δυο θεωρίες, η πνευματική, η οποία βασίζεται στο πόσο καλή ή κακή μπορεί να φανεί σε κάποιον μια μετάφραση. Ωστόσο, αυτή η θεωρία δεν βασίζεται σε συγκεκριμένους κανόνες αλλά σε εντυπώσεις και επίσης, εδώ η μετάφραση εξαρτάται από την αντίληψη και τις εμπειρίες του μεταφραστή (Stolze, 2003). Η άλλη θεωρία βασίζεται στην ανταπόκριση. Στο πλαίσιο αυτής της θεωρίας υπάρχουν οι ερευνητές που δίνουν βάση στη συμπεριφορά του κοινού-στόχου, δηλαδή οι συμπεριφοριστές (behaviorists) και εκείνοι που υπερασπίζονται τη θεωρία του σκοπού, δηλαδή οι λειτουργιστές (functionalists) (House, 2014). Οι πρώτοι θεωρούν ότι απαραίτητη

προϋπόθεση για να θεωρηθεί ένα μετάφρασμα καλό είναι να προκαλεί τα ίδια συναισθήματα στο κοινό-στόχος όπως και το πρωτότυπο στο κοινό, στο οποίο απευθύνεται (Nida, 1964). Από την άλλη, οι λειτουργιστές βασίζονται στη λειτουργικότητα της μετάφρασης και αποστασιοποιούνται από το πρωτότυπο. Σκοπός της μετάφρασης είναι να ανταποκρίνεται στις ανάγκες του κοινού-στόχος παρά να διατηρηθεί ο σκοπός του πρωτότυπου κειμένου και στο μετάφρασμα (Reiß and Vermeer, 1984). Όσον αφορά την Κειμενοστρεφή προσέγγιση εδώ συναντάμε τρεις θεωρίες, α) τις περιγραφικές μεταφραστικές μελέτες, β) την μεταμοντέρνα – αποδομημένη προσέγγιση, και γ) την γλωσσολογική προσέγγιση (House, 2014). Στις περιγραφικές μεταφραστικές μελέτες η αξιολόγηση της μετάφρασης γίνεται από την οπτική γωνία του αποδέκτη και αφορά τη δομή και τη λειτουργία του μεταφράσματος στον πολιτισμό-στόχο. Υπάρχει ισοδυναμία ανάμεσα σε ΚΠ και ΚΣ, η οποία όμως δεν είναι ένα προς ένα, αλλά είναι πλήθος σχέσεων υπό συγκεκριμένες συνθήκες. Ωστόσο, το μετάφρασμα δεν είναι ένα αυτοτελές κείμενο στη ΓΣ, αλλά συνδέεται με ένα προϋπάρχον κείμενο που απευθύνεται σε έναν άλλον πολιτισμό (Toury, 1995). Ο Venuti (2004), όπως κι άλλοι ερευνητές της μεταμοντέρνας προσέγγισης, υποστηρίζουν ότι στη μετάφραση παίζει ρόλο ο παράγοντας του κειμενικού είδους, καθώς σε κάποια είδη γίνεται αλλοίωση του ΚΠ, στο οποίο μπορεί να εκφράζονται κάποιες ιδεολογίες, και να μεταφέρονται διαφορετικά στο ΚΣ, ώστε να ικανοποιηθούν πιθανά συμφέροντα. Για να αξιολογήσει κανείς μια μετάφραση χρειάζεται πολύ καλή κριτική ικανότητα και διεξαγωγή πολλαπλών αναλύσεων. Είναι επίσης απαραίτητο να εξετάσει κανείς ποιες γλωσσικές επιλογές έγιναν στο πρωτότυπο, ποια η σκοπιμότητά τους, τι έγινε με αυτά κατά τη μεταφορά τους στον πολιτισμό-στόχο και ποιος είναι ο αντίκτυπος που έχουν στο κοινό. Η γλωσσολογική προσέγγιση έχει αποτελέσει θέμα πολλών ερευνών, όπως των Hatim and Mason (1997), Doherty (2002), Hatim and Munday (2004), Steiner (2004), Teich (2004), Baker (2011) και άλλων. Στη συγκεκριμένη προσέγγιση, έχουν αναπτυχθεί λεπτομερείς διαδικασίες για ανάλυση και αξιολόγηση και πολλές από αυτές λαμβάνουν υπόψη την σύνδεση κειμένου και συγκεκριμένου, δεδομένης της στενής σύνδεσης γλώσσας και πραγματικού κόσμου (House, 2014).

6. Προσεγγίσεις αξιολόγησης της ποιότητας της μετάφρασης

Όπως έχει προαναφερθεί, οι μελετητές δεν έχουν καταλήξει σε μια κοινή θεωρία για την αξιολόγηση της ποιότητας της μετάφρασης, ούτε υπάρχει κι ένας καθολικός και κοινώς αποδεκτός τρόπος διαχωρισμού των προσεγγίσεων γύρω από την αξιολόγηση, Ωστόσο, ας δούμε αναλυτικά παρακάτω ορισμένες από τις επικρατέστερες προσεγγίσεις αξιολόγησης της ποιότητας της μετάφρασης.

ο Εμπειρική προσέγγιση (Experience-based approach)

Οι εμπειρική προσέγγιση της αξιολόγησης της ποιότητας της μετάφρασης συνήθως δημιουργείται από κάποιον οργανισμό ή εταιρία για την αξιολόγηση της ποιότητας των μεταφράσεων τους (π.χ. LISA QA model). Αυτού του είδους η προσέγγιση είναι ακριβής για τους σκοπούς του οργανισμού, ωστόσο είναι δύσκολο να προσαρμοστεί σε διαφορετικά περιβάλλοντα, καθώς δεν ενέχει ήδη διαμορφωμένα θεωρητικά πλαίσια. Αυτό είναι περιοριστικό, διότι δεν μπορεί να γίνει αναφορά στη θεωρία, ενώ υπάρχει και ο κίνδυνος το αποτέλεσμα της αξιολόγησης να είναι επηρεασμένο από την υποκειμενική κρίση του αξιολογητή, λόγω της έλλειψης πλαισίου.

ο Θεωρητική και ερευνητική προσέγγιση (Theoretical and research-based approach)

Η θεωρητική και ερευνητική προσέγγιση ή αλλιώς προσέγγιση βάσει της ισοδυναμίας (equivalence based approaches) (Lauscher, 2000) αφορά τον αποδέκτη της μετάφρασης και το κείμενο. Αυτού του είδους η προσέγγιση προκύπτει από ένα θεωρητικό πλαίσιο ή μια υπόθεση για τη φύση της μετάφρασης. Ωστόσο, υπάρχουν και αρκετές ανακρίβειες, οι οποίες δεν εξασφαλίζουν την επιτυχία της και τη γενική εφαρμογή της, αλλά επικεντρώνεται σε ορισμένες πτυχές της ποιότητας, ενώ η αξιολόγηση πρέπει να περιλαμβάνει μεγάλος εύρος θεμάτων.

ο Προσέγγιση βάσει της αντίδρασης του αναγνώστη (Reader-response approach)

Αυτή η προσέγγιση αξιολογεί την ποιότητα της μετάφρασης σύμφωνα με την αντίδραση του αποδέκτη της μετάφρασης στον πολιτισμό-στόχος σε σύγκριση με αυτή του αποδέκτη του πρωτοτύπου στον πολιτισμό-

πηγή κατά την ανάγνωση (Nida 1964, Carroll 1966, Nida and Taber 1969). Η συγκεκριμένη προσέγγιση είναι προβληματική, επειδή η αντίδραση του αναγνώστη δεν αποτελεί σημαντικό κριτήριο σε όλα τα κειμενικά είδη, όπως γι' αυτά του δεν απευθύνονται στο αναγνωστικό κοινό (π.χ. κάποια είδη νομικών κειμένων).

ο Κειμενική και πραγματολογική προσέγγιση (Textual and pragmatic approach)

Η κειμενική και πραγματολογική προσέγγιση είναι αρκετά σημαντική, καθώς από την καταμέτρηση λαθών στο επίπεδο λέξης/πρότασης το ενδιαφέρον στρέφεται στην αξιολόγηση ολόκληρων κειμένων και στον στόχο της μετάφρασης, με πια επίκεντρο τον αναγνώστη και την επικοινωνία. Και σ' αυτήν την κατηγορία κάποια μοντέλα συγκρούονται, είτε επειδή ορισμένα επικεντρώνονται περισσότερο στο ΚΠ (Reiß, 1971) δίνοντας μεγαλύτερη βάση στον κειμενικό είδος και σκοπό του πρωτοτύπου, οπότε και η ποιότητα πρέπει να υπολογίζεται βάσει αυτού, είτε επειδή επικεντρώνονται περισσότερο στο ΚΣ (Reiß and Vermeer 1984, Nord 1997), και κατά συνέπεια με τη θεωρία του σκοπού η ποιότητα αξιολογείται σύμφωνα με το κειμενικό είδος και τον σκοπό της μετάφρασης.

ο Λειτουργική προσέγγιση (Functional – Componential approach)

Σύμφωνα με την Colina (2008), δεν υπάρχει κάποια ισορροπία μεταξύ θεωρίας και πράξης ανάμεσα στις διάφορες προσεγγίσεις αξιολόγησης της ποιότητας της μετάφρασης, γι' αυτό και προτείνει μια προσέγγιση, κατάλληλη τόσο για τον επαγγελματικό όσο και για τον ερευνητικό τομέα.

Συγκεκριμένα, προτείνει τις εξής αρχές για την αξιολόγηση:

- Θεωρητικά μοντέλα και προτάσεις που είναι επιβεβαιώσιμες και θέτουν καθαρά ερευνητικά ερωτήματα και υποθέσεις
- Θεωρητικά και εμπειρικά εργαλεία αξιολόγησης με ξεκάθαρα τοποθετημένες υποθέσεις για θεωρητικά ή εμπειρικά μοντέλα
- Προτάσεις ή εργαλεία για αξιολόγηση, τα οποία θέτουν τους στόχους και τα όριά τους
- Μοντέλα – προτάσεις που αναγνωρίζουν πολλές πλευρές της ποιότητας

Η συγκεκριμένη προσέγγιση προσπαθεί να υπερβεί τα όρια που έχουν θέσει οι άλλες προσεγγίσεις, με τους εξής τρόπους: α) βασίζεται στη λειτουργιστική θεωρία και στα κειμενικά μοντέλα της μετάφρασης, β) περιλαμβάνει υπόθεση, η οποία μπορεί να δοκιμαστεί, γ) δηλώνει το περιεχόμενο και τους στόχους της, δ) είναι τμηματική, καθώς περιλαμβάνει κατηγορίες για αξιολόγηση (π.χ. γλωσσολογική μορφή, λειτουργική ακρίβεια κ.α.).

Οι παραπάνω λειτουργική προσέγγιση της Colina (2008) για την αξιολόγηση της ποιότητας της μετάφρασης μπορεί να διαχωριστεί και ως εξής: α) την ολιστική και β) την αναλυτική προσέγγιση, ή αλλιώς ανάλυση βάσει των λαθών. Η ολιστική προσέγγιση εξετάζει το κείμενο στο σύνολό του και βασίζεται σε καθολικά κριτήρια, ενώ η αναλυτική προσέγγιση εξετάζει μεμονωμένα λάθη μέσα σε ένα κείμενο. Με τις δυο αυτές προσεγγίσεις ένα κείμενο εξετάζεται με δύο διαφορετικούς τρόπους, είτε δίνοντας μια καθολική εικόνα του κειμένου και αν αυτό πληροί τα ειδικά κριτήρια (ολιστική προσέγγιση), είτε εντοπίζοντας τα συγκεκριμένα λάθη ενός κειμένου, χωρίς όμως να παρέχεται μια γενικότερη εντύπωση του μεταφράσματος (αναλυτική προσέγγιση) (Lommel and Melby, 2018).

7. Ποιότητα Μηχανικής Μετάφρασης

Η ποιότητα της MM είναι εξίσου σημαντική με την ποιότητα της μετάφρασης γενικότερα και η αξιολόγησή της δεν διαφέρει πολύ σε σχέση με την αξιολόγηση της ποιότητας μιας μετάφρασης που παράγεται από τον άνθρωπο. Υπάρχουν, ωστόσο, ορισμένες διαφορές στη διαδικασία, λόγω της διαφορετικής φύσης της μετάφρασης. Ανάλογα το είδος του, κάθε σύστημα έχει συγκεκριμένη συμπεριφορά στα αποτελέσματα που παράγει, βάσει του κειμενικού είδους, του γλωσσικού συνδυασμού κ.α.. Η αξιολόγηση της ποιότητας των

συστημάτων MM γίνεται είτε ακαδημαϊκούς - ερευνητικούς σκοπούς, είτε για την βελτίωση των ίδιων των συστημάτων και την αξιολόγηση της λειτουργικότητάς τους. Είναι σημαντικό ένα σύστημα να αξιολογείται συστηματικά, ώστε να βελτιώνεται η απόδοσή του, ενώ και στην περίπτωση ενός νέου συστήματος, πρέπει να εξακριβώνεται ότι αποδίδει καλύτερα σε σχέση με το προηγούμενο (Snover, 2011).

Για την αξιολόγηση της ποιότητας της MM υπάρχουν δύο προσεγγίσεις: α) η λεγόμενη «Glass Box» και β) η λεγόμενη «Black Box». Η πρώτη προσέγγιση υπολογίζει την ποιότητα ενός συστήματος βάσει των εσωτερικών λειτουργιών του, ενώ η δεύτερη υπολογίζει την ποιότητα βάσει μόνο του αποτελέσματος της MM. Η δεύτερη προσέγγιση, εντοπίζει μόνο γλωσσολογικά προβλήματα του εκάστοτε συστήματος και εστιάζει στη συμπεριφορά ενός συστήματος σε ένα συγκεκριμένο κείμενο που εξυπηρετεί τους σκοπούς της αξιολόγησης. Η συγκεκριμένη προσέγγιση έχει εφαρμοστεί αρκετά για την σύγκριση της ποιότητας διαφόρων συστημάτων, ώστε να διαπιστωθεί ποιο σύστημα έχει την καλύτερη απόδοση. Αυτό για να επιτευχθεί, πρέπει τα κείμενα, που θα χρησιμοποιηθούν, να είναι του ίδιου χαρακτήρα και να χρησιμοποιηθούν για όλα τα συστήματα υπό εξέταση.

Όσον αφορά την προσέγγιση «Black Box», εδώ υπάρχουν δυο διαφορετικές μετρικές που υπολογίζουν την ακρίβεια και τη χρησιμότητα του αποτελέσματος της MM, η εσωτερική (intrinsic) και η εξωτερική (extrinsic). Η εσωτερική μετρική εστιάζει στην ποιότητα του αποτελέσματος της MM συγκρίνοντας το συχνά με μια ή περισσότερες μεταφράσεις αναφοράς, οι οποίες θεωρούνται υψηλής ποιότητας. Οι ανθρώπινες εσωτερικές μετρικές αξιολογούν την ποιότητα με αντικειμενικότητα, βάσει κάποιων χαρακτηριστικών του αποτελέσματος της MM, συνήθως υπολογίζοντας την ακρίβεια (accuracy) και την ευχέρεια (fluency). Οι αυτόματες εσωτερικές μετρικές, αντίθετα, υπολογίζουν την ομοιότητα των προτάσεων ανάμεσα στο αποτέλεσμα της MM με τις μεταφράσεις αναφοράς (μια ή περισσότερες ανάλογα τη μετρική). Η εξωτερική μετρική εστιάζει σε μια συγκεκριμένη εντολή που έχει δοθεί και βάσει αυτής αξιολογεί την ποιότητα. Σκοπός της είναι να εξετάσει την αποτελεσματικότητα του συστήματος MM σύμφωνα με αυτή τη συγκεκριμένη εργασία.

Πέρα από τα παραπάνω, υπάρχει ακόμα ένας διαχωρισμός της αξιολόγησης της ποιότητας των αποτελεσμάτων της MM. Ο διαχωρισμός αυτός αφορά την αξιολόγηση βάσει μεταφράσεων αναφοράς (Reference-Based MT evaluation) και την αξιολόγηση χωρίς μεταφράσεις αναφοράς (Reference-Less MT evaluation) (Lavie, 2011). Στην πρώτη περίπτωση η αξιολόγηση γίνεται βάσει μεταφράσεων του πρωτοτύπου από τον άνθρωπο και αποτελεί μια μέθοδο που χρησιμοποιείται συχνά, ενώ στη δεύτερη περίπτωση, η αξιολόγηση γίνεται χωρίς μεταφράσεις αναφοράς (Bojar et al. 2015).

Συνοπτικά, λοιπόν, έχουν εφαρμοστεί διάφοροι τρόποι αξιολόγησης της ποιότητας της MM, ωστόσο παρατηρούμε ότι χωρίζονται σε δυο βασικές κατηγορίες: α) την ανθρώπινη και β) την αυτόματη αξιολόγηση. Η ανθρώπινη αξιολόγηση εμφανίστηκε λίγο αργότερα από τη μηχανική μετάφραση με σκοπό την αξιολόγηση της ποιότητας των αποτελεσμάτων των αυτόματων μηχανών μετάφρασης. Αρχικά, για την ανθρώπινη αξιολόγηση χρησιμοποιούσαν υποκειμενικές απόψεις επισημειωτών για να βαθμολογήσουν την ποιότητα της MM, ως προς την ακρίβεια και την ευχέρεια (Snover, 2011), καθώς ειδικά εκπαιδευμένοι για αξιολόγηση άνθρωποι καλούνταν να κρίνουν την ποιότητα της μετάφρασης συγκριτικά με μεταφράσεις που είχαν παραχθεί από επαγγελματίες μεταφραστές.

8. Μέθοδοι αξιολόγησης της ποιότητας της Μηχανικής Μετάφρασης

Ξεκινώντας από την ανθρώπινη αξιολόγηση, μιας και χρονικά προηγείται της αυτόματης, ας δούμε παρακάτω αναλυτικά ορισμένες μεθόδους αξιολόγησης της ποιότητας της Μηχανικής Μετάφρασης.

i. Ακρίβεια και Ευχέρεια

Στην κατηγορία της ανθρώπινης αξιολόγησης MM δυο είναι οι πιο συνήθεις μετρικές, η ακρίβεια (accuracy) και η ευχέρεια (fluency) (White et al., 1994). Υπολογίζοντας την ακρίβεια παρατηρείται αν το αποτέλεσμα της MM περιέχει τις απαραίτητες πληροφορίες σύμφωνα με το πρωτότυπο. Κατά τον υπολογισμό της ακρίβειας, ο επισημειωτής δεν χρειάζεται να λαμβάνει υπόψη αν το αποτέλεσμα της MM είναι γραμματικά και συντακτικά σωστό, αλλά να επικεντρώνεται στη σωστή μεταφορά του νοήματος. Ο επισημειωτής στην

προκειμένη περίπτωση μπορεί να είναι δίγλωσσος ή μονόγλωσσος και θα πρέπει να γνωρίζει καλά τη ΓΠ και τη ΓΣ για να μπορεί να εντοπίσει αν μεταφέρεται σωστά το νόημα. Αντίθετα, για την αξιολόγηση της ευχέρειας είναι αρκετός ένας ομιλητής της ΓΣ, καθώς σε αυτή την μετρική αξιολογείται η γραμματικότητα του αποτελέσματος της MM. Συγκεκριμένα, ελέγχεται αν το αποτέλεσμα της MM είναι άρτιο σε γραμματικό και συντακτικό επίπεδο, ανεξάρτητα με αν το νόημα του πρωτοτύπου έχει αποτυπωθεί και στο μετάφρασμα που προέρχεται από MM. Συνηθέστερα, η αξιολόγηση της ακρίβειας και της ευχέρειας γίνεται σε επίπεδο πρότασης, δηλαδή η κάθε πρόταση αξιολογείται ξεχωριστά, σε μια κλίμακα 5 (ή και λιγότερων ή περισσότερων) βαθμίδων, για το κατά πόσο είναι ακριβής νοηματικά ή σωστή γραμματικά και συντακτικά αντίστοιχα (Przybocki et al., 2008).

ii. Κατάταξη

Άλλη μια μέθοδος αξιολόγησης της ποιότητας των αποτελεσμάτων της MM είναι η κατάταξη (ranking) (Callison-Burch et al., 2007). Σ' αυτήν την περίπτωση, οι επισημειωτές καλούνται να κατατάξουν αντιπαραβολικά διάφορες μεταφράσεις και πάλι σε επίπεδο πρότασης. Αυτή η μέθοδος αξιολόγησης θα μπορούσε να θεωρηθεί μια απλοποιημένη μορφή της αξιολόγησης της ακρίβειας και της ευχέρειας που είδαμε παραπάνω, με τη διαφορά ότι χρησιμοποιείται συγκριτικά με άλλες μεταφράσεις. Η μέθοδος αυτή, συνήθως, χρησιμοποιείται συμπληρωματικά, διότι τις περισσότερες φορές οι επισημειωτές δυσκολεύονται να διαχωρίσουν τα κριτήρια που απαιτούνται για να αξιολογήσουν την ακρίβεια και την ευχέρεια ξεχωριστά.

