



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

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΜΕΡΟΠΗ ΜΑΡΙΑ ΛΑΜΠΡΟΠΟΥΛΟΥ
Ξ2018304

**Ανθρώπινη αξιολόγηση μεταφρασμένων κειμένων που έχουν
παραχθεί από διαφορετικά συστήματα μηχανικής
μετάφρασης**

ΜΕΛΗ ΕΠΙΤΡΟΠΗΣ ΕΠΟΠΤΕΙΑΣ:

ΚΑΤΙΑ ΛΗΔΑ ΚΕΡΜΑΝΙΔΟΥ – Α' ΕΠΙΒΛΕΠΩΝ
ΒΙΛΕΛΜΙΝΗ ΣΩΣΩΝΗ – Β' ΕΠΙΒΛΕΠΩΝ
ΓΕΩΡΓΙΟΣ ΚΕΝΤΡΩΤΗΣ – Γ' ΕΠΙΒΛΕΠΩΝ

ΚΕΡΚΥΡΑ
ΣΕΠΤΕΜΒΡΙΟΣ 2018

**Ανθρώπινη αξιολόγηση μεταφρασμένων κειμένων που έχουν
παραχθεί από διαφορετικά συστήματα μηχανικής
μετάφρασης**

Περιεχόμενα

1.	Εισαγωγή.....	4
2.	Σύντομη Ιστορική Αναδρομή.....	5
3.	Συστήματα Μηχανικής Μετάφρασης.....	5
3.1	Συστήματα Μηχανικής Μετάφρασης Βάσει Κανόνων (Rule-Based Machine Translation, RBMT).....	5
3.2	Συστήματα Μηχανικής Μετάφρασης Βάσει Παραδειγμάτων (Example-Based Machine Translation, EBMT).....	6
3.3	Στατιστικά Συστήματα Μηχανικής Μετάφρασης (Statistical Machine Translation, SMT).....	7
3.4	Υβριδικά Συστήματα Μηχανικής Μετάφρασης (Hybrid Machine Translation).....	7
3.5	Νευρωνικά Συστήματα Μηχανικής Μετάφρασης (Neural Machine Translation, NMT).....	8
4.	Ποιότητα της Μετάφρασης.....	8
4.1	Εξασφάλιση της ποιότητας της μετάφρασης.....	10
4.2	Μέθοδοι αξιολόγησης της ποιότητας της μετάφρασης.....	11
5.	Αξιολόγηση της ποιότητας της Μηχανικής Μετάφρασης.....	13
5.1	Ακρίβεια και Ευχέρεια.....	14
5.2	Κατάταξη.....	14
5.3	Κατηγοριοποίηση των λαθών.....	15
5.4	Τυπολογία λαθών DQF.....	17
5.5	Εναρμονισμένο πρότυπο MQM-DQF.....	18
6.	Inter-annotator agreement.....	18
7.	Αυτόματη Αξιολόγηση της ποιότητας της Μηχανικής Μετάφρασης.....	19
8.	Μεθοδολογία της Έρευνας.....	23
9.	Αποτελέσματα της έρευνας.....	27
10.	Συζήτηση αποτελεσμάτων.....	35
11.	Επίλογος.....	36
12.	Παράρτημα.....	38
13.	Βιβλιογραφικές αναφορές.....	44

1. Εισαγωγή

Η παρούσα διπλωματική εργασία αποτελεί μια συγκριτική μελέτη στον τομέα της μετάφρασης, και συγκεκριμένα στην επανάσταση που έχει επιφέρει η Μηχανική Μετάφραση (MM) κατά τη μεταφραστική διαδικασία. Η δημιουργία μηχανών αυτόματης μετάφρασης, των οποίων η αρχιτεκτονική βασίζεται σε νευρωνικά δίκτυα, ξεκίνησε την τελευταία δεκαετία και αποτελεί μεγάλη καινοτομία στην ιστορία της MM. Τα συστήματα που υπήρχαν ως τότε ήταν τα στατιστικά συστήματα (Statistical Machine Translation, SMT), συστήματα με χρήση παραδειγμάτων (Example-Based Machine Translation, EBMT), συστήματα με χρήση κανόνων (Rule-Based Machine Translation, RBMT) και κάποια υβριδικά συστήματα (Hybrid Systems), τα οποία αποτελούνται από δυο διαφορετικούς τύπους συστημάτων. Κανένα σύστημα, όμως, δεν μπορούσε, και δεν μπορεί μέχρι σήμερα, να προσφέρει μια άρτια μετάφραση χωρίς να είναι απαραίτητη η επέμβαση του ανθρώπου στο τελικό αποτέλεσμα. Γι' αυτόν τον λόγο έχουν διαμορφωθεί κανόνες μετεπεξεργασίας (Post-Editing, PE) του κειμένου που προέρχεται από MM για να ελέγχεται και να γίνεται λειτουργικό, καθώς και μέθοδοι αξιολόγησης των συστημάτων MM, έτσι ώστε να επιδέχονται βελτίωση και να εντοπίζονται οι αδυναμίες τους. Τα τελευταία χρόνια, πλήθος εταιριών χρησιμοποιεί τη MM σε πολλά έργα, όπως φαίνεται και στην έρευνα του Common Sense Advisory (CSA) (Common Sense Advisory, 2016), η οποία αποδεικνύει ότι το 80% των εταιριών προσφέρουν υπηρεσίες μετεπεξεργασίας (PE) κειμένου που έχει προκύψει από MM. Όλα αυτά δείχνουν ότι είναι απαραίτητη η βελτίωση των συστημάτων MM, η αξιολόγηση τους και η εκπαίδευση των μεταφραστών για σωστή χρήση τους έτσι ώστε να εισχωρήσουν στη μεταφραστική διαδικασία στη χώρα μας και να αυξήσουν την παραγωγικότητα. Λόγω της πρόσφατης δημιουργίας νέου τύπου δικτύου στη MM, των νευρωνικών συστημάτων μηχανικής μετάφρασης (Neural Machine Translation, NMT), πολλές έρευνες λαμβάνουν χώρα, έτσι ώστε να αξιολογηθούν τα αποτελέσματα που παράγουν ώστε να βελτιωθούν οι αστοχίες τους, από τη μια, και από την άλλη συγκρίνονται αντιπαραβολικά με παλαιότερα συστήματα έτσι ώστε να αποδειχθεί, ποιο σύστημα είναι πιο λειτουργικό σε ένα ζεύγος γλωσσών υπό συγκεκριμένες συνθήκες. Τέτοιες έρευνες με τη μια γλώσσα υπό εξέταση να είναι τα ελληνικά έχουν διεξαχθεί από τις Sosoni και Stasimioti (2019), Stasimioti et al. (2020) και Castilho et al. (2017). Αποτελέσματα ερευνών όπως οι παραπάνω δείχνουν ότι τα νευρωνικά συστήματα έχουν πολύ καλή απόδοση και η χρήση της MM έχει αυξηθεί κατά τη μεταφραστική διαδικασία. Η παρούσα, λοιπόν, διπλωματική αποτελεί μια συγκριτική μελέτη δυο νευρωνικών συστημάτων MM, έτσι ώστε να αξιολογηθεί η ποιότητα των αποτελεσμάτων τους στο ζεύγος Αγγλικά-Ελληνικά. Επιπλέον θα εξεταστεί ποιο από τα δυο συστήματα δίνει καλύτερα αποτελέσματα σε δυο διαφορετικά είδη κειμένου αλλά και ποιες είναι οι αδυναμίες του κάθε συστήματος για κάθε κειμενικό είδος ξεχωριστά. Η μέθοδος αξιολόγησης που θα εφαρμοστεί είναι η ανθρώπινη αξιολόγηση της MM. Στο θεωρητικό κομμάτι της εργασίας παρουσιάζεται, αρχικά, μια σύντομη ιστορική αναδρομή, τα είδη των συστημάτων MM, στη συνέχεια ορίζεται η ποιότητα της μετάφρασης γενικότερα και παρατίθενται διάφορες θεωρίες, οι οποίες αναπτύχθηκαν γύρω από την αξιολόγηση της αλλά και κάποιες μέθοδοι αξιολόγησης. Αργότερα, αναλύεται η αξιολόγηση της ποιότητας της MM και οι μέθοδοι διεξαγωγής της. Στο ερευνητικό μέρος, παρουσιάζεται η μεθοδολογία της έρευνας, τα αποτελέσματα που προκύπτουν και τα συμπεράσματα, στα οποία καταλήγουμε σύμφωνα με τα αποτελέσματα.

Η καινοτομία της συγκεκριμένης έρευνας είναι η συγκριτική μελέτη των δυο συστημάτων MM (GNMT και PureNMT) στο γλωσσικό συνδυασμό Αγγλικά-Ελληνικά. Εξ όσων γνωρίζω, δεν έχει γίνει παρόμοια συγκριτική μελέτη για τα συγκεκριμένα συστήματα στον συγκεκριμένο γλωσσικό συνδυασμό. Επιπλέον, η μελέτη θα γίνει σε δυο διαφορετικά είδη κειμένου, ένα νομικό και ένα ιατρικό, για να εξεταστεί η αποτελεσματικότητα των συστημάτων σε δυο διαφορετικά κειμενικά είδη. Με λίγα λόγια, η συγκεκριμένη έρευνα μελετά την απόδοση των συστημάτων για κάθε κειμενικό είδος στα ελληνικά με ΓΠ τα αγγλικά, έρευνα που θα μπορούσε να πυροδοτήσει την περαιτέρω και εκτενή μελέτη της απόδοσης των δύο για νομικά και ιατρικά κείμενα.

2. Σύνοψη Ιστορική Αναδρομή

Η Μηχανική Μετάφραση (MT) έκανε την εμφάνισή της στα τέλη του 1940 και ξεκίνησε με αφορμή την αποκωδικοποίηση. Ο Warren Weaver ήταν ο πρώτος που ανέφερε ότι η μηχανική μετάφραση θα μπορούσε να είναι εφικτή, καθώς θεωρούσε πως η μετάφραση αποτελεί η ίδια ένα είδος αποκωδικοποίησης. Παρόλο που ο ίδιος δεν γνώριζε πως ακριβώς να το προσεγγίσει, έθεσε τον προβληματισμό του στον καθηγητή Norbert Wiener, ο οποίος θεώρησε πως λόγω των διαφορετικών δομών των γλωσσών, είναι αδύνατον με τη τεχνολογία της εποχής να δημιουργηθεί μια τέτοια μηχανή. Μάλιστα, ο Weaver έγραψε σε ένα υπόμνημα με προτάσεις για την αντιμετώπιση του προβλήματος που θα πρόκυπτε λόγω της αμφισημίας, χρησιμοποιώντας τις γνώσεις του πάνω στην κρυπτογραφία. Το υπόμνημα αυτό (Weaver, 1949) αποτέλεσε την απαρχή της έρευνας στη μηχανική μετάφραση. Αργότερα, όμως, μια ομάδα ερευνητών αποφάσισε να ασχοληθεί με το μεταφραστικό αυτό πρόβλημα, προσεγγίζοντας το αρχικά από τη δημιουργία ενός “μηχανικού” λεξικού. Το πρώτο σύστημα παρουσιάστηκε στο κοινό, τελικά, το 1954 ήταν ένα σύστημα αυτόματης μετάφρασης μεταξύ δύο μόνο γλωσσών, που δημιουργήθηκε από την ερευνητική ομάδα του Πανεπιστημίου του Georgetown και την IBM (Hutchins, 2004). Αυτή η ανακάλυψη πυροδότησε και την περαιτέρω έρευνα πάνω στη μηχανική μετάφραση μέχρι που το 1966, η επιτροπή ALPAC (Automatic Language Processing Advisory Committee) προέβη στη δημοσίευση μιας δήλωσης, η οποία ουσιαστικά καταδίκασε και την δημιουργία ενός συστήματος που θα παράγει αξιόλογα αποτελέσματα (ALPAC, 1966). Από τότε μειώθηκε η ερευνητική δραστηριότητα, μέχρι το 1980 όπου το ερευνητικό πρόγραμμα EUROTRA (Johnson et al., 1985) έκανε τις πρώτες απόπειρες για ένα σύστημα που θα μετέφραζε στις επίσημες γλώσσες της Ευρωπαϊκής Ένωσης. Το σύστημα που ανέπτυξε το EUROTRA συνδύαζε λεξική, συντακτική και σημασιολογική πληροφορία και ήταν μια αξιόλογη εξέλιξη, έτσι ώστε να ξεκινήσουν ξανά οι έρευνες για πιο αποτελεσματικά συστήματα μηχανικής μετάφρασης. Κατά τη δεκαετία του 1990, με την εξάπλωση του διαδικτύου ξεκίνησαν διάφορες εταιρίες, όπως η IBM, η Altavista, η Systran και η Google, να δημιουργούν τα δικά τους συστήματα τα οποία ήταν στατιστικά συστήματα μηχανικής μετάφρασης (Statistical MT, SMT) και συστήματα με χρήση παραδειγμάτων (Example-based MT, EBM). Καθώς η χρήση του διαδικτύου έγινε πιο εύκολη και ευρέως προσβάσιμη, ταυτόχρονα γινόταν και πιο εκτεταμένη χρήση της αυτόματης μετάφρασης. Όλο και περισσότερες εταιρίες δημιουργούσαν τα δικά τους συστήματα, μέχρι και τη τελευταία δεκαετία όπου έγινε η μεγαλύτερη εξέλιξη στον τομέα αυτό με την εισχώρηση των νευρωνικών δικτύων και της τεχνητής νοημοσύνης στη μηχανική μετάφραση, τα λεγόμενα νευρωνικά συστήματα μηχανικής μετάφρασης (NMT) (Kalchbrenner and Blunsom, 2013).

3. Συστήματα Μηχανικής Μετάφρασης

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

Τα οντολογικά συστήματα (Knowledge-Based Machine Translation) χρησιμοποιούν βάσεις γνώσεων και τεχνικές τεχνητής νοημοσύνης για την μετάφραση ενός κειμένου από μια γλώσσα σε μια άλλη. Τα συστήματα που ανήκουν σε αυτόν τον τύπο είναι τα συστήματα μηχανικής μετάφρασης βάσει κανόνων (Rule-Based Machine Translation, RBMT). Υπάρχουν τρεις τύποι συστημάτων μηχανικής μετάφρασης βάσει κανόνων (RBMT) και αυτά είναι το λεξιλογικό σύστημα (Dictionary Based Machine Translation), το σύστημα μεταφοράς (Transfer Based Machine Translation), και 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. Τα συστήματα μεταφοράς, από την άλλη, λειτουργούν βάσει των διαφορών στη δομή της κάθε γλώσσας. Στο σύστημα αυτό, 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, έτσι ώστε να μεταφερθούν στην άλλη γλώσσα και τέλος γίνεται η παραγωγή στη ΓΣ. Τα διαγλωσσικά συστήματα, τα οποία 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). Το τρίγωνο μετάφρασης του Vauquois (1975) αποτελεί έναν τρόπο αναπαράστασης των διαφορετικών αρχιτεκτονικών αυτού του είδους μηχανικής μετάφρασης.



Εικόνα 1. Το τρίγωνο μετάφρασης Vauquois, (Πηγή: <https://opencourses.ionio.gr/modules/document/file.php/DDI106/9-MT.pdf>)

3.2 Συστήματα Μηχανικής Μετάφρασης Βάσει Παραδειγμάτων (Example-Based Machine Translation, EBMΤ)

Τα συστήματα μηχανικής μετάφρασης βάσει παραδειγμάτων (Example-Based Machine Translation) βασίζονται στη λειτουργία τους με παράλληλα σώματα κειμένου. Το σύστημα εκπαιδεύεται έτσι ώστε 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).

3.3 Στατιστικά Συστήματα Μηχανικής Μετάφρασης (Statistical Machine Translation, SMT)

Τα στατιστικά συστήματα μηχανικής μετάφρασης (Statistical Machine Translation, SMT) βασίζονται στις πιθανότητες (Brown, et.al, 1993). Με βασική ιδέα τη θεωρία της πληροφορίας του Claude Shannon και τη θεωρία των πιθανοτήτων και της στατιστικής, 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. Οι λέξεις αντιστοιχίζονται από τη ΓΠ στην αντίστοιχη της ΓΣ. Η αντιστοιχία αυτή δεν γίνεται βάσει της θέσης των λέξεων στην εκάστοτε πρόταση, αλλά η κάθε λέξη της ΓΠ μπορεί να αντιστοιχεί σε μια ή παραπάνω – και σε ορισμένες περιπτώσεις σε καμία - λέξεις της ΓΣ, λόγω των μορφολογικών διαφοροποιήσεων κάθε γλώσσας.

Η επόμενη εξέλιξη στα στατιστικά συστήματα αφορά την εσωτερική επεξεργασία ενός κειμένου σε επίπεδο φράσης (Phrase-Based Machine Translation, PBMT), η οποία φράση μπορεί να περιλαμβάνει μια ή παραπάνω λέξεις. Το σύστημα επεξεργάζεται την πιθανότητα συνεμφάνισης μιας φράσης σε μια ΓΠ με μια φράση στη ΓΣ (Kalchbrenner, Blunsom, 2013). Σύμφωνα με την έρευνα των Koehn et. Al (2003), στο επίπεδο φράσης, όταν το περιεχόμενο του κειμένου, με το οποίο εκπαιδεύεται το σύστημα είναι απλό, τότε τα αποτελέσματα είναι καλύτερα και επίσης, υπάρχει μεγαλύτερη ακρίβεια όταν η φράση περιλαμβάνει έως και τρεις λέξεις. Η ίδια έρευνα απέδειξε με τη σύγκριση πολλών συστημάτων σε επίπεδο φράσης με παραδοσιακά στατιστικά συστήματα εκπαιδευμένα σε επίπεδο λέξης, ότι η εκπαίδευση σε επίπεδο φράσης δίνει καλύτερα αποτελέσματα. Σημαντικό ρόλο παίζει στα συγκεκριμένα συστήματα η κατανομή βάρους σε συμφράσεις, έτσι ώστε να αυξηθούν οι πιθανότητες επανεμφάνισης τους.

3.4 Υβριδικά Συστήματα Μηχανικής Μετάφρασης (Hybrid Machine Translation)

Η τελευταία εξέλιξη στη ΜΜ, πριν την εμφάνιση των νευρωνικών συστημάτων (Neural Machine Translation, NMT), ήταν τα υβριδικά συστήματα ΜΜ (Hybrid Machine Translation), τα οποία αποτελούνται από δυο τύπους συστημάτων, το στατιστικό σύστημα (Statistical Machine Translation) και το σύστημα ΜΜ βάσει κανόνων (Rule-Based Machine Translation). Υπάρχουν τρία είδη υβριδικών συστημάτων και αυτά είναι η σύζευξη των συστημάτων (coupling of systems), προσαρμογές στην αρχιτεκτονική (architecture adaptations) και τα γνήσια υβριδικά συστήματα (genuine hybrid systems). Η πρώτη κατηγορία είναι η σύζευξη,

κατά την οποία γίνεται ένας συνδυασμός των στατιστικών συστημάτων (SMT) με τα συστήματα μηχανικής μετάφρασης βάσει κανόνων (RBMT). Η σύζευξη λειτουργεί με δυο τρόπους. Από τη μια, η μετεπεξεργασία (Post-Editing, PE) του κειμένου που προέρχεται από το σύστημα βάσει κανόνων (RBMT) γίνεται με τους κανόνες που διέπουν την μετεπεξεργασία (PE) κειμένων που προέρχονται από στατιστικό σύστημα μηχανικής μετάφρασης (SMT). Από την άλλη, η δημιουργία ενός μεταφράσματος καλής ποιότητας γίνεται με τον συνδυασμό μεταφράσεων που προκύπτουν από τα δυο συστήματα MM. Όσον αφορά τις προσαρμογές στην αρχιτεκτονική, η αρχιτεκτονική των συστημάτων ακολουθεί ένα από τα δύο συστήματα, το στατιστικό ή το σύστημα βάσει κανόνων, και προστίθενται πρακτικές του άλλου συστήματος αντίστοιχα. Αυτό επιτυγχάνεται με δυο τρόπους, με την προεπεξεργασία (Pre-Editing) 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).

3.5 Νευρωνικά Συστήματα Μηχανικής Μετάφρασης (Neural Machine Translation, NMT)

Η πιο πρόσφατη εξέλιξη στον τομέα της μηχανικής μετάφρασης αφορά την δημιουργία συστημάτων βάσει νευρωνικών δικτύων. Οι πρώτες αναφορές για χρήση αυτού του τύπου δικτύων στη μηχανική μετάφραση έγιναν στις έρευνες των Kalchbrenner και Blunsom (2013), των Cho et al. (2014b), και των Sutskever et al. (2014). 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 έτσι ώστε να έχουμε σαν αποτέλεσμα 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. Άγνωστη παραμένει, προς το παρόν, η πλήρης λειτουργία των δικτύων αυτών, όμως η απόδοσή τους φαίνεται να δίνει πολύ καλά αποτελέσματα. Τα νευρωνικά συστήματα MM 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, έτσι ώστε να έχουν καλύτερα αποτελέσματα για τα κείμενά τους.

4. Ποιότητα της Μετάφρασης

Για να γίνει κατανοητή η αξιολόγηση της ποιότητας της MM είναι απαραίτητο να οριστεί, αρχικά, η ποιότητα της μετάφρασης γενικότερα, και να γίνει αναφορά στην αξιολόγησή της, η οποία έχει αποδειχθεί ένα δύσκολο εγχείρημα για τους μεταφρασεολόγους. Ο ορισμός της

ποιότητας της μετάφρασης αποτελεί μεγάλο πρόβλημα για τους ερευνητές. Υπάρχουν πολλές αντικρουόμενες θεωρίες για τον ορισμό και πολλοί ερευνητές έχουν αποπειραθεί να την ορίσουν με διάφορους τρόπους. Οι Fields et al. (2014) προσεγγίζουν την ποιότητα της μετάφρασης, αρχικά, μελετώντας τη θεωρία του Garvin (1984) για τον ορισμό της ποιότητας σε γενικότερο επίπεδο, η οποία συνοπτικά αποτελείται από ένα πλαίσιο ορισμού της ποιότητας με πέντε προσεγγίσεις, οι οποίες συναντώνται στους τομείς της φιλοσοφίας, των οικονομικών, του μάρκετινγκ, της λειτουργικής διαχείρισης και της οικονομίας, και καταλήγουν σε δυο ορισμούς, έναν ευρύ ορισμό (broad definition) και έναν στενό ορισμό (narrow definition). Οι δυο αυτοί ορισμοί είναι αντικρουόμενοι και η διαφορά τους είναι ότι ο στενός ορισμός αντιμετωπίζει τη μετάφραση ως μια διαδικασία που επικεντρώνεται αυστηρά στο κείμενο ενώ ο ευρύς έχει ως δεδομένο ότι η μετάφραση αφορά διάφορες διαδικασίες που υπερβαίνουν το κείμενο (π.χ. τοπική προσαρμογή/localization κ.α.) (Koby et al., 2014). Ο ευρύς ορισμός που περιλαμβάνει όλες τις διαδικασίες της μετάφρασης είναι:

«Μια ποιοτική μετάφραση αποδίδει την ακρίβεια και την ευχέρεια που απαιτείται για τον αποδέκτη και τον σκοπό και συμμορφώνεται με τις όλες τις υπόλοιπες προδιαγραφές που τίθενται μεταξύ του αιτούντος και του παρόχου, λαμβάνοντας υπόψη τις ανάγκες του τελικού χρήστη.»¹ (Koby et al., 2014).

Γι' αυτόν τον 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.

Από την άλλη μεριά έχουμε τον στενό ορισμό, ο οποίος δεν αναγνωρίζει τις άλλες διαδικασίες (π.χ. τοπική προσαρμογή, κ.α.) στη μετάφραση. Στην περίπτωση της εφαρμογής Μηχανικής Μετάφρασης στο κείμενο χρειάζεται να γίνει πλήρης μετεπεξεργασία (Full-PE) του μεταφράσματος που προκύπτει 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 υψηλής ποιότητας μετάφραση είναι αυτή, η οποία μεταφέρει ολόκληρο το μήνυμα του κειμένου-πηγή στο κείμενο-στόχος, συμπεριλαμβανομένης της δήλωσης, της συνδύλωσης, της σημασιολογικής διαφοράς και του ύφους, και το κείμενο-πηγή πρέπει να είναι γραμμένο στη γλώσσα-στόχος με χρήση σωστής γραμματικής και σειράς των λέξεων, έτσι ώστε να παραχθεί ένα σωστό πολιτισμικά κείμενο, το οποίο στις περισσότερες περιπτώσεις κατά την ανάγνωση να φαίνεται σαν να προέρχεται από φυσικό ομιλητή της γλώσσας-στόχος και να αφορά το κοινό του πολιτισμού-στόχος.»¹

Ο στενός ορισμός θα μπορούσε να αποτελεί μια συγκεκριμένη περίπτωση του ευρύ ορισμού με διάφορες προκαθορισμένες προδιαγραφές. Παρ' όλα αυτά, 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). Ο ορισμός της ποιότητας της μετάφρασης είναι ένα δύσκολο εγχείρημα και υπάρχουν πολλές διαφορετικές θεωρίες που αφορούν την ποιότητα της μετάφρασης αλλά και την ίδια τη μετάφραση, οι οποίες πολλές φορές συγκρούονται. Αυτό δυσκολεύει την δημιουργία ενός καθολικού

¹ Η μετάφραση των παραθεμάτων είναι της γράφουσας.