Σύμφωνα με τα όσα είδαμε στην αξιολόγηση της ποιότητας της μετάφρασης παραπάνω, οι δύο μέθοδοι αξιολόγησης της MM που είδαμε έως τώρα ανήκουν στην ολιστική μέθοδο αξιολόγησης, μιας και αφορούν την αξιολόγηση του κειμένου σαν σύνολο. Για την αναλυτική μέθοδο της αξιολόγησης της ποιότητας της MM συχνά χρησιμοποιείται η κατηγοριοποίηση των λαθών (error categorization) βάσει κάποιων προτύπων. Μερικά από αυτά είναι η τυπολογία λαθών «MQM (Multidimensional Quality Metrics)» και η τυπολογία λαθών «DQF (Dynamic Quality Framework)» που θα χρησιμοποιηθούν και στην παρούσα εργασία συνδυαστικά.

iii. Κατηγοριοποίηση των λαθών

Σε αντίθεση με άλλες ολιστικές μεθόδους που διαμορφώνουν μια γενική εικόνα της ποιότητας των αποτελεσμάτων, η αναλυτική αυτή μέθοδος αξιολόγησης δίνει μια λεπτομερή εικόνα της απόδοσης ενός συστήματος (Klubička et al., 2018). Με αυτή τη μέθοδο εντοπίζονται τα δυνατά και αδύναμα σημεία ενός συστήματος ανάλογα με το ζεύγος γλωσσών. Γλώσσες με ποικιλία γραμματικών, λεξιλογικών και συντακτικών φαινομένων, για παράδειγμα, παρουσιάζουν πλήθος λαθών ανά σύστημα, γεγονός που απαιτεί μια αναλυτική μέθοδο αξιολόγησης, έτσι ώστε να εντοπιστούν.

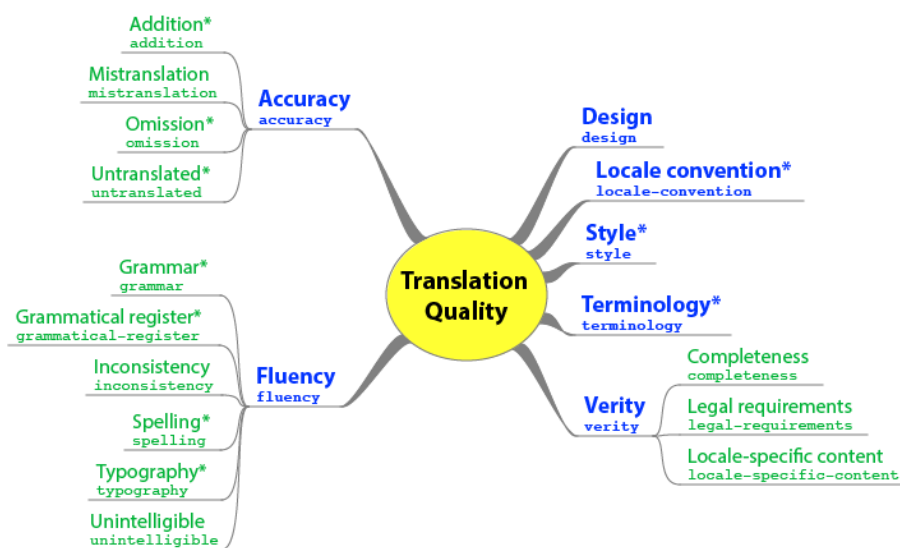
ο Πρότυπο MQM

Η τυπολογία λαθών MQM είναι μια πρωτοβουλία του Γερμανικού Ερευνητικού Κέντρου Τεχνητής Νοημοσύνης (DFKI) σε συνεργασία με το QTLaunchPad. Το συγκεκριμένο πρότυπο αφορά την αξιολόγηση της ποιότητας της μετάφρασης κειμένων ή ακόμα και των ίδιων των πρωτοτύπων, καθώς το πρωτότυπο επηρεάζει το μετάφρασμα.

Η τυπολογία λαθών MQM αποτελείται από 9 βασικές κατηγορίες:

1. Ακρίβεια (Accuracy): αξιολόγηση της σχέσης πρωτοτύπου και μετάφρασης. Σ' αυτή την κατηγορία συμπεριλαμβάνεται και η αλλοίωση του νοήματος, οι προσθήκες ή οι παραλήψεις, και άλλα παρόμοια λάθη.
2. Μορφή (Design): αξιολόγηση της φυσικής παρουσίας του κειμένου.
3. Ευχέρεια (Fluency): αξιολόγηση της γραμματικότητας ενός κειμένου. Αυτή η κατηγορία μπορεί να εξεταστεί ανεξάρτητα με το αν είναι μετάφραση ενός κειμένου ή όχι. Η συγκεκριμένη κατηγορία μπορεί να αφορά και ένα ΚΠ και ένα ΚΣ.

4. Διεθνοποίηση (Internationalization): εντοπισμός λαθών που αφορούν την προετοιμασία ενός ΚΠ για μετάφραση ή τοπική προσαρμογή.
5. Εγχώριες Συμβάσεις (Locale Conventions): συμμόρφωση περιεχομένου με συγκεκριμένες εγχώριες συμβάσεις. Τα λάθη αυτής της κατηγορίας δηλαδή αφορούν την περίπτωση που ένα μετάφρασμα είναι σωστό γλωσσικά και νοηματικά, αλλά δεν περιέχει τις απαραίτητες εγχώριες συμβάσεις που ορίζονται στην μεταφραστική εντολή. Στην κατηγορία αυτή δεν περιλαμβάνεται η αξιολόγηση του περιεχόμενου ως προς την καταλληλότητα του για το κοινό-στόχος.
6. Ύφος (Style): αν και κατά πόσο το κείμενο ακολουθεί τις επίσημες και ανεπίσημες οδηγίες που αφορούν το ύφος ενός κειμένου. Η συγκεκριμένη κατηγορία συνδέεται στενά με την κατηγορία της «Ευχέρειας», παρ' όλα αυτά, αξιολογείται ξεχωριστά.
7. Ορολογία (Terminology): αξιολογείται η σωστή χρήση της ορολογίας σε συγκεκριμένα είδη κειμένων και η συνέπεια χρήσης των συγκεκριμένων όρων σε όλο το κείμενο. Αυτή η κατηγορία αφορά λάθη στην ειδική ορολογία του κειμένου.
8. Εγκυρότητα (Verity): σωστή μεταφορά του περιεχομένου από τον ένα πολιτισμό στον άλλο. Αυτή η κατηγορία δεν έχει να κάνει με την ευχέρεια ή την ακρίβεια αλλά με το αν το περιεχόμενο είναι κατάλληλο για το κοινό-στόχος. Για παράδειγμα, αν ένα πρωτότυπο αγγλικό κείμενο αναφέρει στοιχεία που βρίσκονται μόνο στο Ηνωμένο Βασίλειο, είναι αστοχία να μεταφραστούν στην ελληνική γλώσσα για το ελληνικό κοινό.
9. Άλλο (Other): λάθη που δεν μπορούν να συμπεριληφθούν σε κάποια από τις υπόλοιπες κατηγορίες.



Εικόνα 1: Απεικόνιση της τυπολογίας λαθών MQM

Οι κατηγορίες Μορφή (Design), Ευχέρεια (Fluency), Ύφος (Style), Ορολογία (Terminology) και Εγκυρότητα (Verity) μπορούν να χρησιμοποιηθούν για την αξιολόγηση και του ΚΠ και του ΚΣ. Η Ακρίβεια (Accuracy) αφορά μόνο μεταφράσματα, καθώς αφορά τη σχέση ΚΠ και ΚΣ. Οι Εγχώριες Συμβάσεις (Locale Conventions) αφορούν περισσότερο το ΚΣ, ενώ στη Διεθνοποίηση (Internationalization) τα λάθη που εμπίπτουν βρίσκονται στο ΚΠ.

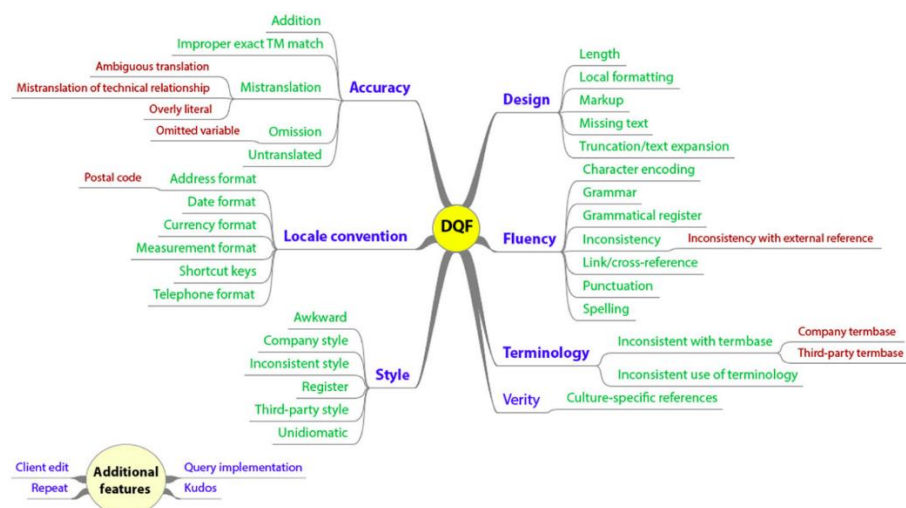
Στην Εικόνα 1 απεικονίζονται αναλυτικά οι 9 κατηγορίες για την αξιολόγηση της ποιότητας MM και οι υποκατηγορίες τους. Σύμφωνα με το παραπάνω σχήμα, στην κατηγορία της Ακρίβειας (Accuracy) συμπεριλαμβάνονται τα εξής: Προσθήκες (Additions), Παρανοήσεις (Mistranslations), Παραλήψεις (Omissions) και Αμετάφραστο κείμενο (Untranslated). Όλες αυτές οι υποκατηγορίες λαθών συμβάλουν στην αξιολόγηση της σωστής μεταφοράς του νοήματος. Στην κατηγορία της Ευχέρειας (Fluency) ανήκουν τα

λάθη που εντοπίζονται στη Γραμματική (Grammar), Γραμματικό Ύφος (Grammatical Register), Ασυνέπεια (Inconsistency), Ορθογραφία (Spelling), Τυπογραφικά λάθη (Typography) και Ακατανόητα λάθη (Unintelligible). Τέλος, στην κατηγορία της Εγκυρότητας (Verity) υπάρχουν τρεις υποκατηγορίες λαθών: Πληρότητα (Completeness), Νομικές Προϋποθέσεις (Legal requirements) και Τοπικά σχετικό περιεχόμενο (Locale-specific content).

ο Πρότυπο DQF

Το πρότυπο «DQF (Dynamic Quality Framework)» είναι προϊόν της TAUS (Translation Automation User Society) και είναι ένα αναγνωρισμένο υποσύνολο του προτύπου MQM, στο οποίο έχουν προστεθεί κατηγορίες από τα μέλη της TAUS. Στην παρακάτω εικόνα (Εικόνα 2) απεικονίζεται το πρότυπο DQF με τις επιπρόσθετες κατηγορίες.

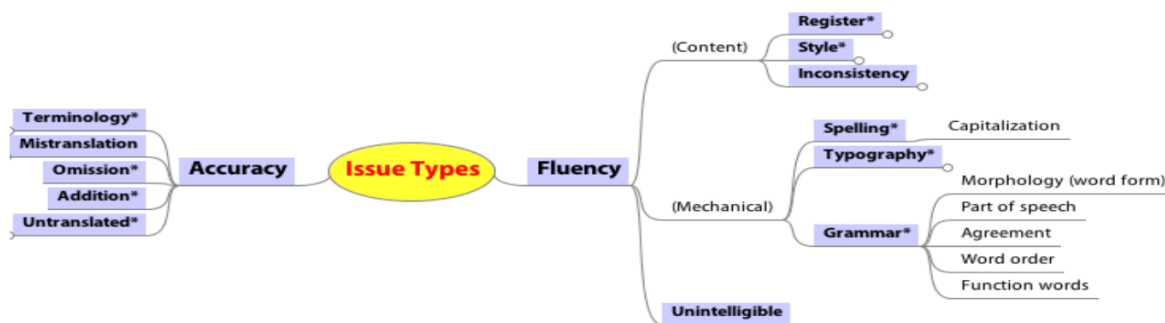
Στο συγκεκριμένο πρότυπο έχουν προστεθεί επίσης κάποια Επιπλέον Χαρακτηριστικά (Additional features), τα οποία έχουν θετική επίδραση στην αξιολόγηση, διότι αφορούν ειδικές περιπτώσεις στην μεταφραστική διαδικασία, π.χ. οδηγίες του πελάτη.



Εικόνα 2: Απεικόνιση της τυπολογίας λαθών DQF

ο Εναρμονισμένο πρότυπο MQM-DQF

Οι Lommel και Melby (2018) προτείνοντας μια σύζευξη των δυο προτύπων καταλήγουν σε ένα εναρμονισμένο πρότυπο τυπολογίας λαθών. Το συγκεκριμένο πρότυπο περιλαμβάνει μέρη και των δύο προτύπων (MQM και DQF) και στοχεύει στην αναγνώριση της φύσης των λαθών, ώστε να διορθωθούν ή να αποφευχθούν εντελώς. Ένα υποσύνολο αυτού του προτύπου αφορά την αξιολόγηση της ποιότητας της MM, λόγω της ευελιξίας στις κατηγορίες των λαθών και είναι και αυτό που θα χρησιμοποιηθεί στην παρούσα μελέτη. Στην Εικόνα 4 παρουσιάζεται το υποσύνολο που αφορά την αξιολόγηση των αποτελεσμάτων της MM. Το πρότυπο χωρίζει τα λάθη σε δυο μεγάλες κατηγορίες, την Ακρίβεια (Accuracy) και την Ευχέρεια (Fluency). Στην Ακρίβεια (Accuracy) περιέχονται τα λάθη Ορολογίας (Terminology), οι Παρανοήσεις (Mistranslations), οι Παραλήψεις (Omissions), οι Προσθήκες (Additions) και το Αμετάφραστο κείμενο (Untranslated). Στην κατηγορία της Ευχέρειας (Fluency) περιέχονται οι υπόλοιπες κατηγορίες που έχουν να κάνουν με τη γραμματικότητα ενός κειμένου, όπως είδαμε και στο πρότυπο MQM με την προσθήκη της κατηγορίας του Ύφους (Style), το οποίο ενώ στο MQM ήταν μια ξεχωριστή κατηγορία, πλέον αποτελεί υποκατηγορία της Ευχέρειας (Fluency) σ' αυτό το εναρμονισμένο πρότυπο.



Εικόνα 3: Απεικόνιση εναρμονισμένου προτύπου τυπολογίας λαθών MQM-DQF

9. Inter-annotator agreement

Στην περίπτωση της ανθρώπινης αξιολόγησης και καθώς επαναλαμβάνεται η διαδικασία της αξιολόγησης της ποιότητας των μεταφράσεων της MM, υπάρχει αρκετή απόκλιση ανάμεσα στις αξιολογήσεις των επισημειωτών (Przybowski et al., 2008). Σε αυτό συμβαίνει σημαντικά ο παράγοντας της υποκειμενικότητας των ανθρώπων κατά τη διαδικασία. Στην περίπτωση που υπάρχουν πολλαπλοί επισημειωτές, οι οποίοι δίνουν διαφορετικά αποτελέσματα, τότε η απόκλιση τους υπολογίζεται με τη διαδικασία του «inter-annotator agreement». Στην περίπτωση που γίνονται ετεροχρονισμένες αξιολογήσεις, για να υπολογιστεί η απόκλιση σε αυτά χρησιμοποιείται η διαδικασία «intra-annotator agreement». Ουσιαστικά, και στις δυο περιπτώσεις υπολογίζονται κατά μέσο όρο οι συσχετιζόμενες αξιολογήσεις των επισημειωτών. Υπάρχουν τρεις βασικοί τρόποι υπολογισμού ανάλογα με τα δεδομένα: α) η συσχέτιση του Pearson (Pearson's correlation) που χρησιμοποιείται όταν τα δεδομένα είναι αριθμητικά, β) το Cohen's Kappa που χρησιμοποιείται όταν τα αποτελέσματα προέρχονται από δυο επισημειωτές (Cohen, 1960) και γ) το Fleiss' Kappa, για να υπολογιστεί η συμφωνία περισσότερων από δυο επισημειωτών (Fleiss, 1969) (Fleiss, 1971) [οι δύο τελευταίοι τρόποι χρησιμοποιούνται για τα κατηγοριοποιημένα δεδομένα που βρίσκονται σε ονομαστική κλίμακα (nominal scale)].

10. Αυτόματη Αξιολόγηση της ποιότητας της Μηχανικής Μετάφρασης

Πέρα από τις διάφορες τεχνικές που έχουν αναλυθεί παραπάνω για την εξασφάλιση της ποιότητας της μηχανικής μετάφρασης, υπάρχει και η αυτόματη αξιολόγηση. Η αυτόματη αξιολόγηση ξεκίνησε σαν διαδικασία μεταγενέστερα και γίνεται με αυτόματες μετρικές. Αυτές οι αυτόματες μετρικές δημιουργήθηκαν για να μειωθεί το κόστος και ο όγκος της εργασίας που απαιτεί η ανθρώπινη αξιολόγηση, ώστε να είναι ευκολότερο να επαναλαμβάνεται συχνότερα η διαδικασία, για να εξασφαλιστεί η αντικειμενικότητα κατά την αξιολόγηση, να μειωθεί ο χρόνος που χρειάζεται η ανθρώπινη αξιολόγηση και να γίνει αυτοματοποίηση της διαδικασίας (Och, 2003). Αυτές οι μετρικές αξιολογούν την ποιότητα συγκρίνοντας τα αποτελέσματα της MM με μια ή περισσότερες μεταφράσεις αναφοράς των δεδομένων του πρωτοτύπου.