ορισμού που θα μπορέσει να προσδιορίσει τους μελετητές, έτσι ώστε να προχωρήσουν στη δημιουργία μιας θεωρίας για την αξιολόγηση της ποιότητας της μετάφρασης.

4.1 Εξασφάλιση της ποιότητας της μετάφρασης

Η εξασφάλιση της ποιότητας ενός μεταφράσματος αποτελεί ένα αναπόσπαστο κομμάτι της μεταφραστικής διαδικασίας. Για την αξιολόγηση της ποιότητας μιας μετάφρασης έχουν αναπτυχθεί πολλές θεωρίες και έχει απασχολήσει αρκετά την ακαδημαϊκή κοινότητα. Οι μελετητές, όχι μόνο δεν έχουν καταλήξει σε μια καθολική θεωρία για την αξιολόγηση, αλλά ούτε έχουν καταλήξει σε ένα καθολικό τρόπο διαχωρισμού των θεωριών και των προσεγγίσεων που υπάρχουν. Η Lauscher (2000) παρουσιάζει έρευνες από διάφορες οπτικές με επίκεντρο είτε τον λειτουργικό χαρακτήρα του πρωτοτύπου, είτε βασιζόμενες στη θεωρία του σκοπού, που έχει αναλυθεί από διάφορους μελετητές (Reiß & Vermeer 1984; Holz-Mänttari 1984; Vermeer 1989, 1996; Nord 1993). Οι τελευταίες αυτές 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 πρωτοτύπου κειμένου και στο μετάφρασμα. Οι δύο αυτές διαφορετικές προσεγγίσεις ονομάζονται προσέγγιση βάσει της ισοδυναμίας (Equivalence-Based approaches) και λειτουργικές προσεγγίσεις (Functional approaches), αντίστοιχα. Οι προσεγγίσεις βάσει της ισοδυναμίας υποστηρίζουν ότι η μετάφραση είναι μια προσπάθεια αναπαραγωγής του ΚΠ όσο πιο κοντά γίνεται στο πρωτότυπο, ενώ οι λειτουργικές θέλουν τη μετάφραση να είναι η διαδικασία της παραγωγής ενός κειμένου βάσει ενός πρωτοτύπου, όμως το ΚΣ αποτελεί αυτοτελές κείμενο (Lauscher, 2000).

Άλλος ένας διαχωρισμός των θεωριών έγινε από διαφορετικές οπτικές. Οι δύο αυτές οπτικές είναι η Ψυχολογική-Κοινωνική Προσέγγιση και η Κειμενοστρεφής Προσέγγιση. Στην Ψυχολογική-Κοινωνική Προσέγγιση υπάρχουν δυο θεωρίες, η πνευματική, η οποία βασίζεται στο πόσο καλή ή κακή μπορεί να φανεί σε κάποιον μια μετάφραση, αλλά και υποστηρίζουν ότι η μετάφραση εξαρτάται από την αντίληψη και τις εμπειρίες του μεταφραστή (Stolze, 2003). Παρ' όλα αυτά, αυτή η θεωρία δεν βασίζεται σε κάποιους συγκεκριμένους κανόνες αλλά σε εντυπώσεις. Η άλλη θεωρία βασίζεται στην ανταπόκριση. Στο πλαίσιο αυτής της θεωρίας υπάρχει ένας διαχωρισμός των ερευνητών, σε αυτούς που δίνουν βάση στη συμπεριφορά του κοινού-στόχος, τους συμπεριφοριστές (behaviourists) και σε αυτούς που υπερασπίζονται τη θεωρία του σκοπού, τους 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 πρωτοτύπου κειμένου και στο μετάφρασμα (Reiß and Vermeer, 1984). Όσον αφορά την Κειμενοστρεφή Προσέγγιση έχουμε τρεις θεωρίες, τις περιγραφικές μεταφραστικές μελέτες, την μεταμοντέρνα – αποδομημένη προσέγγιση, και 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. Παρ' όλα αυτά, το μετάφρασμα δεν είναι ένα αυτοτελές κείμενο στη ΓΣ αλλά συνδέεται με ένα προϋπάρχον κείμενο που απευθύνεται σε έναν άλλον πολιτισμό (Toury, 1995). Οι ερευνητές της μεταμοντέρνας προσέγγισης, όπως ο Venuti (2004), υποστηρίζουν ότι στη μετάφραση παίζει ρόλο ο παράγοντας του κειμενικού είδους, καθώς σε κάποια είδη γίνεται αλλοίωση του ΚΠ, στο οποίο εκφράζονται κάποιες ιδεολογίες, και μεταφέρονται διαφορετικά στο ΚΣ, έτσι ώστε να ικανοποιηθούν πιθανά συμφέροντα. Για να αξιολογήσει κανείς μια μετάφραση θα πρέπει να έχει μια πολύ καλή κριτική ικανότητα και να διεξάγει πολλαπλές αναλύσεις, καθώς και να εξετάσει 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. Τέλος, 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 πολλών άλλων. Στη συγκεκριμένη προσέγγιση, οι μελετητές ανέπτυξαν λεπτομερείς διαδικασίες για ανάλυση και αξιολόγηση. Πολλές από αυτές λαμβάνουν υπόψη την σύνδεση κειμένου και συγκεκριμένου καθώς η γλώσσα και ο πραγματικός κόσμος είναι άρρηκτα συνδεδεμένα (House, 2014).

4.2 Μέθοδοι αξιολόγησης της ποιότητας της μετάφρασης

Εμπειρικές προσεγγίσεις (Experience-based approaches)

Οι εμπειρικές προσεγγίσεις της αξιολόγησης της ποιότητας είναι αυτές που έχουν δημιουργηθεί από κάποιον οργανισμό ή εταιρία για την αξιολόγηση της ποιότητας των μεταφράσεων τους (π.χ. LISA QA model). Αυτού του είδους οι μετρικές είναι ακριβείς για τους σκοπούς των οργανισμών, όμως είναι δύσκολο να προσαρμοστούν σε διαφορετικά περιβάλλοντα γιατί δεν ενέχουν ήδη διαμορφωμένα θεωρητικά πλαίσια. Από τη μια, αυτό μας περιορίζει, διότι δεν μπορεί να γίνει αναφορά στη θεωρία και από την άλλη, λόγω της έλλειψης πλαισίου υπάρχει ο κίνδυνος το αποτέλεσμα της αξιολόγησης να είναι επηρεασμένο από την υποκειμενική κρίση του αξιολογητή.

Θεωρητικές και ερευνητικές προσεγγίσεις (Theoretical and research based approaches)

Οι συγκεκριμένες προσεγγίσεις ή αλλιώς προσεγγίσεις βάσει της ισοδυναμίας (equivalence-based approach) (Lauscher, 2000) βασίζονται στον αποδέκτη της μετάφρασης και στο κείμενο. Αυτές προκύπτουν από ένα θεωρητικό πλαίσιο ή μια υπόθεση για τη φύση της μετάφρασης. Αυτές, όμως, οι προσεγγίσεις έχουν και αρκετές ανακρίβειες, οι οποίες δεν εξασφαλίζουν την επιτυχία τους και τη γενική εφαρμογή τους. Αντιθέτως, επικεντρώνονται σε ορισμένες πτυχές της ποιότητας, ενώ η αξιολόγηση πρέπει να περιλαμβάνει μεγάλο εύρος θεμάτων.

Προσέγγισεις βάσει της αντίδρασης του αναγνώστη (Reader-response approaches)

Τέτοιου είδους προσεγγίσεις αξιολογούν την ποιότητα σύμφωνα με την αντίδραση του αποδέκτη της μετάφρασης στον πολιτισμό-στόχο 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).

Κειμενικές και πραγματολογικές προσεγγίσεις (Textual and pragmatic approaches)

Η συγκεκριμένη προσέγγιση είναι πολύ σημαντική καθώς στρέφει την προσοχή από την καταμέτρηση λαθών στο επίπεδο λέξης/πρότασης στην αξιολόγηση ολόκληρων κειμένων και στον στόχο της μετάφρασης, βάζοντας στο επίκεντρο τον αναγνώστη και την επικοινωνία. Κάποια μοντέλα συγκρούονται, είτε επειδή κάποια 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.

Λειτουργική προσέγγιση (Functional – Componential approach)

Σύμφωνα με τη Colina (2008), οι μέθοδοι αξιολόγησης δεν έχουν επιτύχει μια ισορροπία μεταξύ θεωρίας και πράξης και προτείνει μια μέθοδο, η οποία μπορεί να εφαρμοστεί και στον επαγγελματικό κλάδο και στον 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:

- Βασίζεται στη λειτουργιστική θεωρία και στα κειμενικά μοντέλα της μετάφρασης
- Περιλαμβάνει υπόθεση, η οποία μπορεί να δοκιμαστεί
- Δηλώνει το περιεχόμενο και τους στόχους της
- Είναι τμηματική. Περιλαμβάνει κατηγορίες για αξιολόγηση (π.χ. γλωσσολογική μορφή, λειτουργική ακρίβεια κ.α.)

Οι λειτουργικές θεωρίες της Colina (2008) για την αξιολόγησης της ποιότητας της μετάφρασης μπορούν να συνοψιστούν σε δυο γενικότερες κατηγορίες μεθόδων, την ολιστική προσέγγιση και την αναλυτική, ή αλλιώς ανάλυση βάσει των λαθών. Η ολιστική μέθοδος αξιολόγησης εξετάζει το κείμενο στο σύνολό του και βασίζεται σε καθολικά κριτήρια. Εν αντιθέσει η αναλυτική προσέγγιση εξετάζει μεμονωμένα τυχόντα λάθη μέσα σε ένα κείμενο. Η δυο αυτές προσεγγίσεις εξετάζουν διαφορετικά ένα κείμενο, με την ολιστική μέθοδο να δίνει μια καθολική εικόνα του κειμένου και αν αυτό πληροί τα ειδικά κριτήρια, ενώ η αναλυτική προσέγγιση εντοπίζει τα συγκεκριμένα λάθη ενός κειμένου χωρίς όμως να παρέχει μια γενικότερη εντύπωση του μεταφράσματος (Lommel and Melby, 2018).

5. Αξιολόγηση της ποιότητας της Μηχανικής Μετάφρασης

Η αξιολόγηση της ποιότητας της MM δεν έχει μεγάλες διαφορές σε σχέση με την αξιολόγηση της ποιότητας μιας μετάφρασης που δημιουργείται από τον άνθρωπο. Παρ' όλα αυτά, υπάρχουν διαφορές στη διαδικασία, λόγω της διαφορετικής φύσης της μετάφρασης. Ένα σύστημα, ανάλογα το είδος του, έχει συγκεκριμένη συμπεριφορά στα αποτελέσματα που παράγει, βάσει του κειμενικού είδους, του γλωσσικού συνδυασμού κ.α.. Η αξιολόγηση της ποιότητας των συστημάτων MM γίνεται για διάφορους λόγους, 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, έτσι ώστε να βελτιώνεται η απόδοσή του και στην περίπτωση ενός νέου συστήματος, πρέπει να εξακριβώνεται ότι αποδίδει καλύτερα σε σχέση με το προηγούμενο (Snover, 2011).

Συνηθέστερα, έχουμε δύο τύπους αξιολόγησης της ποιότητας της MM. Η μία προσέγγιση υπολογίζει την ποιότητα ενός συστήματος βάσει των εσωτερικών λειτουργιών του και ονομάζεται «Glass Box», ενώ η άλλη υπολογίζει την ποιότητα βάσει μόνο του αποτελέσματος της MM και ονομάζεται «Black Box». Στη δεύτερη προσέγγιση, εντοπίζονται 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.