Η αυτόματη αξιολόγηση συστημάτων είναι η διαδικασία κατά την οποία αξιολογείται η μετάφραση που προκύπτει από συστήματα μηχανικής μετάφρασης χωρίς την ανθρώπινη παρέμβαση κατά την αξιολόγηση ως διαδικασία και η μόνη παρέμβαση που γίνεται από τον άνθρωπο είναι μόνο για την ανάπτυξη του συστήματος (Χατζηκουμή, 2015). Σύμφωνα με τους Banerjee and Lavie (2005) τα συστήματα αυτόματης αξιολόγησης πρέπει να τηρούν κάποιες προϋποθέσεις, έτσι ώστε να είναι λειτουργικά και αυτές είναι:

- οι μετρικές να έχουν υψηλή συσχέτιση με ποσοτικοποιημένες ανθρώπινες θεωρήσεις για την ποιότητα της MM,
- να λαμβάνουν υπόψη, όσο περισσότερο γίνεται, τις διαφοροποιήσεις της ποιότητας της MM ανάλογα το σύστημα ή τις διαφορετικές εκδόσεις ενός συστήματος, με το οποίο έχει γίνει μετάφραση,

- να υπάρχει συνέπεια αποτελεσμάτων. Σε διαφορετικά κείμενα που ανήκουν στο ίδιο κειμενικό είδος θα πρέπει να υπάρχουν παρόμοια αποτελέσματα ανά είδος συστήματος.
- να είναι αξιόπιστα. Αν το κάθε σύστημα έχει παρόμοιες βαθμολογίες, τότε μπορούμε να συμπεράνουμε ότι αποδίδει την ανάλογη ποιότητα κάθε φορά.
- να καλύπτει ένα ευρύ φάσμα πεδίων και σεναρίων.

Η πλειοψηφία των αυτόματων μετρικών χρησιμοποιεί για την αξιολόγηση της ποιότητας μία τεχνική, η οποία βασίζεται στην απόσταση της διόρθωσης (edit distance), επίσης γνωστή ως απόσταση κατά Levenshtein (Levenshtein distance) (Levenshtein, 1966). Για τον υπολογισμό της απόστασης διόρθωσης αναζητείται ο ελάχιστος αριθμός ενεργειών, έτσι ώστε να δυο προτάσεις να γίνουν ίδιες. Κατά τον Levenshtein (1966), οι ενέργειες που απαιτούνται για τη διόρθωση είναι οι προσθήκες (insertions), οι διαγραφές (deletions) και οι αντικαταστάσεις (substitutions). Στην αξιολόγηση της μηχανικής μετάφρασης, ουσιαστικά, η διόρθωση γίνεται στο επίπεδο της λέξης και αφορά τις απαραίτητες διορθώσεις που πρέπει να γίνουν, ώστε από το αποτέλεσμα της MM να φτάσουμε στη μετάφραση αναφοράς (Fordyce, 2007).

Δύο είναι οι κυριότερες μετρικές κατά την αυτόματη αξιολόγηση: η ακρίβεια (precision) και η ανάκληση (recall). Αυτές οι δυο μετρικές χρησιμοποιούνται κυρίως για την ανάκτηση πληροφοριών, ωστόσο, εφαρμόζονται και στην αυτόματη αξιολόγηση της μηχανικής μετάφρασης (Lavie, 2011). Η ακρίβεια υπολογίζεται από τα «σωστά» n-grams (= ακολουθία n-μονάδων) προς το σύνολο των n-grams του αποτελέσματος της MM και η ανάκληση από τα «σωστά» n-grams προς το σύνολο των n-grams της μετάφρασης-αναφοράς (Lavie, 2011). Ουσιαστικά, στην πρώτη περίπτωση υπολογίζεται το ποσοστό των παραχθειςών λέξεων που είναι σωστές, ενώ στην άλλη πόσες από τις επιθυμητές λέξεις παρήχθησαν.

Υπάρχουν πολλά και διαφορετικά είδη αυτόματων μετρικών, συμπεριλαμβανομένων πλέον και των μετρικών που βασίζονται σε νευρωνικά δίκτυα. Παρακάτω αναφέρονται ορισμένα παραδείγματα και αναλυτικά οι μετρικές που θα χρησιμοποιηθούν στην παρούσα εργασία: BLEU, chrF και TER.

Η μετρική Word Error Rate (WER) είναι η πιο συνηθισμένη για την αξιολόγηση της ποιότητας της αυτόματης αναγνώρισης ομιλίας και μια από τις πρώτες μετρικές που χρησιμοποιήθηκαν στην αξιολόγηση της αυτόματης μετάφρασης. Για τον υπολογισμό αυτής της μετρικής το αποτέλεσμα της MM συγκρίνεται με μία μετάφραση αναφοράς, η οποία είναι μια μετάφραση που γίνεται από τον άνθρωπο και θεωρείται υψηλής ποιότητας. Το WER υπολογίζεται όπως η απόσταση Levenshtein (Levenshtein, 1966) και ουσιαστικά εντοπίζει την καλύτερη αντιστοιχία κάθε λέξης του αποτελέσματος της MM με μία ή καμία λέξη της μετάφρασης αναφοράς και αντίστροφα. Στην περίπτωση που μία λέξη της μετάφρασης αναφοράς δεν αντιστοιχεί σε καμία λέξη της MM, τότε υπάρχει διαγραφή (deletion) και αντίστοιχα σε περίπτωση που μια λέξη του αποτελέσματος της MM δεν αντιστοιχεί σε καμία λέξη της μετάφρασης αναφοράς, τότε υπάρχει προσθήκη (insertion). Όταν μια λέξη από τη MM αντιστοιχεί στην ίδια λέξη της μετάφρασης αναφοράς υπάρχει αντιστοιχία (match), ενώ στην αντίθετη περίπτωση αντικατάσταση (substitution).

Μια εναλλακτική του WER, που εφαρμόζεται όταν υπάρχουν περισσότερες από μία μεταφράσεις αναφοράς, είναι το MWER (Multi-Reference WER), το οποίο εξετάζει το αποτέλεσμα της MM ως προς κάθε μετάφραση αναφοράς (Nießsen et al., 2000). Ουσιαστικά, το MWER κάνει τους ίδιους υπολογισμούς με το WER ανάμεσα στο κείμενο που προκύπτει από MM με την πιο κοντινή σε αυτό μετάφραση αναφοράς, καθώς η κάθε μετάφραση αναφοράς είναι ανεξάρτητη με την άλλη. Άλλος ένας τύπος WER είναι το PER (Position-Independent Error Rate), όπου οι λέξεις από το ένα κείμενο αντιστοιχούν σε λέξεις του άλλου ανεξάρτητα από τη θέση που βρίσκονται. Κανένας από τους τύπους του WER δεν αποτελεί, ωστόσο, κατάλληλη μετρική για την αξιολόγηση μεταφράσεων, καθώς όλοι οι παραπάνω εμφανίζουν αδυναμίες στο σύνολο τους.

Η μετρική METEOR (Metric for Evaluation of Translation with Explicit Ordering) συνδυάζει μετρικές ακρίβειας, ανάκλησης και ευθυγράμμισης. Λαμβάνει υπόψη το stemming και τη συνωνυμία και ενσωματώνει μια μέθοδο για τη ρητή αξιολόγηση της σειράς των λέξεων, παρέχοντας μια πιο ολοκληρωμένη αξιολόγηση από ορισμένες άλλες μετρικές. Η μετρική METEOR υπολογίζει έναν αρμονικό μέσο όρο της ακρίβειας και της ανάκλησης, με βάρη για το stemming και τη συνωνυμία (Banerjee, 2005).

Η διαφορετική μορφή των φράσεων/προτάσεων λόγω των διαφορετικών θέσεων που μπορούν να έχουν οι λέξεις, σύμφωνα με τους κανόνες της εκάστοτε γλώσσας, είναι μια δυσκολία που εμφανίζεται στις περισσότερες αυτόματες μετρικές μιας και δεν μπορούν να εντοπίσουν τις διαφοροποιήσεις αυτές. Μια μετρική που σχεδιάστηκε με σκοπό να καλύψει αυτή την αδυναμία είναι το GTM (General Text Matcher) (Turian, 2003). Η συγκεκριμένη μετρική επιχειρεί να μοντελοποιήσει τις εναλλαγές στη θέση των λέξεων κατά τη μετάφραση, χρησιμοποιώντας το μέγιστο αριθμό αντιστοιχιών (maximum matching size), για να υπολογίζεται καλύτερα η ποιότητα της μετάφρασης.

Η μετρική ROUGE (Recall-Oriented Understudy for Gisting Evaluation) αρχικά σχεδιάστηκε για την περίληψη κειμένων, ωστόσο, χρησιμοποιείται επίσης για την αξιολόγηση της μηχανικής μετάφρασης. Αξιολογεί τον αριθμό των n-grams και άλλων μονάδων (όπως ακολουθίες λέξεων) μεταξύ των μεταφράσεων αναφοράς και των αποτελεσμάτων MM. Υπολογίζει την ακρίβεια, την ανάκληση και το F1-score με βάση τις αντιστοιχισμένες μονάδες.⁵

Η μετρική CIDEr (Αξιολόγηση περιγραφής εικόνας με βάση τη συναίνεση) αναπτύχθηκε αρχικά για τη δημιουργία λεζάντας εικόνας και προσαρμόστηκε για την αξιολόγηση της μηχανικής μετάφρασης. Λαμβάνει υπόψη τη συναίνεση σε πολλαπλές μεταφράσεις αναφοράς και δίνει έμφαση στην ποικιλομορφία των υποψήφιων μεταφράσεων. Η μετρική CIDEr υπολογίζει μια βαθμολογία ομοιότητας με βάση τα n-grams και τη στάθμιση IDF (Inverse Document Frequency).⁶

Η μετρική BLEU (Bilingual Evaluation Understudy) είναι αυτή που έχει χρησιμοποιηθεί περισσότερο απ' όλες και υπολογίζει την ακρίβεια (precision). Συγκεκριμένα, αυτή η μετρική «μετράει» τη διαφορά των n-grams μεταξύ των αποτελεσμάτων της MM και των μεταφράσεων αναφοράς, οι οποίες δημιουργούνται και σ' αυτήν την περίπτωση από τον άνθρωπο (Parineni et al., 2002). Το BLEU έχει ένα κοινό με το MWER κι αυτό είναι ότι μπορεί να αξιολογεί ως προς πολλαπλές μεταφράσεις αναφοράς. Η συγκεκριμένη μετρική υπολογίζει τον αριθμό των n-grams ή των ακολουθιών λέξεων (μέχρι 4 λέξεις), οι οποίες συμπίπτουν με τις μεταφράσεις αναφοράς. Στοχεύει περισσότερο στην ακρίβεια, και μετρά κατά πόσο το μετάφρασμα που έχει παραχθεί από τη MM είναι σωστό και όχι αν η μετάφραση αναφοράς είναι ίδια με το μετάφρασμα από τη MM. Ουσιαστικά, το BLEU υπολογίζει την εγγύτητα της μηχανικής μετάφρασης με τη μετάφραση αναφοράς μετρώντας τα n-grams της MM που συναντώνται σε οποιαδήποτε από τις μεταφράσεις αναφοράς προς τον συνολικό αριθμό των n-grams του αποτελέσματος της MM. Σε περίπτωση που το αποτέλεσμα της MM είναι μικρότερο από τις μεταφράσεις αναφοράς, τότε η μετρική πολλαπλασιάζει τα αποτελέσματα με μια ποινή συντομίας (brevity penalty). Η βαθμολογία που δίνει το BLEU κυμαίνεται από το 0 μέχρι το 1 με το 1 να είναι το καλύτερο, δηλαδή όσο πιο κοντά βρίσκεται η βαθμολογία στο 1, τόσο καλύτερο είναι το αποτέλεσμα της MM (Parineni et al., 2002). Η συγκεκριμένη μετρική λειτουργεί καλύτερα σε μεγάλα σώματα κειμένου, ώστε να υπολογίζεται ο μέσος όρος των βαθμολογιών για περισσότερες προτάσεις, ενώ για επιμέρους προτάσεις, η μετρική αυτή, θεωρείται αναξιόπιστη. Το NIST αποτελεί μια εναλλακτική έκδοση του BLEU, του οποίου η μόνη διαφορά είναι ότι δεν χρησιμοποιεί γεωμετρικούς υπολογισμούς όπως το BLEU, αλλά αριθμητικούς. Άλλες παραλλαγές του BLEU θεωρούνται μη αξιόπιστες, καθώς δεν μπορούν να εντοπίσουν την αντιστοιχία συνώνυμων λέξεων ανάμεσα στα μεταφράσματα και τις πιθανούς διαφορετικούς συνδυασμούς των λέξεων σε μια γλώσσα, όσον αφορά τη σειρά τους σε ένα κείμενο. Έτσι το BLEU παραμένει η μόνη μετρική προς σύγκριση των άλλων μετρικών που αναπτύσσονται (Snover et al., 2011).

Το TER (Translation Error Rate) είναι, επίσης, μια μετρική που επιχειρεί να λύσει το πρόβλημα της διαφορετικής σειράς των λέξεων συμπληρώνοντας την αδυναμία του WER, με το να επιτρέπει τις μετακινήσεις (shifts) λέξεων στο αποτέλεσμα της MM, χωρίς όμως να δίνεται κάποια βαρύτητα σε αυτές (Snover et al., 2006). Η συγκεκριμένη μετρική κάνει μια αναζήτηση για να συλλέξει όλες τις λέξεις, οι οποίες πρόκειται να μετακινηθούν και τους περεταίρω περιορισμούς των λέξεων που πρόκειται να μετακινηθούν.

⁵ Two minutes NLP — Learn the ROUGE metric by examples, Fabio Chiusano, 19/01/2022 medium.com (<https://medium.com/nlplanet/two-minutes-nlp-learn-the-rouge-metric-by-examples-f179cc285499>), accessed on: September 2023

⁶ Consensus-based Image Description Evaluation (CIDEr), OECD.AI, 2021 (<https://oecd.ai/fr/catalogue/metrics/consensus-based-image-description-evaluation-%28cider%29>), accessed on: September 2023

Αυτοί οι περιορισμοί μιμούνται, ουσιαστικά, τον τρόπο, με τον οποίο σκέφτεται ένας επιμελητής να επιλέξει τις λέξεις που πρέπει να μετακινηθούν. Συγκεκριμένα προκύπτουν τέσσερις τρόποι: α) οι μετακινήσεις επιλέγονται από έναν αλγόριθμο, ο οποίος επιλέγει αυτή την μετακίνηση που μειώνει περισσότερο το WER μεταξύ του αποτελέσματος της MM και της μετάφρασης αναφοράς, β) η ακολουθία των λέξεων του αποτελέσματος της MM πρέπει να μετακινηθεί έτσι ώστε να αντιστοιχίζεται ακριβώς στην ακολουθία των λέξεων της μετάφρασης-αναφοράς, γ) οι λέξεις που πρόκειται να μετακινηθούν πρέπει να περιέχουν τουλάχιστον ένα σφάλμα, σύμφωνα με το WER πριν μετακινηθούν, έτσι ώστε να αποφεύγεται η μετακίνηση των ήδη σωστών λέξεων και δ) οι λέξεις που αντιστοιχίζονται στη μετάφραση αναφοράς, οι οποίες μετακινήθηκαν πρέπει, επίσης, να περιέχουν τουλάχιστον ένα σφάλμα. Αυτό αποτρέπει τη μετακίνηση για να αντιστοιχηθούν λέξεις, οι οποίες είναι ήδη σωστά αντιστοιχημένες. Ο υπολογισμός του TER είναι το σύνολο των αλλαγών (προσθήκες, διαγραφές, αντικαταστάσεις και μετακινήσεις) προς τον μέσο όρο των λέξεων όλων των μεταφράσεων αναφοράς (Snover et al., 2011). Χαμηλότερη βαθμολογία TER υποδηλώνει και καλύτερη ποιότητα μετάφρασης.

Όπως όλες οι μετρικές που είδαμε μέχρι τώρα, έτσι κι αυτή έχει κάποιες αδυναμίες. Η αδυναμία της συγκεκριμένης μετρικής είναι ότι δεν διαχειρίζεται όλες τις μεταφράσεις αναφοράς, και ούτε έχει γλωσσολογική πληροφορία, όπως για παράδειγμα το METEOR με την αντιστοιχία συνωνύμων. Αυτές, λοιπόν, οι αδυναμίες του συστήματος διευθετούνται με μια παραλλαγή του TER, η οποία χρησιμοποιεί στοχευμένες μεταφράσεις αναφοράς που δημιουργούνται για ένα συγκεκριμένο αποτέλεσμα της MM από τον άνθρωπο. Αυτό μετατρέπει τη μετρική από αυτόματη σε ημι-αυτόματη, διότι η παρέμβαση του ανθρώπου είναι μεγαλύτερη από αυτή στις αυτόματες μετρικές και ονομάζεται HTER (Human-mediated Translation Error Rate) (Snover et al., 2006). Οι στοχευμένες αυτές μεταφράσεις αναφοράς δημιουργούνται από μονόγλωσσους επισημειωτές για ένα συγκεκριμένο σύστημα και αποτελούνται από τις λιγότερες δυνατές αλλαγές για να είναι γραμματικά σωστό, αλλά και να μεταφέρεται το νόημα των άλλων μεταφράσεων αναφοράς. Οι συγκεκριμένες μεταφράσεις μπορούν να χρησιμοποιηθούν και στις υπόλοιπες αυτόματες μετρικές. Παρ' όλα αυτά, επειδή αυτές οι μεταφράσεις αναφοράς έχουν δημιουργηθεί για ένα συγκεκριμένο σύστημα MM, δεν μπορούν να χρησιμοποιηθούν ως μεταφράσεις αναφοράς για αποτελέσματα άλλων συστημάτων MM, αλλά ούτε και για άλλες εκδόσεις του ίδιου τύπου συστημάτων. Παρόλο που το HTER έχει δημιουργηθεί βάσει της ανθρώπινης κρίσης, είναι μια καθαρά ποσοτική μετρική, η οποία βαθμολογεί ισάξια όλες τις διορθώσεις, ενώ κάποιες είναι περισσότερο σημαντικές από άλλες και μπορεί να αλλάζουν ριζικά το νόημα σε σχέση με άλλες πιο «ελαφριές» (Snover et al., 2011).

Το chrF (character F-score) είναι μια μετρική αξιολόγησης της μηχανικής μετάφρασης που επικεντρώνεται στην ομοιότητα σε επίπεδο χαρακτήρων μεταξύ της μετάφρασης αναφοράς και της μετάφρασης που έχει παραχθεί από ένα αυτόματο σύστημα MM. Έχει σχεδιαστεί για να είναι ανεξάρτητη από τη γλώσσα και ανθεκτική σε σενάρια όπου τα όρια των λέξεων μπορεί να είναι διαφορετικά. Η μετρική chrF βασίζεται στο F-score, το οποίο είναι ο αρμονικός μέσος όρος της ακρίβειας και της ανάκλησης. Στο πλαίσιο του chrF, η ακρίβεια και η ανάκληση υπολογίζονται σε επίπεδο χαρακτήρα. Το chrF «μετρά» τον αριθμό των χαρακτήρων που συμπίπτουν μεταξύ των μεταφράσεων αναφοράς και των αποτελεσμάτων MM. Λαμβάνει υπόψη τόσο την ακρίβεια (την αναλογία των σωστά προβλεφθέντων χαρακτήρων στην υποψήφια μετάφραση) όσο και την ανάκληση (την αναλογία των σωστά προβλεφθέντων χαρακτήρων στη μετάφραση αναφοράς). Το F-score υπολογίζεται στη συνέχεια ως ο αρμονικός μέσος όρος της ακρίβειας και της ανάκλησης, παρέχοντας μια ενιαία μετρική που εξισορροπεί και τις δύο πτυχές. Το chrF περιλαμβάνει μια ρυθμιζόμενη παράμετρο (beta) που επιτρέπει στους χρήστες να ελέγχουν την ισορροπία μεταξύ ακρίβειας και ανάκλησης με βάση τις συγκεκριμένες απαιτήσεις αξιολόγησής τους. Ένα από τα δυνατά σημεία του chrF είναι η ανεξαρτησία του από τις γλώσσες. Δεδομένου ότι λειτουργεί σε επίπεδο χαρακτήρων, επηρεάζεται λιγότερο από ζητήματα τμηματοποίησης λέξεων που μπορεί να προκύψουν σε γλώσσες με διαφορετικά μορφολογικά χαρακτηριστικά (Ρορονιό, 2015).