Στη προσέγγιση «Black Box» υπάρχουν δυο διαφορετικές μετρικές, οι οποίες υπολογίζουν της ακρίβεια και τη χρησιμότητα του αποτελέσματος της MM, η εσωτερική (intrinsic) και η εξωτερική (extrinsic). Η εσωτερική εστιάζει στην ποιότητα του αποτελέσματος της MM και συχνά συγκρίνουν αυτό με μια ή περισσότερες μεταφράσεις-αναφορές, οι οποίες θεωρούνται υψηλής ποιότητας. Από τη μια, οι ανθρώπινες εσωτερικές μετρικές αξιολόγησης υπολογίζουν την ποιότητα με αντικειμενικότητα, βάσει κάποιων χαρακτηριστικών του αποτελέσματος της MM, συνηθέστερα υπολογίζοντας την ακρίβεια (accuracy) και την ευχέρεια (fluency). Από την άλλη, οι αυτόματες εσωτερικές μετρικές υπολογίζουν την ομοιότητα των προτάσεων ανάμεσα στο αποτέλεσμα της MM με τις μεταφράσεις αναφοράς (μια ή περισσότερες ανάλογα τη μετρική). Αντιθέτως, 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.

Υπάρχει άλλος ένας διαχωρισμός της αξιολόγησης της ποιότητας των αποτελεσμάτων της MM και αυτός είναι αξιολόγηση βάσει μεταφράσεων-αναφοράς (Reference-Based MT evaluation) και αξιολόγηση χωρίς μεταφράσεις αναφοράς (Reference-Less MT evaluation) (Lavie, 2011). Κατά την πρώτη περίπτωση η αξιολόγηση γίνεται βάσει μεταφράσεων του πρωτοτύπου από τον άνθρωπο και είναι μια μέθοδος που χρησιμοποιείται συχνά στην αξιολόγηση. Κατά τη δεύτερη περίπτωση, η αξιολόγηση γίνεται χωρίς μεταφράσεις-αναφοράς (Bojar et al. 2015).

Πιο πρακτικά, λοιπόν, έχουν εφαρμοστεί διάφοροι τρόποι αξιολόγησης της ποιότητας της MM. Αυτοί χωρίζονται σε δυο βασικούς και αυτοί είναι η ανθρώπινη και η αυτόματη

αξιολόγηση. Η ανθρώπινη αξιολόγηση ξεκινά λίγο αργότερα από τη MM για την αξιολόγηση της ποιότητας των αποτελεσμάτων των αυτόματων μηχανών μετάφρασης, καθώς ανέθεταν σε εκπαιδευμένους για αξιολόγηση ανθρώπους να κρίνουν την ποιότητα της μετάφρασης συγκριτικά με μεταφράσεις που είχαν παραχθεί από επαγγελματίες μεταφραστές. Χρησιμοποιούσαν, δηλαδή, υποκειμενικές απόψεις επισημειωτών για να βαθμολογήσουν την ποιότητα της MM, ως προς την ακρίβεια και την ευχέρεια (Snover, 2011).

5.1 Ακρίβεια και Ευχέρεια

Στην ανθρώπινη αξιολόγηση δυο είναι οι πιο συνήθεις μετρικές, η ακρίβεια (accuracy) και η ευχέρεια (fluency) (White et al., 1994). Ο υπολογισμός της ακρίβειας υποδεικνύει αν το αποτέλεσμα της MM περιέχει τις απαραίτητες πληροφορίες σύμφωνα με το πρωτότυπο. Στον υπολογισμό της ακρίβειας, ο επισημειωτής δεν πρέπει να λαμβάνει καθόλου υπόψη αν το αποτέλεσμα της MM είναι γραμματικά και συντακτικά σωστό, αλλά να επικεντρώνεται στη σωστή μεταφορά του νοήματος. Ο επισημειωτής στην προκειμένη πρέπει να γνωρίζει καλά τη ΓΠ και τη ΓΣ για να μπορεί να εντοπίσει αν μεταφέρεται σωστά το νόημα. Παρ' όλα αυτά, αν ένας μονόγλωσσος επισημειωτής αξιολογήσει την ακρίβεια, πρέπει να γνωρίζει άπταιστα τη ΓΣ και να συγκρίνει το αποτέλεσμα της MM με υψηλής ποιότητας μεταφράσεις που έχουν παραχθεί από τον άνθρωπο. Για την αξιολόγηση της ευχέρειας είναι αρκετός ένας ομιλητής της γλώσσας-στόχος, 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 αν το νόημα του πρωτοτύπου έχει αποτυπωθεί και στο μετάφρασμα που προέρχεται από MM. Συνηθέστερα, η αξιολόγηση της ακρίβειας και της ευχέρειας γίνεται στο επίπεδο της πρότασης. Δηλαδή η κάθε πρόταση ξεχωριστά κρίνεται σε μια κλίμακα 5 (ή λιγότερων ή περισσότερων) βαθμίδων για το κατά πόσο είναι ακριβής νοηματικά ή γραμματικά και συντακτικά άρτια αντίστοιχα σε κάθε περίπτωση (Przybocki et al., 2008).

5.2 Κατάταξη

Ένας άλλος τρόπος αξιολόγησης της ποιότητας των αποτελεσμάτων της MM είναι η κατάταξη (ranking) (Callison-Burch et al., 2007). Γι' αυτή τη διαδικασία, οι επισημειωτές καλούνται να κατατάξουν (συνήθως από το καλύτερο προς το χειρότερο) αντιπαραβολικά διάφορες μεταφράσεις μιας πρότασης. Ουσιαστικά, αυτή η μέθοδος αξιολόγησης είναι μια απλοποιημένη μορφή της αξιολόγησης της ακρίβειας και της ευχέρειας, και χρησιμοποιείται συγκριτικά με άλλες μεταφράσεις. Και αυτή η μέθοδος γίνεται στο επίπεδο πρότασης, καθώς οι ισοδύναμες προτάσεις των μεταφράσεων αντιστοιχίζονται, έτσι ώστε να καταταχθούν αντιπαραβολικά. 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 και καταλήγουν μέσα από αυτές τις μετρικές να καταλήγουν σε μια κατάταξη των μεταφράσεων.

Όπως είδαμε στην αξιολόγηση της ποιότητας της μετάφρασης παραπάνω, αυτοί οι μέθοδοι αξιολόγησης για την MM ανήκουν στην ολιστική μέθοδο αξιολόγησης, καθώς αφορούν την αξιολόγηση του κειμένου σαν σύνολο, δίνοντας κάποιες βαθμολογίες για την ακρίβεια και

την ευχέρεια των αποτελεσμάτων. 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 «LISA QA Model» (LISA, 2004), η τυπολογία λαθών «MQM (Multidimensional Quality Metrics)»², η οποία θα χρησιμοποιηθεί στην παρούσα εργασία με κάποιες παραλλαγές, κ.α.

5.3 Κατηγοριοποίηση των λαθών

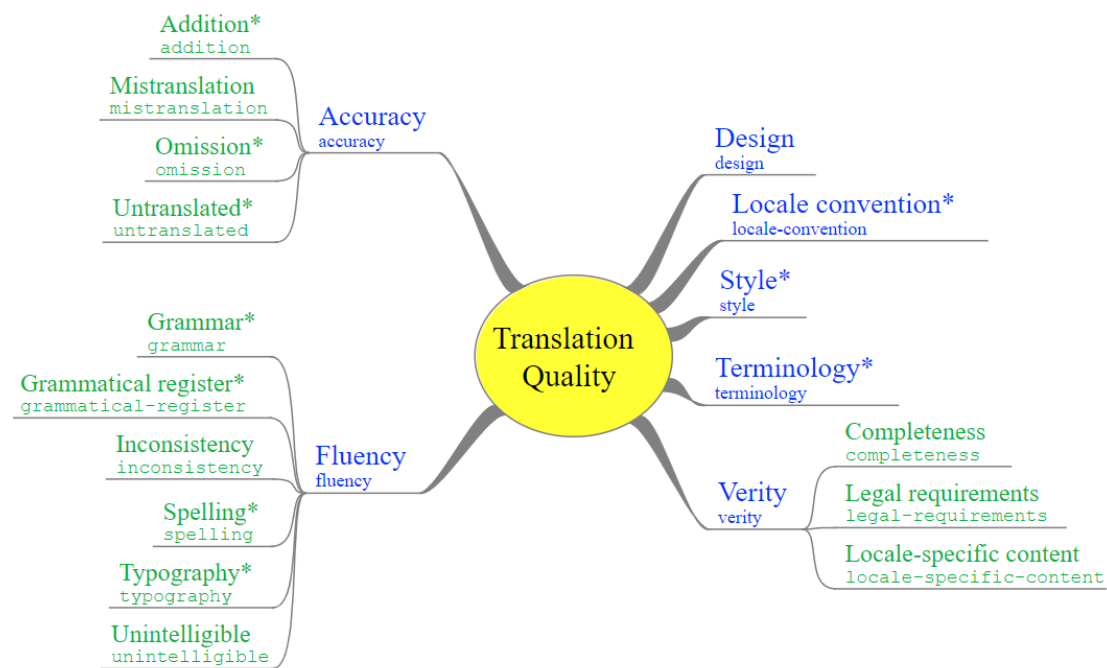
Η αναλυτική αυτή μέθοδος αξιολόγησης δίνει μια λεπτομερή εικόνα της απόδοσης ενός συστήματος, ενώ οι υπόλοιπες διαμορφώνουν μια γενική εικόνα της ποιότητας των αποτελεσμάτων (Klubička et al., 2018). Με αυτόν τον τρόπο εντοπίζονται τα δυνατά και αδύναμα σημεία ενός συστήματος ανάλογα με το ζεύγος γλωσσών. Γλώσσες με ποικιλία γραμματικών, λεξιλογικών και συντακτικών φαινομένων παρουσιάζουν πλήθος λαθών ανά σύστημα, γεγονός που απαιτεί μια αναλυτική μέθοδο αξιολόγησης, έτσι ώστε να εντοπιστούν. Το πρότυπο, το οποίο θα αναλυθεί σε αυτό το σημείο είναι η τυπολογία λαθών MQM¹, μια πρωτοβουλία του Γερμανικού Ερευνητικού Κέντρου Τεχνητής Νοημοσύνης (DFKI) σε συνεργασία με το QTLaunchPad. Το συγκεκριμένο πρότυπο δημιουργήθηκε για να εντοπίζονται λάθη σε μεταφράσματα. Στα μεταφράσματα συμπεριλαμβάνονται και αυτά που έχουν παραχθεί από διάφορες διαδικασίες, όπως κείμενα τοπικής προσαρμογής (localization) ή μεταδημιουργίας (transcreation). Γενικότερα, το συγκεκριμένο πρότυπο αφορά την αξιολόγηση της ποιότητας της μετάφρασης κειμένων ή και των ίδιων των πρωτοτύπων, καθώς το πρωτότυπο επηρεάζει το μετάφρασμα.

Η τυπολογία λαθών MQM αποτελείται από 9 κατηγορίες. Αυτές είναι:

- Ακρίβεια (Accuracy), με την οποία αξιολογούμε τη σχέση πρωτοτύπου και μετάφρασης. Σ' αυτή την κατηγορία περιλαμβάνεται η αλλοίωση του νοήματος, οι προσθήκες ή οι παραλήψεις, και άλλα παρόμοια λάθη.
- Μορφή (Design), η οποία αφορά τη φυσική παρουσία του κειμένου.
- Ευχέρεια (Fluency), η οποία αφορά τη γραμματικότητα ενός κειμένου, η οποία μπορεί να αξιολογηθεί ένα κείμενο ανεξάρτητα με το αν είναι μετάφραση ενός κειμένου ή όχι. Η συγκεκριμένη κατηγορία μπορεί να αφορά και ένα ΚΠ και ένα ΚΣ.
- Διεθνοποίηση (Internationalization), στην οποία εντοπίζονται λάθη που έχουν να κάνουν με την προετοιμασία ενός ΚΠ για μετάφραση ή για τοπική προσαρμογή. Λάθη αυτής της κατηγορίας μπορούν να εντοπιστούν στο ΚΣ και κυρίως εντοπίζονται μέσα από τις υποκατηγορίες της κατηγορίας «Εγχώριες Συμβάσεις»
- Εγχώριες Συμβάσεις (Locale Conventions), οι οποίες ορίζουν ότι το περιεχόμενο πρέπει να συμμορφώνεται σε συγκεκριμένες εγχώριες συμβάσεις. Αν 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 περιεχομένου ως προς την καταλληλότητα του για το κοινό-στόχος.
- Ύφος (Style), το οποίο αφορά 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.