Υπάρχουν επίσης μετρικές που έχουν σχεδιαστεί ειδικά για την αξιολόγηση της ποιότητας των μηχανικών μεταφράσεων που παράγονται από νευρωνικά δίκτυα. Οι παραδοσιακές μετρικές όπως οι BLEU και METEOR χρησιμοποιούνται ευρέως, αλλά οι ερευνητές έχουν επίσης αναπτύξει μετρικές βασισμένες σε

νευρωνικά δίκτυα για την καλύτερη ευθυγράμμιση με τις δυνατότητες των νευρωνικών μοντέλων αυτόματης μετάφρασης.

Το ReVal (Gupta et al., 2015) είναι η πρώτη νευρωνική μετρική που προτείνεται για την αξιολόγηση της ποιότητας της μηχανικής μετάφρασης. Η μετρική ReVal αποτέλεσε σημαντική βελτίωση σε σχέση με τις παραδοσιακές μετρικές με σημαντικά καλύτερη συσχέτιση με την ανθρώπινη αξιολόγηση. Η μετρική βασίζεται σε ένα μοντέλο LSTM και είναι πολύ απλή, αλλά δεν έχει χρησιμοποιηθεί ποτέ στην έρευνα για τη μηχανική μετάφραση, ενώ έχει πια ξεπεραστεί από πιο πρόσφατες μετρικές.

Η μετρική YiSi (Chi-kiu Lo, 2019) είναι ένα πολύ πιο ευέλικτο. Εκμεταλλεύεται κυρίως ένα μοντέλο ενσωμάτωσης, αλλά μπορεί να εμπλουτιστεί με διάφορους πόρους, όπως ένας σημασιολογικός αναλυτής, ένα μεγάλο γλωσσικό μοντέλο (BERT), ακόμη και χαρακτηριστικά από το αρχικό κείμενο και τη γλώσσα προέλευσης. Η χρήση όλων αυτών των επιλογών μπορεί να την καταστήσει αρκετά πολύπλοκη και μειώνει το πεδίο εφαρμογής της σε λίγα ζεύγη γλωσσών. Επιπλέον, τα κέρδη όσον αφορά τη συσχέτιση με τις ανθρώπινες κρίσεις κατά τη χρήση όλων αυτών των επιλογών δεν είναι προφανή. Παρ' όλα αυτά, η ίδια η μετρική, χρησιμοποιώντας μόνο το αρχικό μοντέλο ενσωμάτωσης, παρουσιάζει πολύ καλή συσχέτιση με την ανθρώπινη αξιολόγηση.

Άλλη μετρική σχεδιασμένη ειδικά για την αξιολόγηση της ποιότητας των μηχανικών μεταφράσεων που παράγονται από νευρωνικά δίκτυα είναι το BERTScore. Το BERTScore είναι μια μετρική που χρησιμοποιεί προ-εκπαιδευμένες συγκεκριμενοποιημένες ενσωματώσεις από το BERT (Bidirectional Encoder Representations from Transformers) για τη μέτρηση της ομοιότητας μεταξύ των μεταφράσεων αναφοράς και των αποτελεσμάτων MM. Αυτή η μετρική λαμβάνει υπόψη τα συμφραζόμενα και τη σημασιολογική σημασία, καθιστώντας τη ιδιαίτερα κατάλληλη για την αξιολόγηση της ποιότητας των νευρωνικών μηχανικών μεταφράσεων (Zhang et al., 2020).

Στο έργο τους που προτείνει το Prism, οι Thompson και Post (2019) υποστηρίζουν διαισθητικά ότι η μηχανική μετάφραση και η αξιολόγηση της παράφρασης είναι πολύ παρόμοιες εργασίες. Η μόνη διαφορά τους είναι ότι η γλώσσα προέλευσης δεν είναι η ίδια. Το Prism είναι μια νευρωνική μετρική που εκπαιδεύεται σε ένα μεγάλο πολύγλωσσο παράλληλο σύνολο δεδομένων μέσω ενός πολύγλωσσου πλαισίου νευρωνικής μηχανικής μετάφρασης. Στη συνέχεια, κατά τη στιγμή της εξαγωγής συμπερασμάτων, το εκπαιδευμένο μοντέλο χρησιμοποιείται ως παραφραστική μηδενικού αποτελέσματος για να βαθμολογήσει την ομοιότητα μεταξύ ενός κειμένου πηγής (η μεταφραστική υπόθεση) και του κειμένου-στόχου (η μετάφραση αναφοράς) που είναι και τα δύο στην ίδια γλώσσα. Το κύριο πλεονέκτημα αυτής της προσέγγισης είναι ότι το Prism δεν χρειάζεται δεδομένα εκπαίδευσης ανθρώπινης αξιολόγησης ούτε δεδομένα εκπαίδευσης παραφράσεων. Η μόνη απαίτηση είναι να υπάρχουν παράλληλα δεδομένα για τις γλώσσες που θα αξιολογηθούν.

Η μετρική COMET (Cross-lingual Order-aware Multiple Edit distance for Translation) είναι άλλη μια μετρική που έχει σχεδιαστεί για την αυτόματη αξιολόγηση της μηχανικής μετάφρασης. Επικεντρώνεται στην καταγραφή διαφόρων πτυχών της ποιότητας της μετάφρασης, συμπεριλαμβανομένης της ευχέρειας, της πιστότητας στην πηγή και της επάρκειας. Η μετρική COMET συνδυάζει διάφορα βασικά στοιχεία, όπως η απόσταση επεξεργασίας, η σειρά των λέξεων και η διαγλωσσική ομοιότητα, για να παρέχει μια ολοκληρωμένη αξιολόγηση της ποιότητας της μηχανικής μετάφρασης. Επίσης, υπολογίζει τις διαγλωσσικές ενσωματώσεις για τη μέτρηση της σημασιολογικής ομοιότητας, λαμβάνοντας υπόψη τη σημασία των λέξεων τόσο στη γλώσσα προέλευσης όσο και στη γλώσσα-στόχο. Παρόμοια με άλλες μετρικές, η μετρική COMET παρέχει μια βαθμολογία για την ποσοτικοποίηση της ποιότητας της μηχανικής μετάφρασης. Υψηλότερη βαθμολογία COMET υποδηλώνει καλύτερη ποιότητα μετάφρασης (Rei et al., 2020).

Αυτές οι μετρικές αποσκοπούν στην αντιμετώπιση ορισμένων από τους περιορισμούς των παραδοσιακών μετρικών όταν εφαρμόζονται σε νευρωνικές μηχανικές μεταφράσεις. Οι μετρικές που βασίζονται σε νευρωνικά δίκτυα συχνά λαμβάνουν υπόψη τη σημασιολογική ομοιότητα, την ευχέρεια και την κατανόηση των συμφραζόμενων, οι οποίες αποτελούν κρίσιμες πτυχές για την αξιολόγηση της απόδοσης των σύγχρονων νευρωνικών συστημάτων αυτόματης μετάφρασης. Οι ερευνητές συνεχίζουν να διερευνούν και

να αναπτύσσουν νέες μετρικές για την καλύτερη αποτύπωση των περιπλοκών των μοντέλων νευρωνικής μετάφρασης.

Είναι σημαντικό να σημειωθεί ότι καμία μεμονωμένη μετρική δεν είναι τέλεια και ο συνδυασμός πολλαπλών μετρικών ή η χρήση τους σε συνδυασμό με την ανθρώπινη αξιολόγηση μπορεί να προσφέρει μια πιο ολοκληρωμένη αξιολόγηση της ποιότητας της μηχανικής μετάφρασης.

11. Πλατφόρμα αυτόματης αξιολόγησης MutNMT

Για τους σκοπούς της παρούσας εργασίας η αυτόματη αξιολόγηση των νευρωνικών συστημάτων MM έγινε χρησιμοποιώντας την πλατφόρμα MutNMT. Σύμφωνα με τους δημιουργούς της πλατφόρμας αυτής, η στρατηγική σύμπραξη «MultiTraiNMT - Εκπαίδευση στη Μηχανική Μετάφραση για πολύγλωσσους πολίτες» (MultiTraiNMT - “Machine Translation training for multilingual citizens”) έχει ως στόχο την ανάπτυξη, αξιολόγηση και διάδοση υλικού ανοικτής πρόσβασης που θα οδηγήσει στην ενίσχυση της διδασκαλίας και της μάθησης σχετικά με τη μηχανική μετάφραση μεταξύ μαθητών, καθηγητών ξένων γλωσσών, εκπαιδευόμενων μεταφραστών, καθηγητών μετάφρασης και επαγγελματιών μεταφραστών σε ολόκληρη την Ευρώπη.⁷

Το MultiTraiNMT συγκεντρώνει εμπειρογνώμονες από τέσσερα ευρωπαϊκά πανεπιστήμια – το Universitat Autònoma de Barcelona, το Université Grenoble-Alpes, το Dublin City University και το Universitat d'Alacant, και δύο επιχειρήσεις -- Prompsit Language Engineering και Xcelerator Machine Translations, και υποστηρίζεται από περισσότερους από είκοσι συνεργάτες στην εκπαίδευση και τη μεταφραστική βιομηχανία, οι οποίοι ενδιαφέρονται για τη διδασκαλία και τη μάθηση σχετικά με τη χρήση της νευρωνικής μηχανικής μετάφρασης. Η σύμπραξη στοχεύει συγκεκριμένα στην ανάπτυξη ενός καινοτόμου προγράμματος σπουδών στη μηχανική μετάφραση, και ειδικότερα στη μηχανική μετάφραση που βασίζεται σε δημοφιλείς σήμερα τεχνικές βαθιάς μάθησης, γνωστές και ως «νευρωνική μηχανική μετάφραση». Με την ολοκλήρωσή αυτής της πρωτοβουλίας θα έχουμε οδηγηθεί στην ανάπτυξη ενός σύγχρονου προγράμματος σπουδών στη μηχανική μετάφραση για χρήση στην ευρωπαϊκή τριτοβάθμια εκπαίδευση και αλλού, το οποίο θα επιτρέψει στους φοιτητές να αναπτύξουν τις τεχνικές και ηθικές δεξιότητες και ικανότητες που απαιτούνται για να γίνουν ενημερωμένοι, κριτικοί χρήστες της σύγχρονης μηχανικής μετάφρασης στη δική τους γλωσσική εκμάθηση και μεταφραστική πρακτική. Με τον τρόπο αυτό, θα έχουμε επίσης ανοίξει τον κόσμο της μηχανικής μάθησης στους φοιτητές γλωσσών και μετάφρασης, στους καθηγητές τους και σε άλλους, ενισχύοντας την ικανότητά τους να λειτουργούν ως τεχνολογικά ικανοί, ενημερωμένοι πολίτες σε μια πολύγλωσση Ευρώπη.⁸

Για την αξιολόγηση του κάθε νευρωνικού συστήματος MM, το MutNMT χρειάζεται το πρωτότυπο κείμενο, τη μετάφραση που έχει παραχθεί από ένα νευρωνικό σύστημα μηχανικής μετάφρασης και μία μετάφραση αναφοράς. Αυτό που υπολογίζει είναι οι μετρικές BLEU και TER και η αξιολόγηση γίνεται σε επίπεδο πρότασης. Με το συνδυασμό αυτών των δύο μετρικών έχουμε μία ολοκληρωμένη εικόνα και ένα πιο αξιολόγο αποτέλεσμα ως προς την ποιότητα της μετάφρασης του κάθε συστήματος.

12. Μεθοδολογία της έρευνας

Η παρούσα έρευνα επικεντρώνεται στην αυτόματη και ανθρώπινη αξιολόγηση ενός μεταφρασμένου κειμένου που παράγεται από δύο διαφορετικά συστήματα νευρωνικής μηχανικής μετάφρασης, για τα ζεύγη γλωσσών Αγγλικά-Ελληνικά και Αγγλικά-Ισπανικά. Στόχος είναι η σύγκριση των δύο διαφορετικών συστημάτων νευρωνικής μηχανικής μετάφρασης για τα προαναφερθέντα ζεύγη γλωσσών αντιπαραβάλλοντας τις μεταφράσεις που παράγονται από αυτά με μια ανθρώπινη μετάφραση και αξιολογώντας την ακρίβεια και την ευχέρεια των μεταφράσεων που παράγονται και από τα δύο. Η σύγκριση γίνεται τόσο με αυτόματες όσο και με ανθρώπινες μετρικές, ώστε να διαπιστωθεί ποιο από τα δυο συστήματα παράγει τα καλύτερα αποτελέσματα. Για την μελέτη αυτή χρησιμοποιήθηκε ένα κείμενο, το ίδιο και για τους δύο γλωσσικούς συνδυασμούς. Το κείμενο είναι ένα δελτίο τύπου που συζητά την επίτευξη της

⁷ A MultiTraiNMT outcome, MutNMT homepage (<https://ntradumatica.uab.cat/index>), accessed on: May 2023

⁸ Meet MutNMT – About the Project, MutNMT (<https://ntradumatica.uab.cat/about>), accessed on: May 2023

πιστοποίησης της Welocalize ως εικονικής τοποθεσίας σύμφωνα με το νέο μοντέλο ISO. Το περιεχόμενο υπογραμμίζει την επιτυχή εφαρμογή του νέου μοντέλου από την εταιρεία, δίνοντας έμφαση στη στροφή προς την πιστοποίηση διαδικασιών και συστημάτων ποιότητας που βασίζονται στο cloud αντί για φυσικές τοποθεσίες. Το κείμενο είναι πληροφοριακό. Μεταφέρει πραγματικές πληροφορίες σχετικά με την επίτευξη της πιστοποίησης εικονικής τοποθεσίας της Welocalize σύμφωνα με το νέο μοντέλο ISO, συζητώντας τις επιπτώσεις στην ασφάλεια δεδομένων, τις διαδικασίες ποιότητας και τη συμμόρφωση σε εικονικά περιβάλλοντα εργασίας. Όσον αφορά το ύφος, το κείμενο ακολουθεί ένα επίσημο και επαγγελματικό ύφος που συνήθως συνδέεται με εταιρικές ανακοινώσεις, δελτία τύπου ή επίσημες ανακοινώσεις. Η γλώσσα είναι δομημένη, αντικειμενική και επικεντρώνεται στην παροχή σαφών πληροφοριών σχετικά με το επίτευγμα της Welocalize και τη σημασία του. Ο τόνος είναι έγκυρος και στοχεύει να μεταδώσει την τεχνογνωσία και τα επιτεύγματα της εταιρείας στον τομέα των πιστοποιήσεων ISO και των εικονικών περιβαλλόντων εργασίας.

Η αξιολόγηση των αποτελεσμάτων της MM διεξήχθη με ένα συνδυασμό μεθόδων, έτσι ώστε να διασφαλισθεί η ποιότητα, και όσο είναι δυνατόν η αντικειμενικότητα κατά τη διαδικασία. Αρχικά, για την αυτόματη αξιολόγηση, όπως έχει προαναφερθεί χρησιμοποιήθηκε η πλατφόρμα MutNMT. Εκτός από τις δύο μεταφράσεις των νευρωνικών συστημάτων για την επιτυχή αξιολόγηση χρειάζεται και μία μετάφραση αναφοράς από επαγγελματία μεταφραστή. Για να χρησιμοποιηθεί το MutNMT τα κείμενα πρέπει να είναι σε μορφή αρχείου .txt με τις προτάσεις του κειμένου τη μία κάτω από την άλλη, έτσι ώστε και τα τρία αρχεία που θα χρησιμοποιηθούν για την αυτόματη αξιολόγηση να έχουν τον ίδιο αριθμό σειρών.

Η ανθρώπινη αξιολόγηση έγινε με τις μετρικές της κατάταξης, ακρίβειας και ευχέρειας. Εδώ χρησιμοποιήθηκε μια ολιστική μέθοδος που περιλαμβάνει την κατάταξη των αποτελεσμάτων από τα δυο συστήματα MM και την ανθρώπινη μετάφραση από το καλύτερο προς το χειρότερο. Καταρτίστηκαν τρία ξεχωριστά ερωτηματολόγια για κάθε ζεύγος γλωσσών. Ως αξιολογητές επιλέχθηκαν φυσικοί ομιλητές της ελληνικής και της ισπανικής γλώσσας, οι οποίοι, αν και δεν είναι απαραίτητα επαγγελματίες μεταφραστές, παρείχαν αυθεντικές γνώσεις. Κάθε ερωτηματολόγιο ζητά την κατάταξη των μεταφράσεων σε κλίμακα όπου το 1 σήμαινε την καλύτερη μετάφραση και το 3 τη χειρότερη. Συγκεκριμένα, για την κατάταξη (ranking) η εκφώνηση στα ερωτηματολόγια ήταν η εξής: «Section 1 - In this section you are kindly asked to rank from best to worst the translations given below, sentence by sentence, as follows: 1-Best and 3-Worst. Two of these are produced by two different neural machine translation (NMT) systems and one of them is a human translation» και για κάθε πρόταση «Please rank the following translations from best to worst for the English sentence». Για την αξιολόγηση της ακρίβειας και της ευχέρειας η εκφώνηση περιελάμβανε και μια σύντομη επεξήγηση των όρων, καθώς όπως σημειώθηκε παραπάνω οι επισημειωτές δεν ήταν απαραίτητα επαγγελματίες του κλάδου της μετάφρασης και γνώστες τέτοιας ορολογίας. Για την ακρίβεια (accuracy) η εκφώνηση στα ερωτηματολόγια ήταν η εξής: «Section 1 - In this section you are kindly asked to rate from best to worst the accuracy of the translations given below, sentence by sentence, as follows: 1-Best and 3-Worst. Two of these are produced by two different neural machine translation (NMT) systems and one of them is a human translation. Accuracy: How much of the meaning expressed in the source (English text) is also expressed in the target text (Greek translation) Accuracy issues address the relationship of the target text to the source text and can be assessed only by considering this relationship. Changes in intended meaning, addition and omission of content, and similar issues are considered in it» και για κάθε πρόταση: «Please rate from best to worst the accuracy the English sentence». Αντίστοιχα, για την ευχέρεια (fluency) η εκφώνηση στα ερωτηματολόγια ήταν η εξής: «Section 1 - In this section you are kindly asked to rate from best to worst the fluency of the translations given below, sentence by sentence, as follows: 1-Best and 3-Worst. Two of these are produced by two different neural machine translation (NMT) systems and one of them is a human translation. Fluency: To what extent the translation is one that is well-formed grammatically, contains correct spelling, adheres to common use of terms, titles and names, is intuitively acceptable and can be sensibly interpreted by a native speaker. Fluency includes those issues about the linguistic “well-formedness” of the text that can be assessed without regard to whether the text is a translation or not» και για κάθε πρόταση: «Please rate from best to worst the fluency of the English sentence».

Τα κριτήρια αξιολόγησης που επιλέχθηκαν αποσκοπούν στην καταγραφή διαφόρων πτυχών της ποιότητας της μετάφρασης. Η συνιστώσα της κατάταξης μετρά τη συνολική προτίμηση, η αξιολόγηση της ακρίβειας

εξετάζει την πιστότητα στο αρχικό κείμενο και η αξιολόγηση της ευχέρειας εξετάζει τη φυσικότητα και τη συνοχή των μεταφράσεων. Η μεθοδολογία αξιολόγησης ανά πρόταση διασφαλίζει μια λεπτομερή ανάλυση, επιτρέποντας την αποχρωματισμένη κατανόηση των δυνατών και αδύνατων σημείων κάθε συστήματος MM, καθώς και μια σύγκριση με τις μεταφράσεις που δημιουργούνται από τον άνθρωπο. Η χρήση φυσικών ομιλητών ως αξιολογητών προσθέτει μια πολύτιμη διάσταση, καταγράφοντας γλωσσικές αποχρώσεις και προτιμήσεις που μπορεί να μην είναι διακριτές μέσω αμιγώς αυτοματοποιημένων μετρήσεων. Ο ολοκληρωμένος χαρακτήρας αυτής της μεθοδολογίας αποσκοπεί στην παροχή μιας ολιστικής κατανόησης της συγκριτικής απόδοσης των δύο συστημάτων MM στα συγκεκριμένα ζεύγη γλωσσών.