² Πηγή: <http://www.qt21.eu/mqm-definition/definition-2015-12-30.html#creating-mqm-metrics>

- Ορολογία (Terminology). Σε αυτή την κατηγορία αξιολογείται η σωστή χρήση της ορολογίας σε συγκεκριμένα είδη κειμένων και η συνέπεια των συγκεκριμένων όρων σε όλο το κείμενο. Τα λάθη ορολογίας δεν πρέπει να συμπεριλαμβάνονται στην κατηγορία των παρανοήσεων (mistranslation), γιατί αυτή η κατηγορία περιλαμβάνει λάθη στη γενική γλώσσα του κειμένου και όχι λάθη στην ειδική ορολογία του κειμένου.
- Εγκυρότητα (Verity), η οποία αφορά τη 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.
- Συμβατότητα (Compatibility), η οποία περιλαμβάνει λάθη από άλλες μετρικές οι οποίες δεν είναι κατάλληλες για χρήση στο MQM (επειδή αυτές οι κατηγορίες δεν υπάρχουν στο MQM, όπως διορίες, λειτουργία λογισμικού κλπ. Συμπεριλαμβάνονται μόνο για συμβατότητα με τις παλαιότερες μετρικές και δεν πρέπει να χρησιμοποιούνται για νέες μετρικές MQM.
- Άλλο (Other). Σε αυτή την κατηγορία κατατάσσονται τα λάθη που δεν μπορούν να συμπεριληφθούν σε κάποια από τις υπόλοιπες κατηγορίες. Η κατηγορία «Άλλο» δεν πρέπει να χρησιμοποιείται συχνά.



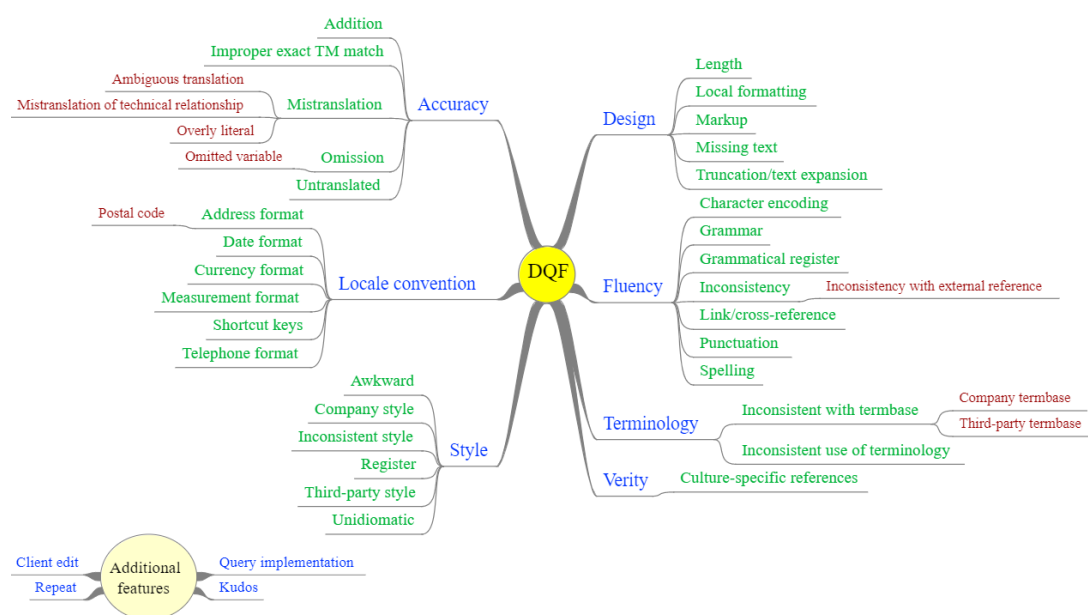
Εικόνα 2. Απεικόνιση της τυπολογίας λαθών MQM με τις βασικές κατηγορίες και τις υποκατηγορίες τους (Πηγή: <http://www.gt21.eu/mqm-definition/definition-2015-12-30.html#creating-mqm-metrics>)

Οι κατηγορίες Μορφή (Design), Ευχέρεια (Fluency), Ύφος (Style), Ορολογία (Terminology) και Εγκυρότητα (Verity) μπορούν να χρησιμοποιηθούν στην 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 τα ΚΣ ενώ στη Διεθνοποίηση (Internationalization) τα λάθη που εμπίπτουν βρίσκονται στο ΚΠ.

Στην Εικόνα 2 βλέπουμε τις 9 κατηγορίες για την αναλυτική αξιολόγηση της ποιότητας και τις υποκατηγορίες τους. Σύμφωνα με το παραπάνω σχήμα, στην κατηγορία της Ακρίβειας (Accuracy) βρίσκουμε τα εξής λάθη: 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) τα λάθη που εντοπίζονται είναι στη Γραμματική (Grammar), Γραμματικό Ύφος (Grammatical Register)³, Ασυνέπεια (Inconsistency), Ορθογραφία (Spelling), Τυπογραφικά λάθη (Typography) και Ακατανόητα λάθη (Unintelligible). Τέλος, στην κατηγορία της Εγκυρότητας έχουμε τρεις υποκατηγορίες λαθών και αυτές είναι: Πληρότητα (Completeness), Νομικές Προϋποθέσεις (Legal requirements) και Τοπικά σχετικό περιεχόμενο (Locale-specific content).

5.4 Τυπολογία λαθών DQF

Το πρότυπο της TAUS (Translation Automation User Society) για την τυπολογία των λαθών που ονομάζεται «DQF (Dynamic Quality Framework)» και είναι ένα αναγνωρισμένο υποσύνολο του MQM, στο οποίο έχουν προστεθεί κατηγορίες από τα μέλη της TAUS. Στην παρακάτω εικόνα (Εικόνα 3.) απεικονίζεται το πρότυπο DQF με τις επιπρόσθετες κατηγορίες.



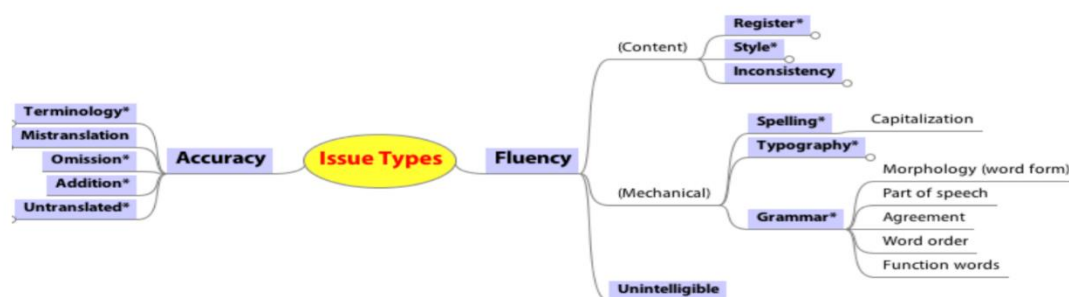
Εικόνα 3. Απεικόνιση της τυπολογίας λαθών DQF με τις βασικές κατηγορίες και τις υποκατηγορίες τους (Πηγή: <http://www.qt21.eu/mqm-definition/definition-2015-12-30.html#creating-mqm-metrics>)

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

³ Η κατηγορία αυτή αφορά την αξιολόγηση της σωστής χρήσης της γραμματικής ανάλογα με το ύφος του κειμένου. Π.χ. όταν σε ένα κείμενο απαιτείται η επίσημη μορφή των ρημάτων και όχι στοιχεία προφορικότητας.

5.5 Εναρμονισμένο πρότυπο MQM-DQF

Οι Lommel και Melby (2018) προτείνουν μια σύζευξη των δυο προτύπων έτσι ώστε να δημιουργήσουν ένα εναρμονισμένο πρότυπο τυπολογίας λαθών. Το συγκεκριμένο πρότυπο αποτελείται από ένα μέρος του κάθε προτύπου (των MQM και DQF) και έχει ως στόχο να αναγνωρίσει τη φύση των λαθών έτσι ώστε να αποφευχθούν ή να διορθωθούν και να τα συσχετίσει με ευρέως γνωστούς τύπους λαθών. Ένα υποσύνολο αυτού του προτύπου αφορά την αξιολόγηση της ποιότητας της MM, λόγω της ευελιξίας στις κατηγορίες των λαθών. Το συγκεκριμένο υποσύνολο θα μας απασχολήσει στη συγκεκριμένη μελέτη, καθώς είναι και αυτό που θα χρησιμοποιηθεί για την αξιολόγηση. 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).



Εικ.4. Απεικόνιση υποσυνόλου της τυπολογίας λαθών MQM-DQF που αφορά τη MM με τις κατηγορίες και τις υποκατηγορίες τους (Πηγή: <https://www.aclweb.org/anthology/W18-1925.pdf>)

6. Inter-annotator agreement

Καθώς επαναλαμβάνεται η διαδικασία της αξιολόγησης της ποιότητας των μεταφράσεων της MM, υπάρχει αρκετή απόκλιση ανάμεσα στις αξιολογήσεις των επισημειωτών (Przybowski et al., 2008). Αυτό συμβαίνει λόγω της υποκειμενικότητας των ανθρώπων κατά τη διαδικασία. Όταν υπάρχουν πολλαπλοί επισημειωτές, οι οποίοι δίνουν διαφορετικά αποτελέσματα, τότε υπολογίζουμε την απόκλιση τους με τη διαδικασία του «inter-annotator agreement», ενώ όταν γίνονται ετεροχρονισμένες αξιολογήσεις, για να υπολογιστεί η απόκλιση σε αυτά χρησιμοποιούμε τη 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. Όταν τα δεδομένα είναι αριθμητικά, χρησιμοποιούμε την συσχέτιση του Pearson (Pearson's correlation). Για τα κατηγοριοποιημένα δεδομένα που βρίσκονται σε ονομαστική κλίμακα (nominal scale) χρησιμοποιούμε το Cohen's Kappa όταν τα αποτελέσματα προέρχονται από δυο επισημειωτές (Cohen, 1960) ή το Fleiss' Kappa, για να υπολογιστεί η συμφωνία περισσότερων από δυο επισημειωτών (Fleiss, 1969) (Fleiss, 1971).

7. Αυτόματη Αξιολόγηση της ποιότητας της Μηχανικής Μετάφρασης

Όπως αναφέρθηκε και παραπάνω, για την αξιολόγηση της ποιότητας της μετάφρασης που προκύπτει από MM έχουν προταθεί και εφαρμοστεί διαφορετικές τεχνικές, έτσι ώστε να εξασφαλιστεί η ποιότητα, καθώς και η χρησιμότητα των συστημάτων κατά τη μεταφραστική διαδικασία. Η άλλη κατηγορία που θα αναλυθεί σε αυτό το κομμάτι είναι η αυτόματη αξιολόγηση. Η αυτόματη αξιολόγηση γίνεται με αυτόματες μετρικές. Οι αυτόματες μετρικές δημιουργήθηκαν για να μειωθεί το κόστος, ο όγκος της εργασίας που απαιτεί η ανθρώπινη αξιολόγηση ώστε να είναι ευκολότερο να επαναλαμβάνεται συχνότερα η διαδικασία, 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, οι οποίες αναφέρονται ως υποψήφιος ή υποθέσεις, 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). Τα συστήματα αυτόματης αξιολόγησης πρέπει να τηρούν κάποιες προϋποθέσεις, έτσι ώστε να είναι λειτουργικά και αυτές είναι: (Banerjee and Lavie, 2005)

- 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.

Το μεγαλύτερο μέρος των αυτόματων μετρικών χρησιμοποιεί για την αξιολόγηση της ποιότητας μια 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 αναζητούμε τον ελάχιστο αριθμό 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 έτσι ώστε από το αποτέλεσμα της MM να φτάσουμε στη μετάφραση-αναφοράς (Fordyce, 2007).

Στην αυτόματη αξιολόγηση χρησιμοποιούνται, κυρίως, δυο μετρικές η *ακρίβεια* (*precision*) και η *ανάκληση* (*recall*). Αυτές οι δυο μετρικές χρησιμοποιούνται για την ανάκτηση πληροφοριών, αλλά εφαρμόζονται και στην αυτόματη αξιολόγηση της μηχανικής μετάφρασης (Lavie, 2011). Η ακρίβεια υπολογίζεται από τα “σωστά” n-grams⁴ 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). Με λίγα λόγια, στη πρώτη περίπτωση υπολογίζουμε το ποσοστό των παραχθειςών λέξεων που είναι σωστές και στην άλλη πόσες από τις επιθυμητές λέξεις παρήχθησαν.