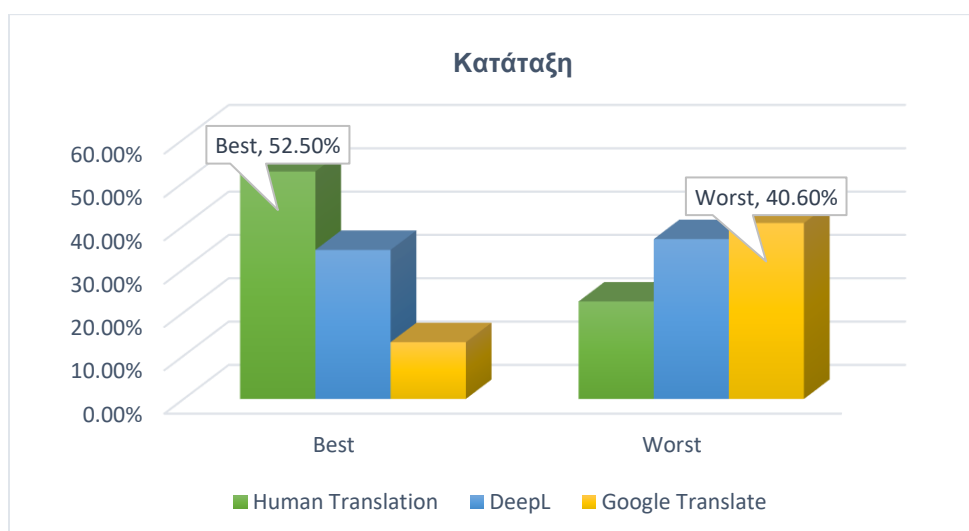
Οι επισημειωτές που κλήθηκαν να αξιολογήσουν τα αποτελέσματα της MM είναι φυσικοί ομιλητές της ελληνικής και της ισπανικής γλώσσας, και δεν είναι απαραίτητα επαγγελματίες μεταφραστές. Η διαδικασία της αξιολόγησης γίνεται με τη χρήση ερωτηματολογίων σε μορφή Google Forms, ώστε να είναι ευκολότερη η διαδικασία για τους επισημειωτές, τα οποία είναι ειδικά διαμορφωμένα για τους σκοπούς της παρούσας έρευνας. Η κατάθεση των ερωτηματολογίων έγινε ανώνυμα και δέκα επισημειωτές έλαβαν μέρος για κάθε γλωσσικό συνδυασμό. Για να ελεγχθεί αν τα αποτελέσματα που προέκυψαν είναι αντιπροσωπευτικά των συστημάτων υπολογίζεται το «Inter-annotator agreement», που δείχνει την συμφωνία των επισημειωτών, όπως είδαμε και στη θεωρία παραπάνω.

13. Αποτελέσματα της έρευνας

i. Ανθρώπινη αξιολόγηση

Η ανθρώπινη αξιολόγηση έγινε σε επίπεδο πρότασης με τις μετρικές της κατάταξης (ranking), ακρίβειας (accuracy) και ευχέρειας (fluency). Οι επισημειωτές έλαβαν ένα ερωτηματολόγιο με τρεις επιλογές για την κάθε πρόταση, μία μετάφραση του συστήματος DeepL, μία μετάφραση του συστήματος Google Translate καθώς και μία μετάφραση από επαγγελματία μεταφραστή και κλήθηκαν να τις κατατάξουν από την καλύτερη στη χειρότερη.

Για το γλωσσικό ζεύγος Αγγλικά – Ελληνικά στην κατηγορία της κατάταξης (ranking), το 52.5% των απαντήσεων καταχωρούν ως καλύτερη μετάφραση την ανθρώπινη μετάφραση, το 34.4% τη μετάφραση του DeepL και μόλις το 13.1% τη μετάφραση του Google Translate. Αντίστοιχα, σύμφωνα με τις απαντήσεις των επισημειωτών το Google Translate έχει κατά 40.6% χειρότερο αποτέλεσμα.



Γράφημα 1: Ποσοστιαία κατάταξη από το καλύτερο προς το χειρότερο των μεταφράσεων που έχουν παραχθεί από επαγγελματία μεταφραστή, το DeepL και το Google Translate για το γλωσσικό ζεύγος Αγγλικά - Ελληνικά

Η συμφωνία των επισημειωτών υπολογίστηκε με την μέθοδο του Inter-Annotator Agreement (IAA) και για την κατάταξη το μέτρο συμφωνίας Fleiss' Kappa βρίσκεται στο 0.167 που δείχνει ελαφρώς αυξημένη, αλλά

γενικά χαμηλή συμφωνία. Αυτό υποδηλώνει ότι, παρόλο που υπάρχει κάποια συνοχή στις αξιολογήσεις, οι αξιολογητές δεν συμφωνούν πλήρως στην κατάταξη των μεταφράσεων. Αντίστοιχα, τα μέτρα συμφωνίας Mean Pearson (0.228) και Cohen's Kappa (0.1875) υποδηλώνουν μέτρια συμφωνία, επιβεβαιώνοντας την ύπαρξη κάποιας συνέπειας στις αξιολογήσεις αλλά και την ανάγκη για περαιτέρω συμφωνία.

Για το γλωσσικό ζεύγος Αγγλικά – Ελληνικά στην κατηγορία της ακρίβειας (accuracy), το 52.5% των απαντήσεων καταχωρούν ως καλύτερη μετάφραση την ανθρώπινη μετάφραση, το 36.25% τη μετάφραση του DeepL και μόλις το 11.25% τη μετάφραση του Google Translate. Αντίστοιχα, σύμφωνα με τις απαντήσεις των επισημεωτών το Google Translate έχει κατά 55.6% χειρότερο αποτέλεσμα.



Γράφημα 2: Ποσοστιαία κατάταξη από το καλύτερο προς το χειρότερο των μεταφράσεων που έχουν παραχθεί από επαγγελματίες μεταφραστές, το DeepL και το Google Translate ως προς την ακρίβειά τους για το γλωσσικό ζεύγος Αγγλικά - Ελληνικά

Για την αξιολόγηση της ακρίβειας το μέτρο Fleiss' Kappa βρίσκεται στο 0.083. Αν και παρουσιάζει μια ελαφρώς υψηλότερη συμφωνία από την κατάταξη, παραμένει σε χαμηλά επίπεδα, αντικατοπτρίζοντας την ποικιλομορφία των απόψεων ανάμεσα στους αξιολογητές. Αντίστοιχα, τα μέτρα συμφωνίας Mean Pearson (0.178) και Cohen's Kappa (0.25) υποδηλώνουν μέτρια συμφωνία και σ' αυτήν την περίπτωση.

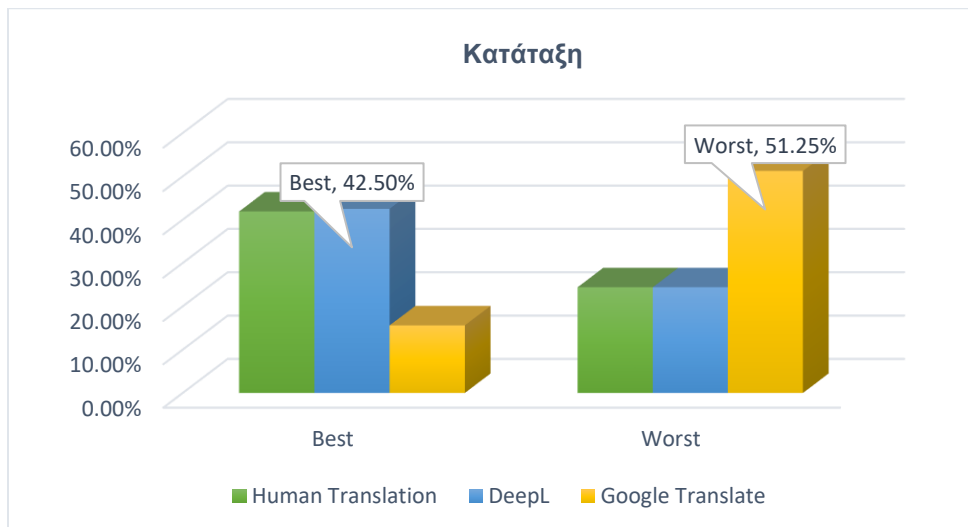
Για το γλωσσικό ζεύγος Αγγλικά – Ελληνικά στην κατηγορία της ευχέρειας (fluency), το 48.75% των απαντήσεων καταχωρούν ως καλύτερη μετάφραση την ανθρώπινη μετάφραση, το 38.75% τη μετάφραση του DeepL και μόλις το 12.5% τη μετάφραση του Google Translate. Αντίστοιχα, σύμφωνα με τις απαντήσεις των επισημεωτών το Google Translate έχει κατά 66.25% χειρότερο αποτέλεσμα.



Γράφημα 3: Ποσοστιαία κατάταξη από το καλύτερο προς το χειρότερο των μεταφράσεων που έχουν παραχθεί από επαγγελματίες μεταφραστές, το DeepL και το Google Translate ως προς την ευχέρειά τους για το γλωσσικό ζεύγος Αγγλικά - Ελληνικά

Για την αξιολόγηση της ευχέρειας, το μέτρο Mean Pearson βρίσκεται στο 0.203. Αυτό δείχνει μια θετική, αλλά σχετικά χαμηλή γραμμική σχέση, υποδηλώνοντας ότι οι βαθμολογίες για τη ρευστότητα των μεταφράσεων δεν είναι πλήρως συνεπείς. Αντίστοιχα, τα μέτρα συμφωνίας Mean Cohen's Kappa (0.080) και Fleiss' Kappa (0.097) υποδηλώνουν μια ελαφρώς πάνω από τυχαία, αλλά γενικά χαμηλή συμφωνία.

Για το γλωσσικό ζεύγος Αγγλικά – Ισπανικά στην κατηγορία της κατάταξης (ranking), το 42.5% των απαντήσεων καταχωρούν ως καλύτερη μετάφραση τη μετάφραση του DeepL, το 41.9% των απαντήσεων καταχωρούν ως την ανθρώπινη μετάφραση, και μόλις το 15.6% τη μετάφραση του Google Translate. Αντίστοιχα, σύμφωνα με τις απαντήσεις των επισημευτών το Google Translate έχει κατά 51.25% χειρότερο αποτέλεσμα.



Γράφημα 4: Ποσοστιαία κατάταξη από το καλύτερο προς το χειρότερο των μεταφράσεων που έχουν παραχθεί από επαγγελματίες μεταφραστές, το DeepL και το Google Translate για το γλωσσικό ζεύγος Αγγλικά - Ισπανικά

Για την κατάταξη το μέτρο συμφωνίας Fleiss' Kappa βρίσκεται στο 0.026. Αυτό επιβεβαιώνει την πολύ χαμηλή συμφωνία, δείχνοντας μεγάλη διαφοροποίηση στις αξιολογήσεις. Αντίστοιχα, το μέτρο Mean

Pearson βρίσκεται στο 0.090 υποδηλώνοντας χαμηλή συμφωνία και το μέτρο Cohen's Kappa βρίσκεται στο -0.156 και δηλώνει διαφωνία μεταξύ των δύο αξιολογητών.

Για το γλωσσικό ζεύγος Αγγλικά – Ισπανικά στην κατηγορία της ακρίβειας (accuracy), το 38.75% των απαντήσεων καταχωρούν ως καλύτερη μετάφραση την ανθρώπινη μετάφραση, το 30.6% τη μετάφραση του DeepL και το 30.6% τη μετάφραση του Google Translate. Αντίστοιχα, σύμφωνα με τις απαντήσεις των επισημεωτών το Google Translate έχει κατά 39.4% χειρότερο αποτέλεσμα.



Γράφημα 5: Ποσοστιαία κατάταξη από το καλύτερο προς το χειρότερο των μεταφράσεων που έχουν παραχθεί από επαγγελματίες μεταφραστές, το DeepL και το Google Translate ως προς την ακρίβειά τους για το γλωσσικό ζεύγος Αγγλικά - Ισπανικά

Για την αξιολόγηση της ακρίβειας τα μέτρα Fleiss' Kappa και Mean Pearson βρίσκονται στο -0.002. Αυτό υποδηλώνει εξαιρετικά χαμηλή συμφωνία, σχεδόν τυχαία, ενώ ο Cohen's Kappa (0.156) δείχνει μια ελαφρώς καλύτερη συμφωνία.

Για το γλωσσικό ζεύγος Αγγλικά – Ισπανικά στην κατηγορία της ευχέρειας (fluency), το 50.6% των απαντήσεων καταχωρούν ως καλύτερη μετάφραση την ανθρώπινη μετάφραση, το 36.9% τη μετάφραση του DeepL και μόλις το 12.5% τη μετάφραση του Google Translate. Αντίστοιχα, σύμφωνα με τις απαντήσεις των επισημεωτών το Google Translate έχει κατά 61.25% χειρότερο αποτέλεσμα.



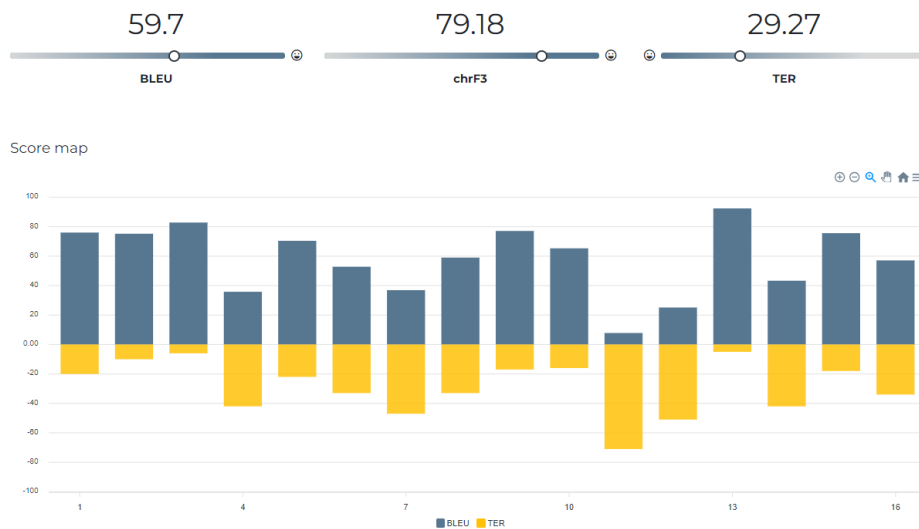
Γράφημα 6: Ποσοστιαία κατάταξη από το καλύτερο προς το χειρότερο των μεταφράσεων που έχουν παραχθεί από επαγγελματίες μεταφραστή, το DeepL και το Google Translate ως προς την ευχέρειά τους για το γλωσσικό ζεύγος Αγγλικά – Ισπανικά

Για την αξιολόγηση της ευχέρειας, το μέτρο Fleiss' Kappa βρίσκεται στο 0.279. Αν και υψηλότερος από τους άλλους δείκτες, παραμένει κάτω από μέτριο, υποδηλώνοντας περιορισμένη συμφωνία στην αξιολόγηση της ρευστότητας των μεταφράσεων. Αντίστοιχα, το μέτρο Mean Pearson (0.319) δηλώνει μέτρια θετική γραμμική σχέση και το μέτρο Cohen's Kappa (0.0625) δηλώνει ελαφρώς πάνω από τυχαία αλλά γενικά χαμηλή συμφωνία.

ii. Αυτόματη αξιολόγηση

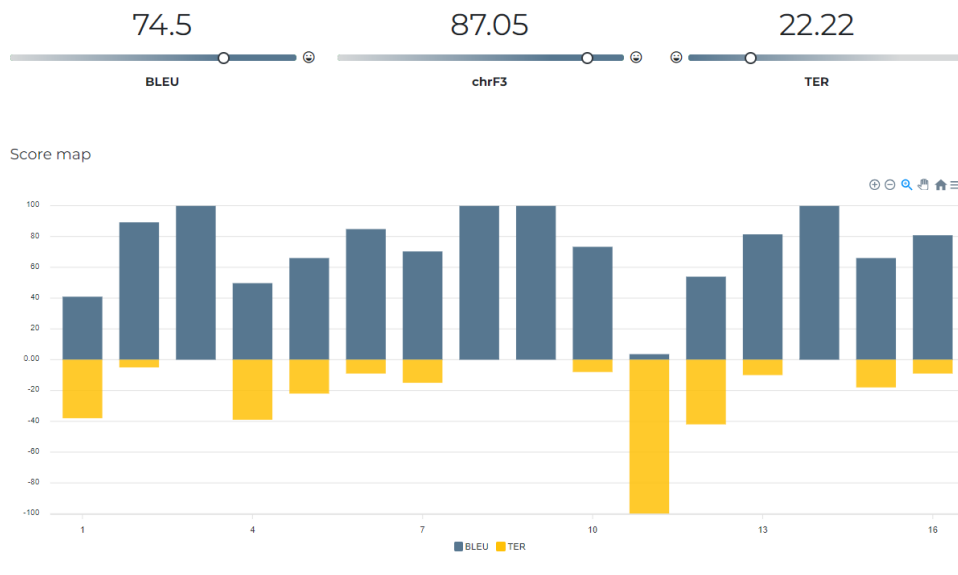
Για την αυτόματη αξιολόγηση, όπως έχει προαναφερθεί χρησιμοποιήθηκε η πλατφόρμα MutNMT. Οι αυτόματες μετρικές που χρησιμοποιούνται για την αυτόματη αξιολόγηση στη συγκεκριμένη πλατφόρμα είναι οι μετρικές BLEU, chrF3 και TER. Αφού έγινε η απαραίτητη προετοιμασία των αρχείων, έγινε η αξιολόγηση ξεχωριστά για κάθε γλωσσικό συνδυασμό και σύστημα μετάφρασης.

Για το γλωσσικό συνδυασμό Αγγλικά – Ελληνικά και το νευρωνικό σύστημα μετάφρασης DeepL, το MutNMT έδειξε τα εξής αποτελέσματα: BLEU = 59.7, chrF3 = 79.18 και TER = 29.27.



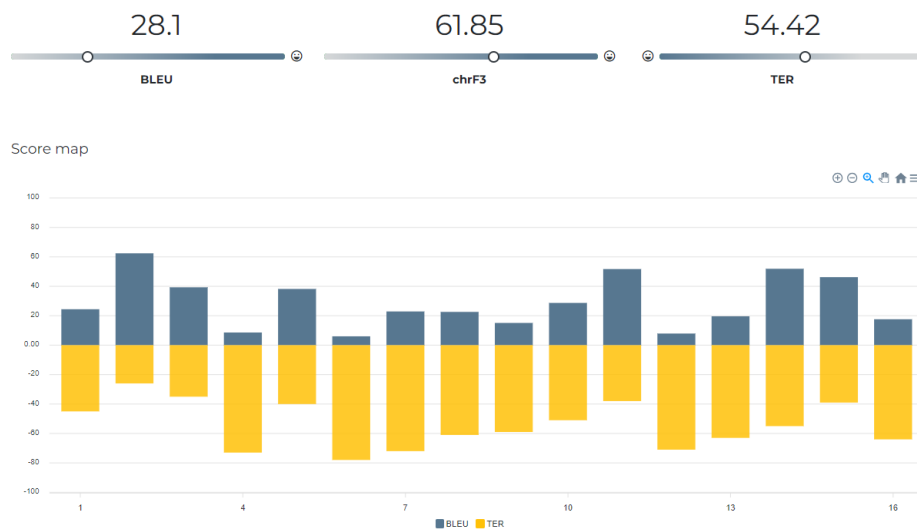
Γράφημα 7: Αποτελέσματα μετρικών BLEU, chrF3 και TER για το γλωσσικό συνδυασμό Αγγλικά – Ελληνικά και το νευρωνικό σύστημα μετάφρασης DeepL

Για το γλωσσικό συνδυασμό Αγγλικά – Ελληνικά και το νευρωνικό σύστημα μετάφρασης Google Translate, το MutNMT έδειξε τα εξής αποτελέσματα: BLEU = 74.5, chrF3 = 87.05 και TER = 22.22.



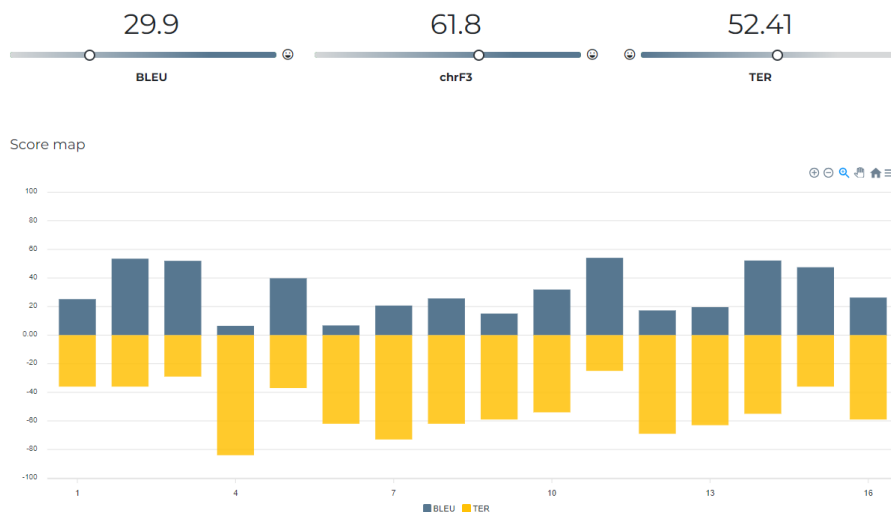
Γράφημα 8: Αποτελέσματα μετρικών BLEU, chrF3 και TER για το γλωσσικό συνδυασμό Αγγλικά – Ελληνικά και το νευρωνικό σύστημα μετάφρασης Google Translate

Για το γλωσσικό συνδυασμό Αγγλικά – Ισπανικά και το νευρωνικό σύστημα μετάφρασης DeepL, το MutNMT έδειξε τα εξής αποτελέσματα: BLEU = 28.1, chrF3 = 61.85 και TER = 54.42.



Γράφημα 9: Αποτελέσματα μετρικών BLEU, chrF3 και TER για το γλωσσικό συνδυασμό Αγγλικά – Ισπανικά και το νευρωνικό σύστημα μετάφρασης DeepL

Για το γλωσσικό συνδυασμό Αγγλικά – Ισπανικά και το νευρωνικό σύστημα μετάφρασης Google Translate, το MutNMT έδειξε τα εξής αποτελέσματα: BLEU = 29.9, chrF3 = 61.8 και TER = 52.41.



Γράφημα 10: Αποτελέσματα μετρικών BLEU, chrF3 και TER για το γλωσσικό συνδυασμό Αγγλικά – Ισπανικά και το νευρωνικό σύστημα μετάφρασης Google Translate

14. Συζήτηση αποτελεσμάτων

Σύμφωνα με τα παραπάνω, παρατηρούμε ότι τόσο η μέθοδος αξιολόγησης όσο και ο γλωσσικός συνδυασμός επηρεάζουν το αποτέλεσμα. Όσον αφορά τη σύγκριση ανάμεσα στις γλώσσες, για το γλωσσικό ζεύγος Αγγλικά – Ελληνικά τα μέτρα συμφωνίας (Fleiss' Kappa, Mean Pearson, Cohen's Kappa) δείχνουν γενικά μέτρια έως χαμηλή συμφωνία, με ελαφρώς αυξημένες τιμές σε σύγκριση με το γλωσσικό ζεύγος Αγγλικά – Ισπανικά. Για το γλωσσικό ζεύγος Αγγλικά – Ισπανικά παρατηρούμε χαμηλότερα επίπεδα

συμφωνίας σε όλες τις μετρήσεις, υποδηλώνοντας περισσότερη διακύμανση στις αξιολογήσεις των επισημειωτών.

Σχετικά με τη σύγκριση μεταξύ των μετρικών, η κατάταξη στο γλωσσικό ζεύγος Αγγλικά – Ελληνικά έχει καλύτερη συμφωνία σε σύγκριση με το με το γλωσσικό ζεύγος Αγγλικά – Ισπανικά, υποδηλώνοντας ότι οι αξιολογητές συμφωνούν περισσότερο στην κατάταξη των μεταφράσεων στα Ελληνικά παρά στα Ισπανικά. Η αξιολόγηση της ακρίβειας δείχνει μια ελαφρώς βελτιωμένη συμφωνία στο γλωσσικό ζεύγος Αγγλικά – Ελληνικά σε σύγκριση με την κατάταξη, ενώ στο γλωσσικό ζεύγος Αγγλικά – Ισπανικά η συμφωνία παραμένει εξαιρετικά χαμηλή. Στο γλωσσικό ζεύγος Αγγλικά – Ελληνικά, η αξιολόγηση της ευχέρειας δείχνει μια μέτρια γραμμική σχέση μεταξύ των βαθμολογιών των αξιολογητών, ενώ στο γλωσσικό ζεύγος Αγγλικά – Ισπανικά, παρά τη μέτρια τιμή του Pearson Correlation, η συμφωνία μεταξύ των αξιολογητών με βάση τον Fleiss' Kappa και τον Cohen's Kappa είναι χαμηλή.

Συνεπώς, παρατηρούμε ότι ενώ υπάρχει κάποια συνέπεια στις αξιολογήσεις ανάμεσα στους 10 αξιολογητές, η γενική συμφωνία είναι χαμηλή, ειδικά για τις μεταφράσεις στο γλωσσικό ζεύγος Αγγλικά – Ισπανικά. Η μεγαλύτερη συμφωνία που παρατηρήθηκε στο γλωσσικό ζεύγος Αγγλικά – Ελληνικά μπορεί να υποδηλώνει μια καλύτερη ομοιογένεια στην κατανόηση και την ερμηνεία των κριτηρίων αξιολόγησης ή στις προσδοκίες από τις μεταφράσεις. Η διαφορά στη συμφωνία ανάμεσα στις διαφορετικές μετρικές (ranking, accuracy, fluency) υποδεικνύει τις προκλήσεις στην ομοιόμορφη αξιολόγηση των διάφορων πτυχών της μετάφρασης και την ανάγκη για περαιτέρω εκπαίδευση ή εναρμόνιση των αξιολογητών.

Σχετικά με τη σύγκριση ανάμεσα στα δύο συστήματα, στην ανθρώπινη αξιολόγηση, όπου υπάρχει και η επιλογή της μετάφρασης από επαγγελματία μεταφραστή παρατηρούμε ότι στο γλωσσικό ζεύγος Αγγλικά – Ελληνικά η τελευταία αξιολογήθηκε ως καλύτερη, με δεύτερο στην κατάταξη το αποτέλεσμα του DeepL και τρίτο το Google Translate και στις τρεις περιπτώσεις. Αντίθετα, στο γλωσσικό ζεύγος Αγγλικά – Ισπανικά το DeepL αξιολογείται ως καλύτερο αποτέλεσμα στην πρώτη κατηγορία της κατάταξης, ενώ εδώ βλέπουμε επίσης και τα δύο νευρωνικά συστήματα να έχουν ίδιο ποσοστό κατά την αξιολόγηση της ακρίβειας. Επιπλέον, τόσο για το γλωσσικό ζεύγος Αγγλικά – Ελληνικά όσο και για το ζεύγος Αγγλικά – Ισπανικά η απόκλιση ανάμεσα στην ανθρώπινη μετάφραση και τη μετάφραση του DeepL είναι σχετικά μικρή στις περιπτώσεις της κατάταξης και της αξιολόγησης της ακρίβειας, ενώ κατά την αξιολόγηση της ευχέρειας παρατηρούμε ότι το Google Translate έχει χειρότερο αποτέλεσμα με διαφορά.

Όσον αφορά την αυτόματη αξιολόγηση, εδώ παρατηρούμε ότι τα αποτελέσματα είναι λίγο διαφορετικά. Εδώ εξετάζουμε μεμονωμένα τις μεταφράσεις που έχουν παραχθεί από το DeepL και το Google Translate ως προς τη μετάφραση αναφοράς, δηλαδή μία μετάφραση επαγγελματία μεταφραστή. Όπως έχει προαναφερθεί, η μετρική BLEU υπολογίζει τον αριθμό λέξεων ή των ακολουθιών λέξεων (μέχρι 4 λέξεις), οι οποίες συμπίπτουν με τη μετάφραση αναφοράς και η μετρική TER υπολογίζει το σύνολο των αλλαγών (προσθήκες, διαγραφές, αντικαταστάσεις και μετακινήσεις) που χρειάζονται ως προς τη μετάφραση αναφοράς. Επομένως, για να θεωρηθεί καλό το αποτέλεσμα των παραπάνω συστημάτων χρειάζεται υψηλό BLEU και χαμηλό TER. Για το γλωσσικό ζεύγος Αγγλικά – Ελληνικά παρατηρούμε υψηλό BLEU και χαμηλό TER και στα δύο συστήματα, ωστόσο το Google Translate εμφανίζει καλύτερο αποτέλεσμα. Αντίθετα, για το γλωσσικό ζεύγος Αγγλικά – Ισπανικά έχουμε αντίστροφα αποτελέσματα, δηλαδή υψηλό TER και χαμηλό BLEU και στα δύο συστήματα με μικρή διαφορά στους αριθμούς (DeepL: BLEU 28.1 ενώ Google Translate: BLEU 29.9 και DeepL: TER 54.42 ενώ Google Translate: TER 52.41).

Συνοπτικά, φαίνεται ότι τα αποτελέσματα της αξιολόγησης δείχνουν ότι η απόδοση των συστημάτων νευρωνικής μηχανικής μετάφρασης ποικίλλει ανάλογα με το ζεύγος γλωσσών και τη χρησιμοποιούμενη μέθοδο αξιολόγησης. Στην ανθρώπινη αξιολόγηση, τα αποτελέσματα της MM αξιολογήθηκαν παράλληλα και επίσης εδώ υπάρχει ο κίνδυνος της υποκειμενικότητας, γι' αυτό και η αξιοπιστία του αποτελέσματος κρίνεται από το IAA. Στην αυτόματη αξιολόγηση, τα αποτελέσματα της MM αξιολογήθηκαν μεμονωμένα και ως προς τη μετάφραση αναφοράς. Σχετικά με τη σύγκριση ανάμεσα στα δύο συστήματα, το DeepL είχε γενικά καλύτερες επιδόσεις στην ανθρώπινη αξιολόγηση και για τα δύο ζεύγη γλωσσών. Οι μετρικές αυτόματης αξιολόγησης παρουσιάζουν διακυμάνσεις στις επιδόσεις, με το DeepL και το Google Translate να έχουν μερικές φορές αντικρουόμενα αποτελέσματα στα BLEU και TER.

Δεδομένης της διαφορετικής δομής, του τρόπου εκπαίδευσης των συστημάτων, καθώς και της διαφοροποιημένης φύσης της ποιότητας της μετάφρασης και των διαφορετικών αναγκών των χρηστών, είναι δύσκολο να δηλώσουμε οριστικά ότι ένα σύστημα είναι καθολικά «καλύτερο» από το άλλο. Χρειάζονται ενδελεχείς δοκιμές και εξέταση των συγκεκριμένων απαιτήσεων και προτεραιοτήτων των μεταφραστικών αναγκών ανά περίπτωση για την επιλογή του κατάλληλου συστήματος. Με βάση τα αποτελέσματα της έρευνας που διεξήχθη για την παρούσα εργασία, εάν η ακρίβεια είναι υψίστης σημασίας, το DeepL μπορεί να προτιμηθεί, καθώς αξιολογήθηκε υψηλότερα κατά την ανθρώπινη αξιολόγηση και για τα δύο γλωσσικά ζεύγη. Το Google Translate παρουσίασε αποκλίσεις, έχοντας σημαντικά χειρότερες επιδόσεις στην ευχέρεια σύμφωνα με την ανθρώπινη αξιολόγηση. Επίσης, το DeepL παρουσίασε σταθερές επιδόσεις σε διάφορες κατηγορίες αξιολόγησης, υποδεικνύοντας σταθερότητα στην ποιότητα της μετάφρασης.

15. Επίλογος

Η παρούσα συγκριτική μελέτη έγινε στο πλαίσιο του ΠΜΣ «Επιστήμη της μετάφρασης» του Ιονίου Πανεπιστημίου και αποτελεί μια ενδεικτική έρευνα για την σύγκριση δυο νευρωνικών συστημάτων μηχανικής μετάφρασης σε δύο διαφορετικούς γλωσσικούς συνδυασμούς, χρησιμοποιώντας ένα συνδυασμό μετρικών, έτσι ώστε να αποδειχθεί, ποιο σύστημα έχει την καλύτερη απόδοση. Σκοπός της έρευνας, όμως, δεν είναι μόνο η ανάδειξη των αποτελεσμάτων, αλλά και η παρουσίαση της μεθοδολογίας της αξιολόγησης της μηχανικής μετάφρασης ως διαδικασία. Μέσω ολοκληρωμένων μεθοδολογιών αξιολόγησης που περιλαμβάνουν ανθρώπινες αξιολογήσεις και αυτόματες μετρικές, παρατηρήσαμε αποχρώσεις διαφορές στις επιδόσεις του DeepL και του Google Translate. Η ασυμφωνία μεταξύ της ανθρώπινης μετάφρασης και των αποτελεσμάτων των νευρωνικών συστημάτων υπογραμμίζει περαιτέρω την πολυπλοκότητα της αξιολόγησης της ποιότητας της μετάφρασης και την ανάγκη για πολύπλευρες προσεγγίσεις αξιολόγησης. Η επιλογή του «καλύτερου» συστήματος MM, συγκεκριμένα εδώ μεταξύ του DeepL και του Google Translate εξαρτάται από ένα συνδυασμό συντελεστών, όπως ο τύπος του κειμένου, ο γλωσσικός συνδυασμός, ο σκοπός του έργου, η δομή και ο τρόπος λειτουργίας του συστήματος, οι συγκεκριμένες προτεραιότητες και προτιμήσεις των χρηστών, και άλλοι παράγοντες. Τα ευρήματα της παρούσας εργασίας είναι αντιπροσωπευτικά και συμβάλλουν στον ευρύτερο διάλογο για την αξιολόγηση της μηχανικής μετάφρασης, ωστόσο για να διαπιστωθεί η αντικειμενικότητά τους, είναι απαραίτητο να γίνουν παρόμοιες μελέτες. Όπως έχει προαναφερθεί, η ανθρώπινη αξιολόγηση είναι περισσότερο αναλυτική και δίνει περισσότερες πληροφορίες, όμως υπάρχει πάντα ο κίνδυνος της υποκειμενικής κρίσης των επισημειωτών, γεγονός που αμφισβητεί την ορθότητα των αποτελεσμάτων. Αντίστοιχα, από το πλήθος των αυτόματων μετρικών που υπάρχουν χρησιμοποιήθηκε ένας συνδυασμός τριών από τις πιο γνωστές, οι οποίες αξιολογούν το αποτέλεσμα της μηχανικής μετάφρασης ως προς μία μετάφραση αναφοράς. Η μεθοδολογία της συγκεκριμένης έρευνας ήταν μια χρονοβόρα και απαιτητική διαδικασία, παρά το γεγονός ότι το κείμενο που επιλέχθηκε ήταν μικρής έκτασης, δεδομένης της συνολικής έκτασης της παρούσας διπλωματικής. Γι' αυτό και η συγκεκριμένη εργασία αποτελεί, κυρίως, μια παρουσίαση της διαδικασίας και δίνει κάποια ενδεικτικά αποτελέσματα της απόδοσης των δύο συστημάτων για την αυτόματη μετάφραση κειμένου στους δύο συγκεκριμένους γλωσσικούς συνδυασμούς. Καθώς περιηγούμαστε στο συνεχώς διευρυνόμενο τοπίο της μεταφραστικής τεχνολογίας, η συνεχής έρευνα και βελτίωση θα είναι καθοριστική για την αξιοποίηση του πλήρους δυναμικού των νευρωνικών συστημάτων αυτόματης μετάφρασης για ποικίλα γλωσσικά πλαίσια και ανάγκες χρηστών. Παρόμοιες μελέτες θα μπορούσαν να διεξαχθούν όχι μόνο για τους ίδιους αλλά και περισσότερους γλωσσικούς συνδυασμούς. Επιπλέον, πέρα από τις μετρικές που χρησιμοποιήθηκαν θα μπορούσαν να προστεθούν και άλλες, όπως η λεπτομερής ανάλυση και κατηγοριοποίηση των μεταφραστικών λαθών, ενώ ανφορικά με την αυτόματη αξιολόγηση θα μπορούσαν να γίνουν μελέτες χρησιμοποιώντας πιο πρόσφατες αυτόματες μετρικές που βασίζονται σε νευρωνικά δίκτυα. Για τους σκοπούς της παρούσας εργασίας χρησιμοποιήθηκε ένα κείμενο μη ειδικό, με γενικό λεξιλόγιο, χωρίς κάποια συγκεκριμένη ορολογία, επομένως, θα μπορούσαν να πραγματοποιηθούν μελέτες ως προς την αξιολόγηση της απόδοσης των συγκεκριμένων συστημάτων μηχανικής μετάφρασης σε εξειδικευμένους τομείς, όπως η ιατρική, η νομική ή η τεχνική μετάφραση για τους ίδιους ή περισσότερους γλωσσικούς συνδυασμούς ή ακόμα και για γλώσσες χαμηλών πηγών, όπου τα διαθέσιμα δεδομένα εκπαίδευσης είναι περιορισμένα.

Ολοκληρώνοντας την παρούσα εργασία, ας υπογραμμίσουμε άλλη μια φορά τον καθοριστικό ρόλο της μηχανικής μετάφρασης στη διευκόλυνση της διαγλωσσικής επικοινωνίας και στην προώθηση της παγκόσμιας συνδεσιμότητας. Καθώς η τεχνολογία συνεχίζει να εξελίσσεται με ταχείς ρυθμούς, η σημασία της μηχανικής μετάφρασης για την άρση των γλωσσικών φραγμών και την προώθηση των διαπολιτισμικών ανταλλαγών δεν μπορεί να υπερεκτιμηθεί. Τα ευρήματα της παρούσας μελέτης υπογραμμίζουν την ανάγκη για συνεχή έρευνα και ανάπτυξη στον τομέα της μηχανικής μετάφρασης για την αντιμετώπιση προκλήσεων όπως η ακρίβεια της μετάφρασης, η ευχέρεια και η πολιτισμική προσαρμογή. Διατηρώντας μια αυστηρή και τεκμηριωμένη προσέγγιση στην αξιολόγηση και τη βελτίωση, οι ερευνητές μπορούν να συμβάλουν στη συνεχή πρόοδο της τεχνολογίας της μηχανικής μετάφρασης. Με το βλέμμα στραμμένο στο μέλλον, είναι σημαντικό τόσο οι μελετητές όσο και οι επαγγελματίες να παραμείνουν σε εγρήγορση στην επιδίωξη της καινοτομίας και της αριστείας στον τομέα της μηχανικής μετάφρασης. Παραμένοντας ενήμεροι για τις αναδυόμενες τάσεις, συνεργαζόμενοι σε διάφορους κλάδους και συμμετέχοντας σε κριτικό διάλογο, είναι εφικτό να διαμορφωθεί συλλογικά το μέλλον της μηχανικής μετάφρασης ώστε να εξυπηρετηθούν καλύτερα οι ανάγκες διαφορετικών γλωσσικών κοινοτήτων και να προωθηθεί η διαπολιτισμική κατανόηση. Σε αυτό το διαρκώς εξελισσόμενο τοπίο της τεχνολογίας και της επικοινωνίας, οι ευκαιρίες για περαιτέρω εξερεύνηση και ανακάλυψη στη μηχανική μετάφραση είναι τεράστιες. Με συντονισμένες προσπάθειες και δέσμευση στην επιστημονική έρευνα, είναι δυνατή η συνεχής διεύρυνση των ορίων του εφικτού στη μηχανική μετάφραση και η συμβολή σε μια πιο διασυνδεδεμένη και χωρίς αποκλεισμούς παγκόσμια κοινωνία.