Μερικές από τις πιο γνωστές αυτόματες μετρικές είναι το Word Error Rate (WER), η οποία είναι η πιο συνηθισμένη που χρησιμοποιείται για την αξιολόγηση της ποιότητας της αυτόματης αναγνώρισης ομιλίας και μια από τις πρώτες που χρησιμοποιήθηκαν στην αυτόματη μετάφραση. Για τον υπολογισμό του συγκρίνει το αποτέλεσμα της MM με την μετάφραση-αναφοράς, η οποία είναι μια μετάφραση που γίνεται από τον άνθρωπο. Οι περισσότερες μετρικές χρησιμοποιούν μετάφραση-αναφοράς, είτε μια, είτε περισσότερες ανάλογα τη μετρική. 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. Στις περιπτώσεις που μια λέξη της μετάφρασης-αναφοράς δεν αντιστοιχεί σε καμία λέξη, τότε έχουμε *διαγραφή* (*deletion*) και αντίστοιχα σε περίπτωση που μια λέξη του αποτελέσματος της MM δεν αντιστοιχεί σε καμία λέξη, τότε έχουμε *προσθήκη* (*insertion*). Όταν η μια λέξη από τη MM αντιστοιχεί στην ίδια λέξη στην μετάφραση-αναφοράς έχουμε *αντιστοιχία* (*match*) ενώ στην αντίθετη περίπτωση έχουμε *αντικατάσταση* (*substitution*). Μια εναλλακτική του WER είναι το MWER (Multi-Reference WER), το οποίο εξετάζει τη διαφορά της MM με κάθε μετάφραση-αναφοράς (Nießen et al., 2000). Ουσιαστικά, το MWER κάνει τους ίδιους υπολογισμούς με το WER ανάμεσα στο κείμενο από MM με την πιο κοντινή σε αυτό μετάφραση-αναφοράς, καθώς η κάθε μετάφραση-αναφοράς είναι ανεξάρτητη με την άλλη. Και οι δυο αυτοί τύποι του WER δεν είναι οι κατάλληλες μετρικές για την αξιολόγηση μεταφράσεων, καθώς δεν λαμβάνουν υπόψη τις διαφορές στη σειρά των λέξεων ανάμεσα σε διαφορετικές γλώσσες. Υπάρχει άλλος ένας τύπος του WER, ο οποίος ονομάζεται PER (Position-Independent Error Rate) και οι λέξεις από το ένα κείμενο αντιστοιχούν σε λέξεις του άλλου ανεξάρτητα από τη θέση που βρίσκονται. Αυτό, δεν θα μπορούσε να είναι καλύτερο από τους υπόλοιπους τύπους, διότι δεν είναι ικανό να ξεχωρίσει μια σωστή μετάφραση από μια άλλη με μπερδεμένες λέξεις (Snover et al., 2011).

Η μετρική BLEU (Bilingual Evaluation Understudy) είναι από τις πιο πολυχρησιμοποιημένες και υπολογίζει την ακρίβεια (*precision*), η οποία ουσιαστικά «μετράει» τη διαφορά των n-grams μεταξύ των αποτελεσμάτων της MM και των μεταφράσεων-αναφοράς, οι οποίες δημιουργούνται από τον άνθρωπο (Papineni et al., 2002). Το BLEU έχει ένα κοινό

⁴ N-grams εννοούμε μια ακολουθία n-μονάδων.

χαρακτηριστικό με το MWER, το οποίο είναι ότι μπορεί να αξιολογεί με 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. Η μετρική αυτή στοχεύει περισσότερο στην ακρίβεια, και μετρά κατά πόσο το μετάφρασμα 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 (Papineni 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 METEOR (Metric for Evaluation of Translation with Explicit ORdering) σχεδιάστηκε σύμφωνα με τις προαναφερθείσες προϋποθέσεις, έτσι ώστε να μπορέσει να καλύψει όλες τις αδυναμίες έχει το BLEU (Banerjee, 2005). Το METEOR είναι μια μετρική βασισμένη στην *ανάκληση*, εν αντιθέσει με το BLEU που βασίζεται στην *ακρίβεια*, και η διαφορά του είναι ότι υπολογίζει και συνδυάζει και τις δυο. Η *ακρίβεια* στο METEOR υπολογίζεται διαιρώντας τον αριθμό των αντιστοιχιών με τον αριθμό των λέξεων του αποτελέσματος της MM, ενώ η *ανάκληση* διαιρώντας το σύνολο των αντιστοιχιών με το σύνολο των λέξεων της μετάφρασης-αναφοράς. Η συγκεκριμένη μετρική χρησιμοποιεί διάφορα στάδια για την αντιστοίχιση των λέξεων του αποτελέσματος της MM και της μετάφρασης-αναφοράς, έτσι ώστε να αντιστοιχηθούν οι προτάσεις μεταξύ τους. Αυτά είναι η *ακριβή αντιστοιχία (exact matching)*, όταν η πρόταση της MM είναι ίδια με τη μετάφραση-αναφοράς, *αντιστοιχία προέλευσης (stem matching)*, ώστε να υπάρχει αντιστοιχία λέξεων με την ίδια μορφολογική ρίζα και η *αντιστοιχία συνωνύμων (synonymy matching)*, όταν να αντιστοιχίζονται οι συνώνυμες λέξεις.⁵ Λέξεις, οι οποίες έχουν ήδη αντιστοιχηθεί σε προηγούμενο στάδιο δεν αντιστοιχίζονται ξανά. Σημαντικό είναι να τονιστεί και εδώ ότι η συγκεκριμένη μετρική λειτουργεί μόνο στο επίπεδο λέξης (Banerjee, 2005).

Η συγκεκριμένη μετρική μπορούμε να πούμε ότι ενέχει κάποιες γλωσσολογικές πληροφορίες μέχρι ένα επίπεδο, καθώς κατά την *αντιστοιχία προέλευσης (stem matching)* και την *αντιστοιχία συνωνύμων (synonymy matching)* αναγνωρίζει ισοδυναμίες ανάμεσα στη αποτέλεσμα της MM και τη μετάφραση-αναφοράς, οι οποίες δεν θα γίνονταν αντιληπτές από

⁵ Η μετάφραση των όρων για τα στάδια της αξιολόγησης είναι ελεύθερη μετάφραση του συντάκτη της παρούσας εργασίας.

άλλες μετρικές. Μια αδυναμία της συγκεκριμένης μετρικής είναι ότι δεν μπορεί να λαμβάνει πληροφορίες από όλες τις μεταφράσεις-αναφορές που χρησιμοποιεί αλλά αντιστοιχεί κάθε πρόταση του αποτελέσματος της MM με την μετάφραση-αναφοράς που δίνει τη καλύτερη βαθμολογία. Αυτό κάνει το METEOR να μην εκμεταλλεύεται αρκετά το πλεονέκτημά του να χρησιμοποιεί πολλές μεταφράσεις-αναφορές. Επιπλέον, άλλο ένα μειονέκτημα αυτής της μετρικής είναι ότι δεν “αντιλαμβάνεται” την διαφορετική σειρά των λέξεων που είναι πολύ πιθανό να συμβεί σε μια γλώσσα, στην οποία μια φράση ή πρόταση μπορεί να γραφεί με πολλούς διαφορετικούς και εξίσου σωστούς τρόπους (Snover et al., 2011).

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 αντιμετωπίζουν οι περισσότερες αυτόματες μετρικές, οι οποίες δεν μπορούν να εντοπίσουν τις διαφοροποιήσεις αυτές. Μια μετρική που σχεδιάστηκε έτσι ώστε 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*) ώστε να υπολογίζεται η ποιότητα της μετάφρασης. Ουσιαστικά εντοπίζει τις μεγαλύτερες ακολουθίες λέξεων, στις οποίες υπάρχει αντιστοιχία στο κείμενο από τη MM και τη μετάφραση-αναφοράς. Η ακρίβεια υπολογίζεται διαιρώντας το μέγιστο αριθμό αντιστοιχιών με το μέγεθος του αποτελέσματος της MM, ενώ η ανάκληση με το μέγεθος της μετάφρασης-αναφοράς. Το GTM μπορεί να διαχειριστεί πολλαπλές μεταφράσεις-αναφορές, όμως δεν περιέχει γλωσσολογική πληροφορία, και έτσι αντιστοιχεί τις λέξεις, οι οποίες είναι πανομοιότυπες. Επιπλέον, υπολογίζει δίνοντας βαρύτητα σε μία λανθασμένη λέξη ανάλογα με το σημείο που βρίσκεται στην εκάστοτε πρόταση, καθώς τα λάθη που βρίσκονται στη μέση μιας πρότασης θεωρούνται πιο σοβαρά από αυτά που βρίσκονται στην αρχή ή στο τέλος αυτής (Snover et al., 2011).

Το TER (Translation Error Rate) είναι, επίσης, μια μετρική που διευθετεί 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. Αυτοί οι περιορισμοί μιμούνται, ουσιαστικά, τον τρόπο, με τον οποίο σκέφτεται ένας επιμελητής να επιλέξει τις λέξεις που πρέπει να μετακινηθούν. Συγκεκριμένα αυτοί είναι τέσσερις. Πρώτον, 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. Δεύτερον, η ακολουθία των λέξεων του αποτελέσματος της MM πρέπει να μετακινηθεί έτσι ώστε να αντιστοιχίζεται ακριβώς στην ακολουθία των λέξεων της μετάφρασης-αναφοράς. Τρίτον, οι λέξεις που πρόκειται να μετακινηθούν πρέπει να περιέχουν τουλάχιστον ένα σφάλμα, σύμφωνα με το WER πριν μετακινηθεί, έτσι ώστε να αποφεύγεται η μετακίνηση των ήδη σωστών λέξεων. Και τέταρτον, 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).

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 έτσι ώστε 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, μα 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).

8. Μεθοδολογία της Έρευνας

Η παρούσα έρευνα επικεντρώνεται στη σύγκριση δυο νευρωνικών συστημάτων MM και την αξιολόγησή τους με ανθρώπινες μετρικές, ώστε να διαπιστωθεί ποιο από τα δυο συστήματα παράγει τα καλύτερα αποτελέσματα για κάθε κειμενικό είδος που εξετάζεται και ποιες κατηγορίες λαθών παρουσιάζει το κάθε ένα ξεχωριστά σε κάθε είδος. Τα δυο αυτά συστήματα που χρησιμοποιήθηκαν για να μεταφράσουν τα δυο κείμενα είναι γενικής χρήσης συστήματα (generic) και συγκεκριμένα το Google translate (GNMT, Google Neural Machine Translation), το οποίο πλέον ανανέωσε το δίκτυό του σε νευρωνικό και το PureNMT. Και τα δυο συστήματα είναι ανοιχτού κώδικα (open-source), ελεύθερα για όλους τους χρήστες του διαδικτύου και υποστηρίζουν την ελληνική γλώσσα, η οποία είναι η γλώσσα στην οποία εξετάζεται η απόδοση των συστημάτων. Το ζεύγος γλωσσών είναι Αγγλικά-Ελληνικά, με το πρωτότυπο να είναι στην αγγλική. Αυτό που διαχωρίζει τα δυο συστήματα και αφορά, κυρίως, την απόδοσή τους, μεταξύ άλλων, είναι τα δεδομένα, με τα οποία εκπαιδεύεται το κάθε νευρωνικό σύστημα για κάθε γλώσσα. Τα δεδομένα εκπαίδευσης των νευρωνικών συστημάτων τύπου γενικής χρήσης (generic), που παρέχονται ελεύθερα στους χρήστες του διαδικτύου, συλλέγονται από τα διαθέσιμα παράλληλα σώματα κειμένου, όμως ο πάροχος του εκάστοτε συστήματος ακολουθεί τη δική του στρατηγική συλλογής δεδομένων για την εκπαίδευση. Το σύστημα MM PureNMT έχει εκπαιδευτεί με παράλληλα σώματα κειμένου ανοιχτού κώδικα, ελεύθερα στο διαδίκτυο. Τα κείμενα που χρησιμοποιούνται, συνήθως, είναι νομικά κείμενα από τις ιστοσελίδες της «Europarl», «JRC» ή της «MultiUN», ομιλίες συνεδρίων «TED» και είναι επιστημονικές παρουσιάσεις και υπότιτλοι, που ανήκουν σε πιο