16. Παράρτημα

Κείμενο – Πηγή

Welocalize Leads the Way with New Virtual-Site ISO Certifications

NEW YORK — What is the impact on data security, quality processes, and compliance in virtual work environments? How are global organizations approaching quality management in a cloud-based ecosystem?

Now possible, the new ISO virtual-site certification model provides validation that organizations—where the central business functions are not tied to physical locations—are adhering to and executing quality processes virtually.

Welocalize, a leader in the global language services industry and ISO 9001 certified for 14 years, recently achieved its first virtual-site certification for ISO 9001 and ISO 13485, two of its five ISO frameworks due for recertification this year.

“The virtual-site model certifies cloud-based quality processes and systems; not physical buildings and locations as was done in the past. It is an added assurance for our clients that our quality processes, procedures, and systems required for their service request are in place and can be executed effectively no matter where the project team member resides,” comments Silvie Peric, Director of Quality Systems & Strategy, Welocalize.

“We take immense pride in being one of the first companies in the language services industry to achieve virtual-site certifications. The certification is important for our clients because they can have confidence that the systems in place are designed to help protect their data in a cloud-based work environment. In addition, this model reduces our audit frequency and cost to once per year, instead of numerous individual audits that were required per physical location until earlier this year,” adds Ms. Peric.

The Path to Virtual-Site Accreditation

Prior to the shift to remote worksites brought on by the pandemic, Welocalize put a priority on migrating from physical location-based IT servers to a virtually enabled and protected work environment. This included implementing Okta Single Sign-On (SSO) functionality to facilitate seamless secure user access to protected data.

“These efforts were rolled out in parallel to our ISO/IEC 27001:2013 certification. ISO/IEC 27001 covers the entire scope of information security and provides guidance on a systematic approach to ensuring the confidentiality, integrity, and availability of sensitive company and customer information. All of these steps helped pave the way to creating processes and procedures for a secure, quality-driven virtual work environment,” adds Aaron Heber, Vice President of Global IT at Welocalize.

Μετάφραση επαγγελματία μεταφραστή Αγγλικά – Ελληνικά

Η Welocalize πρωτοπορεί με νέες πιστοποιήσεις ISO για εικονικούς ιστότοπους

ΝΕΑ ΥΟΡΚΗ - Ποιος είναι ο αντίκτυπος στην ασφάλεια των δεδομένων, τις διαδικασίες ποιότητας και τη συμμόρφωση σε εικονικά περιβάλλοντα εργασίας; Πώς προσεγγίζουν οι παγκόσμιοι οργανισμοί τη διαχείριση ποιότητας σε ένα οικοσύστημα που βασίζεται στο cloud;

Το πλέον δυνατό, νέο μοντέλο πιστοποίησης ISO για εικονικούς ιστότοπους παρέχει επικύρωση ότι οι οργανισμοί, στους οποίους οι κεντρικές επιχειρηματικές λειτουργίες δεν συνδέονται με φυσικές τοποθεσίες, ακολουθούν και εκτελούν διαδικασίες ποιότητας εικονικά.

Η Welocalize, ηγέτης στον κλάδο γλωσσικών υπηρεσιών σε παγκόσμιο επίπεδο και με πιστοποίηση ISO 9001 εδώ και 14 χρόνια, πέτυχε πρόσφατα την πρώτη της πιστοποίηση εικονικού ιστότοπου για ISO 9001 και ISO 13485, δύο από τα πέντε πλαίσια ISO που πρόκειται να επαναπιστοποιηθούν φέτος.

«Το μοντέλο εικονικού ιστότοπου πιστοποιεί διαδικασίες και συστήματα ποιότητας που βασίζονται σε cloud, και όχι φυσικά κτίρια και τοποθεσίες όπως γινόταν στο παρελθόν. Είναι μια επιπλέον διαβεβαίωση για τους πελάτες μας ότι οι διαδικασίες ποιότητας, οι μέθοδοι και τα συστήματά μας, που απαιτούνται για το αίτημα υπηρεσίας, είναι σε ισχύ και μπορούν να εκτελεστούν αποτελεσματικά, ανεξάρτητα από το πού κατοικεί το μέλος της ομάδας του έργου», σχολιάζει η Silvie Peric, Διευθύντρια Συστημάτων Ποιότητας και Στρατηγικής στη Welocalize.

«Είμαστε υπερήφανοι που είμαστε μία από τις πρώτες εταιρείες στον κλάδο των γλωσσικών υπηρεσιών που πέτυχε πιστοποιήσεις εικονικών ιστότοπων. Η πιστοποίηση είναι σημαντική για τους πελάτες μας, επειδή μπορούν να έχουν εμπιστοσύνη ότι τα ισχύοντα συστήματα έχουν σχεδιαστεί για να βοηθούν στην προστασία των δεδομένων τους σε ένα περιβάλλον εργασίας που βασίζεται στο cloud. Επιπλέον, το μοντέλο αυτό μειώνει τη συχνότητα και το κόστος του ελέγχου μας σε μία φορά το χρόνο, αντί για πολυάριθμους μεμονωμένους ελέγχους ανά φυσική τοποθεσία που απαιτούνταν μέχρι τις αρχές του έτους», προσθέτει η κα. Peric.

Ο δρόμος προς την διαπίστευση εικονικών ιστότοπων

Πριν από τη αλλαγή σε απομακρυσμένους χώρους εργασίας ως αποτέλεσμα της πανδημίας, η Welocalize έθεσε προτεραιότητα στη μετάβαση από διακομιστές πληροφορικής που βασίζονται σε φυσική τοποθεσία σε ένα προστατευμένο περιβάλλον εργασίας με εικονικές δυνατότητες. Αυτό περιλάμβανε την εφαρμογή της λειτουργικότητας Okta Single Sign-On (SSO) για τη διευκόλυνση της απρόσκοπτης ασφαλούς πρόσβασης των χρηστών σε προστατευμένα δεδομένα.

«Αυτές οι προσπάθειες υλοποιήθηκαν παράλληλα με την πιστοποίηση ISO/IEC 27001:2013. Η ISO/IEC 27001 καλύπτει όλο το εύρος της ασφάλειας πληροφοριών και παρέχει καθοδήγηση αναφορικά με τη συστηματική προσέγγιση για τη διασφάλιση της εμπιστευτικότητας, της ακεραιότητας και της διαθεσιμότητας ευαίσθητων πληροφοριών της εταιρείας και των πελατών. Όλα αυτά τα βήματα βοήθησαν να ανοίξει ο δρόμος για τη δημιουργία διαδικασιών και μεθόδων για ένα ασφαλές εικονικό περιβάλλον εργασίας με γνώμονα την ποιότητα», προσθέτει ο Aaron Heber, Αντιπρόεδρος της Global IT στη Welocalize.

Αποτέλεσμα αυτόματης μετάφρασης από το DeepL, Αγγλικά – Ελληνικά

Η Welocalize πρωτοπορεί με νέες πιστοποιήσεις ISO για εικονικές τοποθεσίες

NEA YOPKH - Ποιος είναι ο αντίκτυπος στην ασφάλεια των δεδομένων, στις διαδικασίες ποιότητας και στη συμμόρφωση σε εικονικά περιβάλλοντα εργασίας; Πώς προσεγγίζουν οι παγκόσμιοι οργανισμοί τη διαχείριση της ποιότητας σε ένα οικοσύστημα που βασίζεται στο cloud;

Το νέο μοντέλο πιστοποίησης ISO virtual-site είναι πλέον εφικτό και παρέχει επικύρωση ότι οι οργανισμοί - όπου οι κεντρικές επιχειρηματικές λειτουργίες δεν είναι συνδεδεμένες με φυσικές τοποθεσίες- τηρούν και εκτελούν τις διαδικασίες ποιότητας εικονικά.

Η Welocalize, ηγέτης στον κλάδο των παγκόσμιων γλωσσικών υπηρεσιών και πιστοποιημένη κατά ISO 9001 εδώ και 14 χρόνια, πέτυχε πρόσφατα την πρώτη της πιστοποίηση εικονικής τοποθεσίας για τα πρότυπα ISO 9001 και ISO 13485, δύο από τα πέντε πλαίσια ISO που πρόκειται να επαναπιστοποιηθούν φέτος.

"Το μοντέλο virtual-site πιστοποιεί διαδικασίες και συστήματα ποιότητας που βασίζονται στο cloud- όχι φυσικά κτίρια και τοποθεσίες, όπως γινόταν στο παρελθόν. Είναι μια πρόσθετη διαβεβαίωση για τους πελάτες μας ότι οι διεργασίες, οι διαδικασίες και τα συστήματα ποιότητας που απαιτούνται για το αίτημά τους για παροχή υπηρεσιών υπάρχουν και μπορούν να εκτελεστούν αποτελεσματικά, ανεξάρτητα από τον

τόπο διαμονής του μέλους της ομάδας έργου", σχολιάζει η Silvie Peric, Director of Quality Systems & Strategy, Welocalize.

"Είμαστε εξαιρετικά υπερήφανοι που είμαστε μία από τις πρώτες εταιρείες στον κλάδο των γλωσσικών υπηρεσιών που πέτυχαν την πιστοποίηση εικονικής τοποθεσίας. Η πιστοποίηση είναι σημαντική για τους πελάτες μας, επειδή μπορούν να έχουν εμπιστοσύνη ότι τα συστήματα που εφαρμόζουμε είναι σχεδιασμένα να συμβάλλουν στην προστασία των δεδομένων τους σε ένα περιβάλλον εργασίας που βασίζεται στο cloud. Επιπλέον, το μοντέλο αυτό μειώνει τη συχνότητα και το κόστος των ελέγχων μας σε μία φορά το χρόνο, αντί για πολυάριθμους μεμονωμένους ελέγχους που απαιτούνταν ανά φυσική τοποθεσία μέχρι τις αρχές του τρέχοντος έτους", προσθέτει η κα Peric.

Η πορεία προς τη διαπίστευση εικονικής τοποθεσίας

Πριν από τη στροφή σε απομακρυσμένα εργοτάξια που προκάλεσε η πανδημία, η Welocalize έθεσε ως προτεραιότητα τη μετάβαση από τους διακομιστές πληροφορικής που βασίζονται σε φυσικές τοποθεσίες σε ένα εικονικά ενεργοποιημένο και προστατευμένο περιβάλλον εργασίας. Αυτό περιελάμβανε την εφαρμογή της λειτουργικότητας Okta Single Sign-On (SSO) για τη διευκόλυνση της απρόσκοπτης ασφαλούς πρόσβασης των χρηστών σε προστατευμένα δεδομένα.

"Αυτές οι προσπάθειες αναπτύχθηκαν παράλληλα με την πιστοποίησή μας κατά ISO/IEC 27001:2013. Το ISO/IEC 27001 καλύπτει όλο το πεδίο της ασφάλειας πληροφοριών και παρέχει οδηγίες για μια συστηματική προσέγγιση για τη διασφάλιση της εμπιστευτικότητας, της ακεραιότητας και της διαθεσιμότητας ευαίσθητων πληροφοριών της εταιρείας και των πελατών. Όλα αυτά τα βήματα βοήθησαν να ανοίξει ο δρόμος για τη δημιουργία διαδικασιών και διαδικασιών για ένα ασφαλές, ποιοτικό εικονικό περιβάλλον εργασίας", προσθέτει ο Aaron Heber, αντιπρόεδρος του Global IT της Welocalize.

Αποτέλεσμα αυτόματης μετάφρασης από το Google Translate, Αγγλικά – Ελληνικά

Το Welocalize οδηγεί το δρόμο με τις νέες πιστοποιήσεις ISO για εικονικούς ιστότοπους

NEA YOPKH — Ποιος είναι ο αντίκτυπος στην ασφάλεια των δεδομένων, τις διαδικασίες ποιότητας και τη συμμόρφωση σε εικονικά περιβάλλοντα εργασίας; Πώς προσεγγίζουν παγκόσμιοι οργανισμοί τη διαχείριση ποιότητας σε ένα οικοσύστημα που βασίζεται στο cloud;

Τώρα είναι δυνατό, το νέο μοντέλο πιστοποίησης εικονικών τοποθεσιών ISO παρέχει επικύρωση ότι οι οργανισμοί - όπου οι κεντρικές επιχειρηματικές λειτουργίες δεν συνδέονται με φυσικές τοποθεσίες - τηρούν και εκτελούν διαδικασίες ποιότητας εικονικά.

Η Welocalize, ηγέτης στον παγκόσμιο κλάδο γλωσσικών υπηρεσιών και πιστοποιημένη κατά ISO 9001 για 14 χρόνια, πέτυχε πρόσφατα την πρώτη της πιστοποίηση εικονικής τοποθεσίας για ISO 9001 και ISO 13485, δύο από τα πέντε πλαίσια ISO που πρόκειται να επαναπιστοποιηθούν φέτος.

«Το μοντέλο εικονικού ιστότοπου πιστοποιεί διαδικασίες και συστήματα ποιότητας που βασίζονται σε cloud. όχι φυσικά κτίρια και τοποθεσίες όπως γινόταν στο παρελθόν. Είναι μια πρόσθετη διαβεβαίωση για τους πελάτες μας ότι οι διαδικασίες ποιότητας, οι διαδικασίες και τα συστήματά μας που απαιτούνται για το αίτημα εξυπηρέτησης είναι σε ισχύ και μπορούν να εκτελεστούν αποτελεσματικά ανεξάρτητα από το πού κατοικεί το μέλος της ομάδας του έργου», σχολιάζει η Silvie Peric, Διευθύντρια Συστημάτων Ποιότητας & Στρατηγικής, Welocalize.

«Είμαστε πολύ υπερήφανοι που είμαστε μία από τις πρώτες εταιρείες στον κλάδο των γλωσσικών υπηρεσιών που πέτυχε πιστοποιήσεις εικονικών ιστοτόπων. Η πιστοποίηση είναι σημαντική για τους πελάτες μας, επειδή μπορούν να έχουν εμπιστοσύνη ότι τα υπάρχοντα συστήματα έχουν σχεδιαστεί για να βοηθούν στην προστασία των δεδομένων τους σε ένα περιβάλλον εργασίας που βασίζεται στο cloud. Επιπλέον, αυτό το μοντέλο μειώνει τη συχνότητα και το κόστος του ελέγχου μας σε μία φορά το χρόνο, αντί για πολυάριθμους μεμονωμένους ελέγχους που απαιτούνταν ανά φυσική τοποθεσία μέχρι τις αρχές του έτους», προσθέτει η κ. Peric.

The Path to Virtual-Site Accreditation

Πριν από τη στροφή σε απομακρυσμένα εργοτάξια που προκάλεσε η πανδημία, η Welocalize έθεσε προτεραιότητα στη μετάβαση από διακομιστές πληροφορικής που βασίζονται σε φυσική τοποθεσία σε ένα εικονικά ενεργοποιημένο και προστατευμένο περιβάλλον εργασίας. Αυτό περιελάμβανε την εφαρμογή της λειτουργικότητας Okta Single Sign-On (SSO) για τη διευκόλυνση της απρόσκοπτης ασφαλούς πρόσβασης χρηστών σε προστατευμένα δεδομένα.

«Αυτές οι προσπάθειες έγιναν παράλληλα με την πιστοποίηση ISO/IEC 27001:2013. Το ISO/IEC 27001 καλύπτει ολόκληρο το εύρος της ασφάλειας πληροφοριών και παρέχει καθοδήγηση σχετικά με μια συστηματική προσέγγιση για τη διασφάλιση της εμπιστευτικότητας, της ακεραιότητας και της διαθεσιμότητας ευαίσθητων πληροφοριών εταιρείας και πελατών. Όλα αυτά τα βήματα βοήθησαν να ανοίξει ο δρόμος για τη δημιουργία διαδικασιών και διαδικασιών για ένα ασφαλές εικονικό περιβάλλον εργασίας με γνώμονα την ποιότητα», προσθέτει ο Aaron Heber, Αντιπρόεδρος Global IT της Welocalize.

Μετάφραση επαγγελματία μεταφραστή Αγγλικά – Ισπανικά

Welocalize abre el camino con las nuevas certificaciones ISO virtual

Nueva York - ¿Cuál es el impacto en la seguridad de datos, procesos de control de calidad y cumplimiento normativo en los entornos de trabajo virtuales? ¿Cómo abordan las organizaciones globales la gestión de calidad dentro de un ecosistema de la nube?

Ahora sí es posible, ya que el modelo de las nuevas certificaciones ISO virtual proporciona una validación concedida por las organizaciones en las que los servicios centrales de la empresa no se ven físicamente localizados y así los procesos de control de calidad se ejercen virtualmente.

Welocalize, un líder en la industria de servicios lingüísticos internacionales y certificada ISO 9001 desde hace 14 años, recientemente recibió su primera certificación virtual para ISO 9001 y ISO 13485, dos de cinco marcos ISO que deben efectuar su recertificación este año.

“El modelo virtual asegura procesos de control de calidad y sistemas basados en nube, en contraste con las instalaciones y las ubicaciones físicas del pasado. Se consigue una garantía añadida hacia nuestra clientela cuando nuestros procesos de control de calidad, procedimientos y sistemas quedan en orden y se pueden realizar de manera eficaz sin importar la ubicación física del empleado que trabaja en el proyecto” comenta Silvie Peric, Directora de Sistemas de Calidad y Estrategias de Welocalize.

“Es todo un orgullo para nosotros ser una de las primeras empresas de servicios lingüísticos que han recibido certificaciones de espacio virtual. La certificación se ve importante para nuestra clientela porque le da la seguridad y certeza de que los sistemas instalados sean diseñados para proteger sus datos en entornos laborales basados en nube. Cabe añadir que tal modelo reduce nuestra frecuencia de auditoría e incluso el coste a una vez al año, en lugar de las múltiples auditorías individuales exigidas por cada ubicación física hasta comienzos de este año”, añade Sra Peric.

El camino hacia la acreditación virtual

Antes de que la pandemia impusiera a los empleados dirigirse a espacios de trabajo remotos, Welocalize dió prioridad al traslado de servidores informáticos basados en la ubicación a entornos laborales virtualmente activos y protegidos. Dicho cambio incluyó implementar la funcionalidad Okta Single Sign-On (SSO) para facilitar el acceso seguro e integrado del usuario a datos protegidos.

“Tales esfuerzos se desarrollaron al mismo tiempo que nuestra certificación ISO/IEC 27001:2013. El ISO/IEC 27001 cubre todo el ámbito de seguridad de la información y suministra una guía de enfoque sistemático para garantizar la confidencialidad, la integridad y la disponibilidad de la información sensible de la empresa y de sus clientes. Como resultado, los pasos anteriormente mencionados prepararon el

terreno para poder crear procesos y procedimientos con vista a un entorno laboral virtual seguro y de calidad" agrega Aaron Heber, vicepresidente de Global IT en Welocalize.

Αποτέλεσμα αυτόματης μετάφρασης από το DeepL, Αγγλικά – Ισπανικά

Welocalize lidera el camino con nuevas certificaciones ISO de sitios virtuales

NUEVA YORK - ¿Cuál es el impacto en la seguridad de los datos, los procesos de calidad y el cumplimiento normativo en los entornos de trabajo virtuales? ¿Cómo enfocan las organizaciones globales la gestión de la calidad en un ecosistema basado en la nube?

Ahora es posible, el nuevo modelo de certificación ISO de sitio virtual proporciona la validación de que las organizaciones -donde las funciones empresariales centrales no están vinculadas a ubicaciones físicas- se adhieren a los procesos de calidad y los ejecutan virtualmente.

Welocalize, líder en el sector de los servicios lingüísticos globales y con la certificación ISO 9001 desde hace 14 años, obtuvo recientemente su primera certificación de sitio virtual para ISO 9001 e ISO 13485, dos de sus cinco marcos ISO que deben recertificarse este año.