απλό είδος κειμένου και είναι στην καθομιλουμένη. Το συγκεκριμένο, λοιπόν, σύστημα έχει χρησιμοποιήσει μια μίξη όλων στο υλικό για εκπαίδευση, δίνοντας μεγαλύτερο βάρος σε κάποιες κατηγορίες κειμένων ανάλογα την διαθεσιμότητα σε υλικό που έχει η κάθε γλώσσα (για γλώσσες με μεγάλο όγκο δεδομένων μειώθηκε η χρήση του υλικού νομικών κειμένων και χρησιμοποιήθηκαν κείμενα ιστοσελίδων ενημέρωσης) και αυτό χωρίστηκε σε τρία μέρη για κάθε γλωσσικό συνδυασμό. Αυτά είναι ένα βασικό σώμα κειμένου (baseline corpus) με ένα εκατομμύριο προτάσεις, ένα μεσαίου μεγέθους σώμα κειμένου (medium corpus) δυόμιση εκατομμυρίων προτάσεων και ένα μεγάλο (very large corpora) με περισσότερες από 10 εκατομμύρια προτάσεις. Η εκπαίδευση έγινε σταδιακά για να εξεταστεί με ελέγχους η απόδοση του συστήματος σε κάθε γλώσσα (Crego et al., 2016). Το σύστημα αυτόματης μετάφρασης της Google, το GNMT έχει εκπαιδευτεί με προτάσεις και έγγραφα και τη μετάφραση τους, τα οποία βρίσκονται στο διαδίκτυο διαθέσιμα για το ευρύ κοινό. Το σύστημα, που συλλέγει τα δεδομένα από το διαδίκτυο, αναζητά υψηλής ποιότητας δεδομένα, για να εκπαιδευτεί το σύστημα όσο το δυνατόν καλύτερα. Τα δεδομένα που συλλέγονται, περνούν από μια διαδικασία “καθαρισμού” και το σύστημα αφού εκπαιδευτεί με όλα τα δεδομένα, εκπαιδεύεται ξανά με μικρότερες και πιο “καθαρές” υποενότητες των δεδομένων αυτών. Άλλη μια μέθοδος, που χρησιμοποιείται από το σύστημα αυτό είναι να γίνεται μετάφραση ήδη μεταφρασμένων κειμένων χωρίς πρόσβαση στο πρωτότυπο (back-translation), κυρίως για γλώσσες με λίγους πόρους δεδομένων, που δεν υπάρχουν πολλά παράλληλα σώματα κειμένων. Τα υπάρχοντα παραλληλοποιημένα ανά πρόταση σώματα κειμένων συμπληρώνονται με ήδη υπάρχοντα κείμενα στη ΓΠ και η μετάφρασή τους γίνεται με διαφορετικό νευρωνικό σύστημα MM. Βασική, λοιπόν, διαφορά των δυο είναι ο τρόπος εκπαίδευσης του κάθε συστήματος. Επιπλέον, ο τύπος δικτύου του GNMT και του PureNMT είναι ο ίδιος, όμως υπάρχουν διαφορές στον τρόπο, με τον οποίο γίνεται η αποκωδικοποίηση και η κωδικοποίηση ξανά στην άλλη γλώσσα. Για παράδειγμα, το PureNMT περιλαμβάνει κάποια γλωσσικά χαρακτηριστικά (word features) κατά τη διαδικασία της κωδικοποίησης, κι άλλες τέτοιες μικρές διαφορές κατά την λειτουργία, που επηρεάζουν την απόδοση του συστήματος (Crego, et al., 2016). Μ’ αυτόν τον τρόπο, βελτιώνεται η ευχέρεια των αποτελεσμάτων του συστήματος σε γλώσσες με λιγότερους γλωσσικούς πόρους και το σύστημα εκπαιδεύεται σωστότερα με τον μεγαλύτερο όγκο δεδομένων, που προκύπτει (Caswell and Liang, 2020).

Τα κείμενα που χρησιμοποιήθηκαν για την μελέτη είναι δυο διαφορετικών ειδών, με διαφορετικές ιδιαιτερότητες το καθένα. Το πρώτο κείμενο (Κείμενο Νο1) είναι ένα απόσπασμα νομικού περιγραφικού κειμένου, ουσιαστικά μιας ανασκόπησης των πειθαρχικών διαδικασιών των ορκωτών λογιστών, από το Συμβούλιο Χρηματοοικονομικής Αναφοράς (Financial Reporting Council). Το συγκεκριμένο απόσπασμα αφορά το Πειθαρχικό Συμβούλιο των Ελεγκτών Εταιρειών (Companies Auditors Disciplinary Board) και τις δικαιοδοσίες του. Η γλώσσα στο συγκεκριμένο κείμενο είναι περιγραφική, άρα σχετικά απλή και κατανοητή, που όμως περιέχει αρκετή ορολογία. Το ύφος του κειμένου είναι σοβαρό και αυστηρό, όπως συνηθίζεται σε νομικά κείμενα, καθώς έχει συνταχθεί από νομικό όργανο της αυστραλιανής κυβέρνησης. Το συγκεκριμένο απευθύνεται σε κοινό με τουλάχιστον στοιχειώδεις νομικές γνώσεις, συνήθως νομικά όργανα, μπορεί όμως να χρησιμοποιηθεί και από ερευνητές. Το δεύτερο κείμενο (Κείμενο Νο2) ανήκει στο ιατρικό είδος κειμένου και είναι και αυτό ενημερωτικού χαρακτήρα. Το συγκεκριμένο αποτελεί άρθρο στην ιστοσελίδα των Κέντρων Ελέγχου και Πρόληψης Νοσημάτων (Centers for Disease Control and Prevention

- CDC). Το θέμα που πραγματεύεται είναι η ασθένεια του ύπνου που προκαλείται από το τσίμπημα της μύγας τσε-τσε και η γλώσσα στο άρθρο είναι απλή, καθώς είναι και αυτό περιγραφικό, αλλά η ορολογία που υπάρχει είναι μέτριας δυσκολίας. Κάποιοι όροι επεξηγούνται και γίνονται κατανοητοί από τα συμφραζόμενα, όμως υπάρχουν και όροι, όπως θα δούμε παρακάτω, που αποδίδονται δύσκολα στα ελληνικά και είναι δυσνόητοι για το ευρύ κοινό. Το ύφος του κειμένου είναι απλό και κατανοητό, ενημερωτικό και ουδέτερο, σύμφωνα με το στόχο του πρωτοτύπου. Αυτό το κείμενο απευθύνεται τόσο στο ευρύ κοινό, όσο και στο ειδικό, γιατί παρόλο που η ορολογία μπορεί να μην γίνεται κατανοητή από όλους, το νόημα του κειμένου περνά στον απλό αναγνώστη και εξειδικευμένοι αναγνώστες σε ιατρικά ζητήματα μπορούν να ενημερωθούν βαθύτερα για την ασθένεια αντιλαμβανόμενοι την ορολογία και να το μελετήσουν σε ένα πρωταρχικό στάδιο.

Είναι σημαντικό να γίνουν κατανοητά τα είδη των δυο κειμένων και να διασαφηνιστεί, ποιο είναι το επιθυμητό αποτέλεσμα και ποιος είναι ο στόχος της μετάφρασης. Το επικοινωνιακό πλαίσιο, λοιπόν, κατά το οποίο διεξάγεται η μελέτη αφορά την απλή μετάφραση των κειμένων, διατηρώντας το ύφος, έτσι ώστε να είναι κατανοητό για τον Έλληνα αναγνώστη. Το κοινό-στόχος για το κάθε κείμενο είναι το ίδιο με αυτό του πρωτοτύπου, όπως και όλα τα γλωσσικά χαρακτηριστικά, με τη μόνη διαφορά ότι το κείμενο θα πρέπει να ενημερώσει το ελληνικό κοινό. Η επιλογή των συγκεκριμένων κειμένων έγινε γιατί σκοπός της έρευνας είναι να εξεταστεί πως αποδίδουν δυο προσβάσιμα από όλους συστήματα MM σε δυο διαφορετικούς τομείς, με δυο κείμενα, τα οποία, από τη μια, είναι σχετικά απλά, αλλά από την άλλη έχουν συγκεκριμένες ιδιαιτερότητες όσον αφορά το ύφος, τη γλώσσα και την ορολογία.

Η αξιολόγηση των αποτελεσμάτων της MM στα Κείμενα No1 και No2, διεξήχθη με ένα συνδυασμό μεθόδων, έτσι ώστε να διασφαλισθεί η ποιότητα, και όσο είναι δυνατόν η αντικειμενικότητα κατά τη διαδικασία. Για να αξιολογηθεί η ποιότητα χρησιμοποιήθηκε η ανθρώπινη αξιολόγηση των συστημάτων MM και συγκεκριμένα συνδυασμός ολιστικών και αναλυτικών μεθόδων αξιολόγησης. Η αξιολόγηση γίνεται σε επίπεδο πρότασης για κάθε κείμενο. Δηλαδή, για κάθε πρόταση του ΚΠ δίνεται η μετάφραση της αντίστοιχης πρότασης που προκύπτει από κάθε σύστημα. Η ολιστική μέθοδος περιλαμβάνει την κατάταξη των αποτελεσμάτων από τα δυο συστήματα MM από το καλύτερο προς το χειρότερο σύμφωνα με την εξής οδηγία και στις δυο περιπτώσεις: «Σύμφωνα με τα μεταφράσματα που σας δίνονται, τα οποία προέρχονται από δυο διαφορετικά νευρωνικά συστήματα μηχανικής μετάφρασης (NMT), καλείστε να τα βαθμολογήσετε/αξιολογήσετε από το καλύτερο προς το χειρότερο.» Μετά την ολοκλήρωση αυτής της διαδικασίας για κάθε πρόταση του ΚΠ, ακολουθεί η αξιολόγηση της ακρίβειας (accuracy) για κάθε πρόταση, αντιπαραβάλλοντας τα αποτελέσματα της MM και βαθμολογώντας το κάθε αποτέλεσμα της MM σε μια κλίμακα Likert 5 βαθμίδων ακολουθώντας την οδηγία: «Σε αυτήν την ενότητα καλείστε να αξιολογήσετε σε κλίμακα Likert 5 βαθμίδων την ΑΚΡΙΒΕΙΑ της κάθε πρότασης που έχει προκύψει από μηχανική μετάφραση, απαντώντας στην ερώτηση: "Έχει διατηρηθεί το νόημα της αγγλικής πρότασης και στο μετάφρασμα;"». Εδώ οι βαθμίδες αυτές είναι: 1-Καθόλου, 2-Λίγο, 3-Αρκετά, 4-Πολύ, 5-Πάρα πολύ.

Στη συνέχεια, η ίδια διαδικασία ακολουθείται για την ευχέρεια (fluency), με την οδηγία να γίνεται: «Σε αυτήν την ενότητα καλείστε να αξιολογήσετε σε κλίμακα Likert 5 βαθμίδων την ΕΥΧΕΡΕΙΑ της κάθε πρότασης που έχει προκύψει από μηχανική μετάφραση, απαντώντας στην

ερώτηση: "Σύμφωνα μόνο με τους γραμματικούς και συντακτικούς κανόνες η μεταφρασμένη πρόταση είναι:"», και οι βαθμίδες έχουν τη μορφή: 1-Πολύ κακή, 2-Κακή, 3-Καλή, 4-Πολύ καλή, 5-Εξαιρετική. Σε όλες τις παραπάνω διαδικασίες δίνονται και οι προτάσεις του πρωτοτύπου.

Περνώντας στην αναλυτική αξιολόγηση, οι επισημειωτές καλούνται να κατηγοριοποιήσουν τα λάθη που εντοπίζουν σε κάθε αποτέλεσμα της MM. Και εδώ η αξιολόγηση γίνεται σε επίπεδο πρότασης, και βλέποντας τις προτάσεις του πρωτοτύπου, εντοπίζουν τα λάθη που έχει κάθε πρόταση κάθε συστήματος, πλέον όχι αντιπαραβολικά, αλλά για κάθε σύστημα ξεχωριστά. Η εντολή που δόθηκε σε αυτό το στάδιο είναι: «Σε αυτή την ενότητα καλείστε να εντοπίσετε και να κατηγοριοποιήσετε τα λάθη της κάθε πρότασης που προκύπτει από μηχανική μετάφραση σύμφωνα με τις παρακάτω κατηγορίες λαθών.» Οι κατηγορίες λαθών δόθηκαν στους επισημειωτές ως επιλογές και πηγάζουν από το υποσύνολο των Lommel και Melby (2018) από το εναρμονισμένο πρότυπο MQM- DQF, καθώς αυτή η κατηγοριοποίηση αφορά την αξιολόγηση των αποτελεσμάτων της MM, όπως αναλύθηκε και παραπάνω. Από τις κατηγορίες του υποσυνόλου, επιλέχθηκαν οι πιο σημαντικές και αυτές που μας δίνουν πληροφορίες για τα συγκεκριμένα είδη κειμένου. Συγκεκριμένα, οι κατηγορίες είναι: «Προσθήκες», «Παραλήψεις», «Παρανοήσεις», «Αμετάφραστο κείμενο», «Λάθη στην ορολογία», «Γραμματικά Λάθη», «Λάθη στη στίξη», «Λάθη στο ύφος», «Ορθογραφικά Λάθη», «Τυπογραφικά Λάθη». Στην κατηγορία των γραμματικών λαθών περιλαμβάνονται τα λάθη σε μέρη του λόγου, στη σειρά των λέξεων και στις λειτουργικές λέξεις. Η επιλογή των συγκεκριμένων έγινε ώστε να είναι βαθιά η διαδικασία για τους επισημειωτές, αλλά και για να καλυφθούν οι αδυναμίες όλων των προτάσεων των κειμένων μέσα σε αυτές, δίνοντας πάντα και την επιλογή «Άλλο», σε περίπτωση που κάποια πρόταση περιέχει κάποιο άλλο λάθος που δεν καλύπτεται από τις συγκεκριμένες κατηγορίες. Επίσης, τους δόθηκε και η επιλογή «Δεν υπάρχει λάθος» σε περίπτωση που κάποια πρόταση είναι σωστή. Σύμφωνα με συναφείς έρευνες για την ελληνική γλώσσα (Sosoni and Stasimioti, 2019), (Stasimioti et al., 2020) αυτές είναι οι συνήθεις κατηγορίες που εντοπίζονται για τα ελληνικά στα αποτελέσματα της MM.