"El modelo de sede virtual certifica los procesos y sistemas de calidad basados en la nube y no en edificios y ubicaciones físicas, como se hacía antes. Es una garantía añadida para nuestros clientes de que nuestros procesos, procedimientos y sistemas de calidad requeridos para su solicitud de servicio están implantados y pueden ejecutarse eficazmente independientemente de dónde resida el miembro del equipo del proyecto", comenta Silvie Peric, Directora de Sistemas y Estrategia de Calidad de Welocalize.

"Estamos muy orgullosos de ser una de las primeras empresas del sector de los servicios lingüísticos en obtener la certificación de sede virtual. La certificación es importante para nuestros clientes porque pueden confiar en que los sistemas implantados están diseñados para ayudar a proteger sus datos en un entorno de trabajo basado en la nube. Además, este modelo reduce la frecuencia y el coste de nuestras auditorías a una vez al año, en lugar de las numerosas auditorías individuales que se requerían por ubicación física hasta principios de este año", añade la Sra. Peric.

El camino hacia la acreditación de centros virtuales

Antes del cambio a lugares de trabajo remotos provocado por la pandemia, Welocalize dio prioridad a la migración de servidores de TI basados en ubicaciones físicas a un entorno de trabajo virtualmente habilitado y protegido. Esto incluyó la implementación de la funcionalidad de inicio de sesión único (SSO) de Okta para facilitar el acceso seguro y sin interrupciones de los usuarios a los datos protegidos.

"Estos esfuerzos se desplegaron en paralelo a nuestra certificación ISO/IEC 27001:2013. La norma ISO/IEC 27001 abarca todo el ámbito de la seguridad de la información y ofrece orientación sobre un enfoque sistemático para garantizar la confidencialidad, integridad y disponibilidad de la información sensible de empresas y clientes. Todos estos pasos ayudaron a allanar el camino hacia la creación de procesos y procedimientos para un entorno de trabajo virtual seguro y orientado a la calidad", añade Aaron Heber, vicepresidente de TI global de Welocalize.

Αποτέλεσμα αυτόματης μετάφρασης από το Google Translate, Αγγλικά – Ισπανικά

Welocalize lidera el camino con nuevas certificaciones ISO de sitio virtual

NUEVA YORK: ¿Cuál es el impacto en la seguridad de los datos, los procesos de calidad y el cumplimiento en los entornos de trabajo virtuales? ¿Cómo abordan las organizaciones globales la gestión de la calidad en un ecosistema basado en la nube?

Ahora es posible, el nuevo modelo de certificación de sitio virtual ISO brinda validación de que las organizaciones, donde las funciones comerciales centrales no están vinculadas a ubicaciones físicas, se adhieren y ejecutan procesos de calidad virtualmente.

Welocalize, líder en la industria de servicios lingüísticos globales y certificado ISO 9001 durante 14 años, logró recientemente su primera certificación de sitio virtual para ISO 9001 e ISO 13485, dos de sus cinco marcos ISO que deben recertificarse este año.

“El modelo de sitio virtual certifica procesos y sistemas de calidad basados en la nube; no edificios físicos y ubicaciones como se hizo en el pasado. Es una garantía adicional para nuestros clientes que nuestros procesos, procedimientos y sistemas de calidad requeridos para su solicitud de servicio estén implementados y puedan ejecutarse de manera efectiva sin importar dónde resida el miembro del equipo del proyecto”, comenta Silvie Peric, Directora de Sistemas y Estrategia de Calidad. , Localizamos.

“Nos enorgullece enormemente ser una de las primeras empresas en la industria de servicios lingüísticos en lograr certificaciones de sitios virtuales. La certificación es importante para nuestros clientes porque pueden confiar en que los sistemas implementados están diseñados para ayudarlos a proteger sus datos en un entorno de trabajo basado en la nube. Además, este modelo reduce la frecuencia y el costo de nuestras auditorías a una vez al año, en lugar de las numerosas auditorías individuales que se requerían por ubicación física hasta principios de este año”, agrega la Sra. Peric.

El camino hacia la acreditación del sitio virtual

Antes del cambio a lugares de trabajo remotos provocado por la pandemia, Welocalize dio prioridad a la migración de servidores de TI físicos basados en la ubicación a un entorno de trabajo virtualmente habilitado y protegido. Esto incluyó la implementación de la funcionalidad de inicio de sesión único (SSO) de Okta para facilitar el acceso seguro y sin problemas de los usuarios a los datos protegidos.

“Estos esfuerzos se implementaron en paralelo a nuestra certificación ISO/IEC 27001:2013. ISO/IEC 27001 cubre todo el alcance de la seguridad de la información y brinda orientación sobre un enfoque sistemático para garantizar la confidencialidad, integridad y disponibilidad de la información confidencial de la empresa y del cliente. Todos estos pasos ayudaron a allanar el camino para crear procesos y procedimientos para un entorno de trabajo virtual seguro y basado en la calidad”, agrega Aaron Heber, vicepresidente de TI global de Welocalize.

17. Βιβλιογραφικές Αναφορές

Ξένη βιβλιογραφία

Automatic Language Processing Advisory Committee (ALPAC), (1966): Language and Machines: Computers in Translation and Linguistics.

Baker, M. (2011). In Other Words. A Coursebook on Translation, 2nd ed. London: Routledge.

Banerjee, S., Lavie, A. (2005). METEOR: An automatic metric for MT evaluation with improved correlation with human judgments. In: Proceedings of the ACL Workshop on Intrinsic and Extrinsic Evaluation Measures for MT and/or Summarization. Ann Arbor, Michigan, USA, pp. 65–72.

Bojar, O., Chatterjee, R., Federmann, C., Haddow, B., Huck, M., Hokamp, C., Koehn, P., Logacheva, V., Monz, C., Negri, M., Post, M., Scarton, C., Specia, L., Turchi M. (2015). Findings of the 2015 Workshop on Statistical Machine Translation. Proceedings of the Tenth Workshop on Statistical Machine Translation. Association for Computational Linguistics, 1-46.

Brown, P., Della Pietra, S., Della Pietra, V., and Mercer, R. (1993). The mathematics of statistical machine translation: Parameter estimation. Computational Linguistics, 19(2):263–311.

Callison-Burch, C., C. Fordyce, P. Koehn, Monz, C., Schroeder, J. (2007). (meta-) evaluation of machine translation. Proceedings of the Second Workshop on Statistical Machine Translation, pages 136–158, Prague, Czech Republic, June. Association for Computational Linguistics.

Carroll, J. (1966). An Experiment in Evaluating the Quality of Translations, Mechanical Translation 9 (3–4): 55–66. [Google Scholar]

Castilho, S., Moorkens, J., Gaspari, F., Sennrich, R., Sosoni, V., Georgakopoulou, G., Lohar, P., Way, A., Barone, A.V.M., Gialama, M. (2017). A Comparative Quality Evaluation of PBSMT and NMT using Professional Translators. Proceedings of MT Summit XVI, vol.1: Research Track, p. 131-116. Nagoya, Japan.

Chen, Z., Komrusch, S., Tufano, M., Pouchet, L., Poshyvanyk, D. and Monperrus, M. (2019). SEQUENCER: Sequence-to-Sequence Learning for End-to-End Program Repair, [online] available at: <https://arxiv.org/pdf/1901.01808.pdf>

Chi-kiu Lo. 2019. YiSi - a Unified Semantic MT Quality Evaluation and Estimation Metric for Languages with Different Levels of Available Resources. In Proceedings of the Fourth Conference on Machine Translation (Volume 2: Shared Task Papers, Day 1), pages 507–513, Florence, Italy. Association for Computational Linguistics.

Chiusano, F. (2023). Two minutes NLP — Learn the ROUGE metric by examples. Medium, available at: <https://medium.com/nlplanet/two-minutes-nlp-learn-the-rouge-metric-by-examples-f179cc285499>

Cho, K., van Merriënboer, B., Bahdanau, D., and Bengio, Y. (2014b). On the properties of neural machine translation: Encoder–Decoder approaches. In Eighth Workshop on Syntax, Semantics and Structure in Statistical Translation. to appear.

Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. Educational and Psychological Measurement, pp. 37–46.

Colina, S. (2008). Translation Quality Evaluation: Empirical evidence for a Functionalist Approach. The Translator 14:1. 97–134.

Doherty, M. (2002). Language Processing in Discourse: A Key to Felicitous Translation. London: Routledge.

Fields, P., Hague, D., Koby, G.S., Lommel, A., Melby, A. (2014). What is quality? A management discipline and the translation industry get acquainted. *Revista Tradumàtica* 12:404–412. [online] available at: https://ddd.uab.cat/pub/tradumatica/tradumatica_a2014n12/tradumatica_a2014n12p404.pdf.

Fleiss, J. (1971). Measuring Nominal Scale Agreement among Many Raters. *Psychological Bulletin*, Vol. 76.

Fleiss, J. Cohen, J. and Everitt, B.S. (1969). Large Sample Standard Errors of Kappa and Weighted Kappa. *Psychological Bulletin*, Vol. 72.

Fordyce, C. S. (task responsible) (2007). Survey of Machine Translation Evaluation. Statistical and Hybrid Machine Translation Between All European Languages, IST 034291, Deliverable 1.3, EuroMatrix.

Gupta, R., Orasan, C., Van Genabith, J. (2015). ReVal: A Simple and Effective Machine Translation Evaluation Metric, [online] available at: <https://aclanthology.org/D15-1124.pdf>

Based on Recurrent Neural Networks

Hatim, B. and Mason, I. (1997). *The Translation as Communicator*. London: Routledge.

Hatim, B. and Munday, J. (2004). *Translation: An Advanced Resource Book*. Oxford: Routledge.

Holz-Mänttari, J. (1984). *Translatorisches Handeln. Theorie und Methode*, Helsinki: Suomalainen Tiedeakatemia.

Horwood, E. (2020). Machine Translation: past, present, future. [online] [Hutchinsweb.me.uk.](http://www.hutchinsweb.me.uk/), [online] available at: <http://www.hutchinsweb.me.uk/PPF-TOC.htm>

House, J. (2014). Translation Quality Assessment: Past and Present. In: House J. (eds) *Translation: A Multidisciplinary Approach*. Palgrave Advances in Language and Linguistics. Palgrave Macmillan, London

How does deepl work? (2021) DeepL Translate: The world's most accurate translator, available at: <https://www.deepl.com/en/blog/how-does-deepl-work>.

Hutchins, J. (1986). *Machine translation: past, present, future*. Chichester: Ellis Horwood.

Hutchins, W. (1995). MACHINE TRANSLATION: A BRIEF HISTORY. [online] [Hutchinsweb.me.uk.](http://www.hutchinsweb.me.uk/), [online] available at: <http://www.hutchinsweb.me.uk/ConcHistoryLangSci-1995.pdf>

Hutchins, W. (2001). Machine Translation over fifty years, [online] available at: [Hutchinsweb.me.uk.](http://www.hutchinsweb.me.uk/)

Hutchins, J., & Somers, H. (1992). *An Introduction to Machine Translation*. London: Academic Press Limited, [online] available at: <https://www.scirp.org/reference/referencespapers?referenceid=2839626>

Kalchbrenner, N. and Blunsom, P. (2013). Recurrent continuous translation models. In *Proceedings of the ACL Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 1700–1709. Association for Computational Linguistics.

Klubička, F., Toral, A. and Sánchez-Cartagena, V.M. (2018). Quantitative finegrained human evaluation of machine translation systems: a case study on English to Croatian. *Machine Translation*, 32(3), pp. 195–215.

Koby, G.S., Fields, P., Hague, D., Lommel, A., Melby, A. (2014) Defining translation quality. *Revista Tradumàtica* 12:413–420. [online] available at: https://ddd.uab.cat/pub/tradumatica/tradumatica_a2014n12/tradumatica_a2014n12p413.pdf.

Koehn, P., Och, F. J., and Marcu, D. (2003). Statistical phrase-based translation. In *Proceedings of the 2003 Conference of the North American Chapter of the Association for Computational Linguistics on Human*

Language Technology - Volume 1, NAACL '03, pages 48–54, Stroudsburg, PA, USA. Association for Computational Linguistics.

Lauscher, S. (2000). Translation Quality Assessment, *The Translator*, 6:2, 149-168, DOI: 10.1080/13556509.2000.10799063

Lavie, A. (2011). Evaluating the Output of Machine Translation Systems. In: *Proceedings of the 13th MT Summit*. Xiamen, China.

Levenshtein, V. I. (1966). Binary Codes Capable of Correcting Deletions, Insertions, and Reversals. *Soviet Physics Doklady*, 10:707–710.

Localization Industry Standards Association (LISA), (2004): *LISA QA Model 3.0 Product Documentation*. Féchy, Switzerland: LISA.

Lommel, A. and Melby, A.K. (2018). MQM-DQF: A good marriage (Translation quality for the 21st century). Tutorial at the 13th Conference of the Association for Machine Translation in the Americas. Boston, USA, 17 March 2018. AMTA, [online] available at: <https://www.aclweb.org/anthology/W18-1925>.

Mutnmt. (n.d.), available at: <https://ntradumatica.uab.cat/about>

Nida, E. (1964). *Toward a Science of Translation*. Leiden: Brill.

Nida, E. and Taber, C. (1969). *The Theory and Practice of Translation*, Leiden: Brill. [Google Scholar]

Nießen, S., Och, F.J., Leusch, G., Ney, H. (2000). An evaluation tool for machine translation: Fast evaluation for MT research. In *Proceedings of the 2nd International Conference on Language Resources and Evaluation (LREC-2000)*, pages 39–45.

Nirenburg, S. (1989). Knowledge-Based Machine Translation. *Machine Translation* 4 (1989), 5 - 24.

Nord, C. (1993). *Einführung in das funktionale Übersetzen: Am Beispiel von Titeln und Überschriften*. Tübingen: Francke. [UTB, 1734.]

Nord, C. (1997). *Translating as a Purposeful Activity: Functionalist Approaches Explained*, Manchester: St. Jerome. [Google Scholar]

OECD AI Policy Observatory Portal. (n.d.), available at: <https://oecd.ai/fr/catalogue/metrics/consensus-based-image-description-evaluation-%28cider%29>

Papineni, K., Roukos, S., Ward, T., Zhu, W.-J. (2002). BLEU: a method for automatic evaluation of machine translation. *Proceedings of the 40th Annual Meeting of the Association for the Computational Linguistics (ACL)*, Philadelphia, July 2002, pp. 311-318.

Popović, M (2015). chrF: character n-gram F-score for automatic MT evaluation. In *Proceedings of the Tenth Workshop on Statistical Machine Translation*, pages 392–395, Lisbon, Portugal. Association for Computational Linguistics.

Przybocki, M., Peterson, K., Bronsart, S. (2008). Official results of the NIST 2008. Metrics for Machine Translation Challenge (Metrics-MATR08) <http://nist.gov/speech/tests/metricsmatr/2008/results/>, October.

Rei, R., Stewart, C., Farinha, An. C and Lavie, Al. (2020). COMET: A Neural Framework for MT Evaluation. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 2685–2702, Online. Association for Computational Linguistics.

Reiß, K. (1971). *Möglichkeiten und Grenzen der Übersetzungskritik*. München: Hueber

Reiß, K. and Vermeer, H.J. (1984). *Grundlegung einer Allgemeinen Translationstheorie*, Tübingen: Niemeyer

Sfetcu, N. (2009). Approaches in machine translation - SetThings. [online] SetThings, [online] available at: <https://www.setthings.com/en/approaches-in-machine-translation/>

Snover M., Madnani, N., Dorr, B., Schwartz, R. (2011). Chapter 5.2 Automatic and SemiAutomatic Measures. In Handbook of Natural Language Processing and Machine Translation: DARPA Global Autonomous Language Exploitation. New York: Springer, pp. 815-823

Snover, M., Dorr, B., Schwartz, R., Micciulla, L., Makhoul, J. (2006). A Study of Translation Edit Rate with Targeted Human Annotation. Proceedings of Association for Machine Translation in the Americas (AMTA-2006), pages 223–231, Cambridge, Massachusetts.

Sommerlad, Joe (June 19, 2018). "Google Translate: How does the search giant's multilingual interpreter actually work?". The Independent. Archived from the original on November 2, 2020.

Sosoni, V. and Stasimioti, M. (2019). MT output and post-editing effort: Insights from a comparative analysis of SMT and NMT output and implications for training. Fit-For-Market Translator and Interpreter Training in a Digital Age (Language and Linguistics).

Stasimioti, M., Sosoni, V., Mouratidis, D., Kermanidis, K. (2020). Machine Translation Quality: A comparative evaluation of SMT, NMT and tailored-NMT outputs. Proceedings of the 22nd Annual Conference of the European Association for Machine Translation, p. 441–450. Lisboa, Portugal.

Stolze, R. (2003). Hermeneutisches Übersetzen. Tübingen: Narr

Sutskever, I., Vinyals, O., and Le, Q. (2014). Sequence to sequence learning with neural networks. In Advances in Neural Information Processing Systems (NIPS 2014).

Thompson, B. and Post, M. (2020). Automatic Machine Translation Evaluation in Many Languages via Zero-Shot Paraphrasing [online], available at <https://aclanthology.org/2020.emnlp-main.8.pdf>

Thurmair, G. (2009). Comparing different architectures of Hybrid Machine Translation systems. In Proceedings of the MT Summit XII, pages 340–347.

Toury, G. (1995). Descriptive Translation Studies and Beyond. Amsterdam: Benjamins

Turian, J.P., Shen L., Melamed D.I. (2003). Evaluation of machine translation and its evaluation. Proc. MT Summit IX, pages 386–393, New Orleans, LA.

Turovsky, Barak (2016). "Found in translation: More accurate, fluent sentences in Google Translate". The Keyword Google Blog. Google Inc. Archived from the original on April 7, 2017.

Turovsky, Barak (2016). "Ten years of Google Translate". Google Translate Blog. Google Inc. Archived from the original on December 24, 2019.

Vashee, K. (2017). Customization or 'Specializing' Neural Machine Translation in SYSTRAN. Medium, ITNEXT, [online] available at: itnext.io/customization-or-specializing-neural-machine-translation-in-systran-261e67d7fb57.

Venuti, L. (2004). The Translator's Invisibility. London: Routledge

Vermeer, H. J. (1989). Skopos und Translationsauftrag – Aufsätze, Heidelberg: thw.

Vermeer, H. J. (1996). A Skopos Theory of Translation (Some arguments for and against), Heidelberg: TEXTconTEXT-Verlag.

White, J.S., O'Connell, T., O'Mara, F. (1994). The ARPA MT Evaluation Methodologies: Evolution, Lessons, and Future Approaches. Proceedings of the First Conference of the Association for Machine Translation, pages 193–205.

Wyndham, A. (2023, January 27). Inside DeepL: The world's fastest-growing, most secretive machine translation company. Slator, available at: <https://slator.com/inside-deepl-the-worlds-fastest-growing-most-secretive-machine-translation-company/>

Zhang, T., Kishore, V., Wu, F., Weinberger, K. Q., Artzi, Y. (2020). BERTScore: Evaluating Text Generation with BERT [online] available at: <https://arxiv.org/pdf/1904.09675.pdf>.

Ελληνική βιβλιογραφία

Χατζηκουμή Ε. (2015). Αξιολόγηση ποιότητας μηχανικής μετάφρασης. Τέσσερα συστήματα στο μικροσκόπιο της μη αυτοματοποιημένης αξιολόγησης και της αυτόματης μετρικής (Διπλωματική εργασία, Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών, Εθνικό Μετσόβιο Πολυτεχνείο, Ινστιτούτο Επεξεργασίας Λόγου Ερευνητικό Κέντρο Αθήνα, Αθήνα Ελλάδα). Διαθέσιμο στον ιστότοπο: <http://glotta.ntua.gr/nlp/courses/Technoglossia/Dissertations/diplomatiki-Chatzikoumi.pdf>