Οι επισημειωτές που κλήθηκαν να αξιολογήσουν τα αποτελέσματα της MM είναι επαγγελματίες μεταφραστές, τελειόφοιτοι φοιτητές του ΠΜΣ «Επιστήμη της Μετάφρασης» του Ιονίου Πανεπιστημίου, με γνώσεις πάνω στη MM, την μετεπεξεργασία των αποτελεσμάτων MM και την ανθρώπινη αξιολόγηση της MM. 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 δυο επισημειωτές έλαβαν μέρος, έτσι ώστε τα αποτελέσματα να είναι όσο το δυνατόν περισσότερο αντικειμενικά. 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. Εφόσον οι επισημειωτές είναι δυο, ο υπολογισμός θα γίνει με το Cohen's Kappa, του οποίου οι βαθμολογίες διαφέρουν από το 0 μέχρι το 1 με την εξής κατηγοριοποίηση: (Glen, 2014)

0 = καμία συμφωνία (agreement equivalent to chance).

0,1 – 0,20 = ελάχιστη συμφωνία (slight agreement).

0,21 – 0,40 = αρκετή συμφωνία (fair agreement).

0,41 – 0,60 = μέτρια συμφωνία (moderate agreement).

0,61 – 0,80 = καλή συμφωνία (substantial agreement).

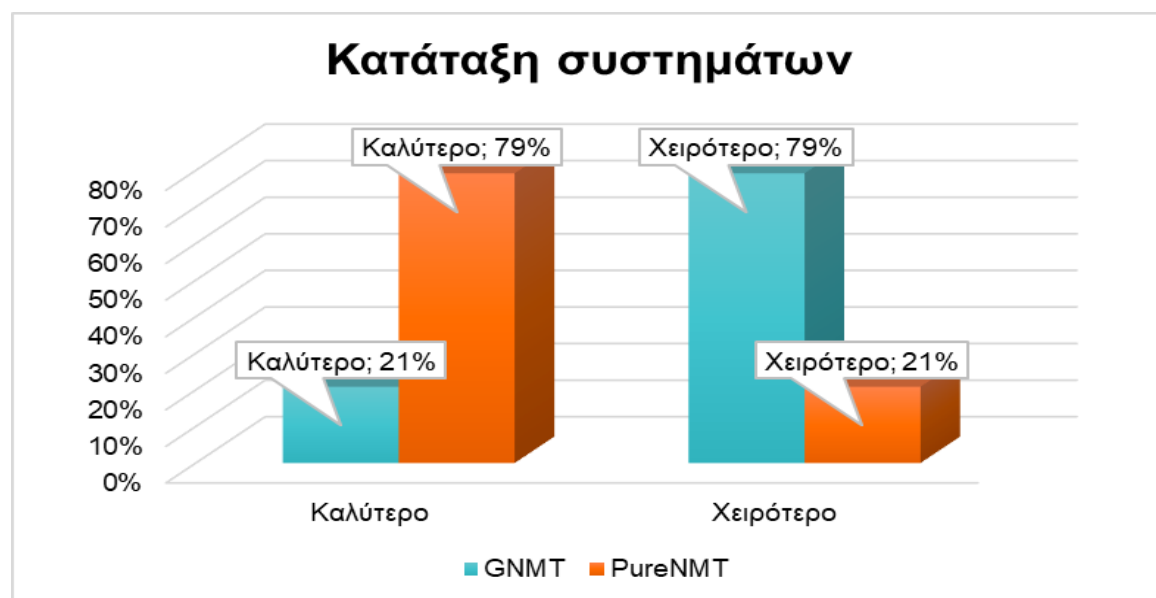
0,81 – 0,99 = πολύ καλή συμφωνία (near perfect agreement).

1 = εξαιρετική συμφωνία (perfect agreement).

9. Αποτελέσματα της έρευνας

Νομικό κείμενο.

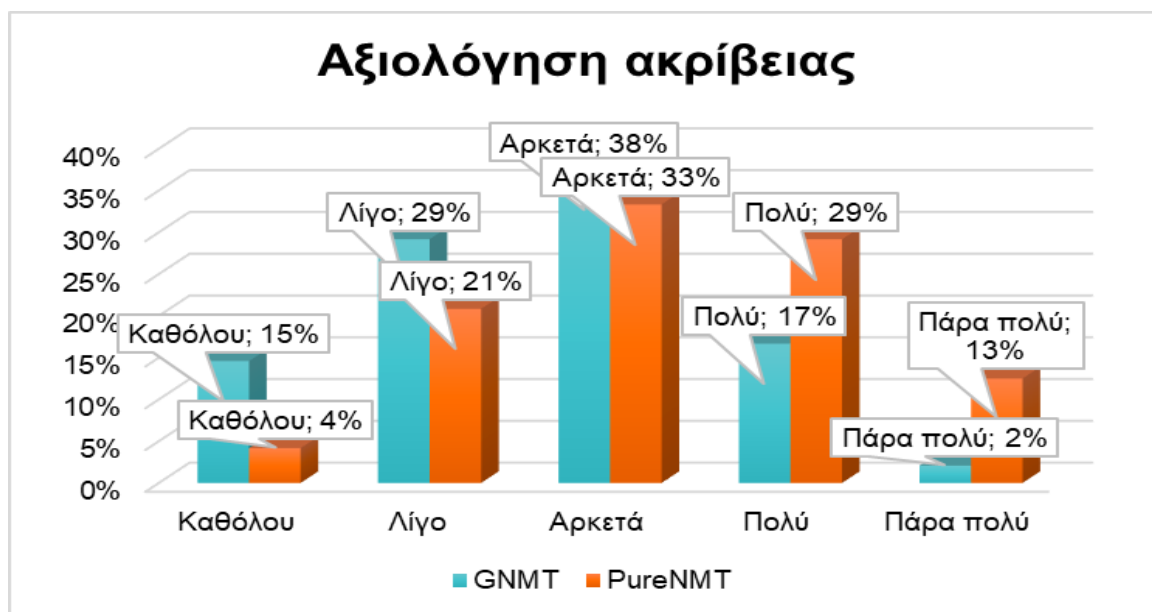
Σύμφωνα με τους δυο επισημειωτές, στο πρώτο στάδιο της κατάταξης των μεταφρασμένων προτάσεων από τα δυο συστήματα MM, GNMT και PureNMT, η καλύτερη απόδοση ήταν αυτή του συστήματος PureNMT. Συγκεκριμένα, συγκρίνοντας τις μεταφρασμένες προτάσεις του αγγλικού κειμένου, οι επισημειωτές έκριναν ότι το 79% των προτάσεων, που προέκυψαν από το PureNMT ήταν καλύτερες από αυτές του GNMT. Πιο αναλυτικά, το 79% των απαντήσεων και των δυο επισημειωτών, αξιολογούσαν το PureNMT ως καλύτερο, ενώ μόνο το 21% ήταν υπέρ του GNMT. Όπως φαίνεται και στο γράφημα 1, υπάρχει μεγάλη διαφορά ανάμεσα στην απόδοση των δύο συστημάτων, όταν αυτά αξιολογούνται συνολικά και αντιπαραβολικά το ένα μετάφρασμα με το άλλο. Η συμφωνία των δυο επισημειωτών υπολογίστηκε με την μέθοδο του Inter-Annotator Agreement (IAA), που εξηγήθηκε παραπάνω και για την κατάταξη, η συμφωνία υπολογίστηκε στο 0,83, αριθμός που δείχνει πολύ καλή συμφωνία (near perfect agreement) μεταξύ τους. Αυτό σημαίνει, ότι μπορούμε να θεωρήσουμε ότι τα δεδομένα βρίσκονται σε πολύ καλό επίπεδο αντικειμενικότητας.



Γράφημα 1. Ποσοστιαία κατάταξη από το καλύτερο προς το χειρότερο των προτάσεων, που έχουν προκύψει από δυο διαφορετικά συστήματα MM.

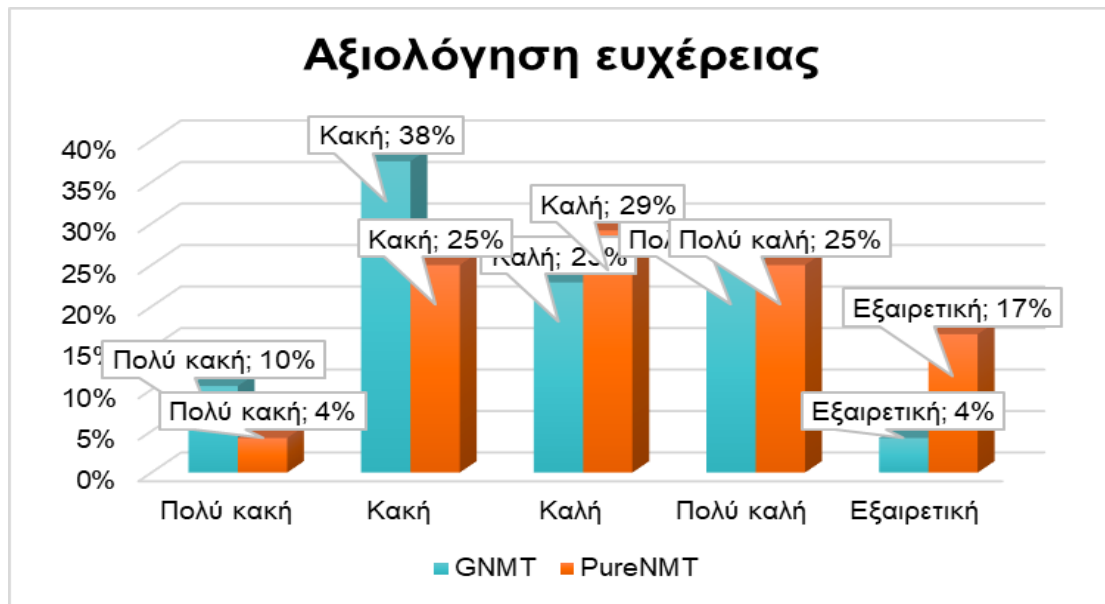
Στη συνέχεια, απαντώντας στην ερώτηση "Έχει διατηρηθεί το νόημα της αγγλικής πρότασης και στο μετάφρασμα;", οι απαντήσεις των επισημειωτών δείχνουν ότι το PureNMT έχει

μεταφέρει καλύτερα το νόημα στα ελληνικά, απ' ότi το GNMT. Για το PureNMT, το 75% των καταχωρήσεων δηλώνει, ότi το νόημα έχει μεταφερθεί «Αρκετά», «Πολύ», «Πάρα πολύ» στις επίπεδα της κλίμακας Likert. Συγκεκριμένα, το 33% των απαντήσεων αναφέρει, ότi έχει μεταφερθεί το νόημα «Αρκετά», το 29% «Πολύ» και το 13% «Πάρα πολύ». Αντιθέτως, το μεγαλύτερο ποσοστό (81%) του GNMT αποδίδεται στο «Καθόλου», «Λίγο» και «Αρκετά», με το 38% να αξιολογεί την απόδοση του νοήματος από το GNMT ως «Αρκετά». Τα ποσοστά φαίνονται αναλυτικά στο παρακάτω γράφημα. Εδώ, η συμφωνία των επισημειωτών, σύμφωνα με το IAA, βρίσκεται στο 0,45, αριθμός που δηλώνει μέτρια συμφωνία (moderate agreement), με κάποιες, όμως, αμφιβολίες για την αντικειμενικότητα των επισημειωτών κατά την αξιολόγηση.

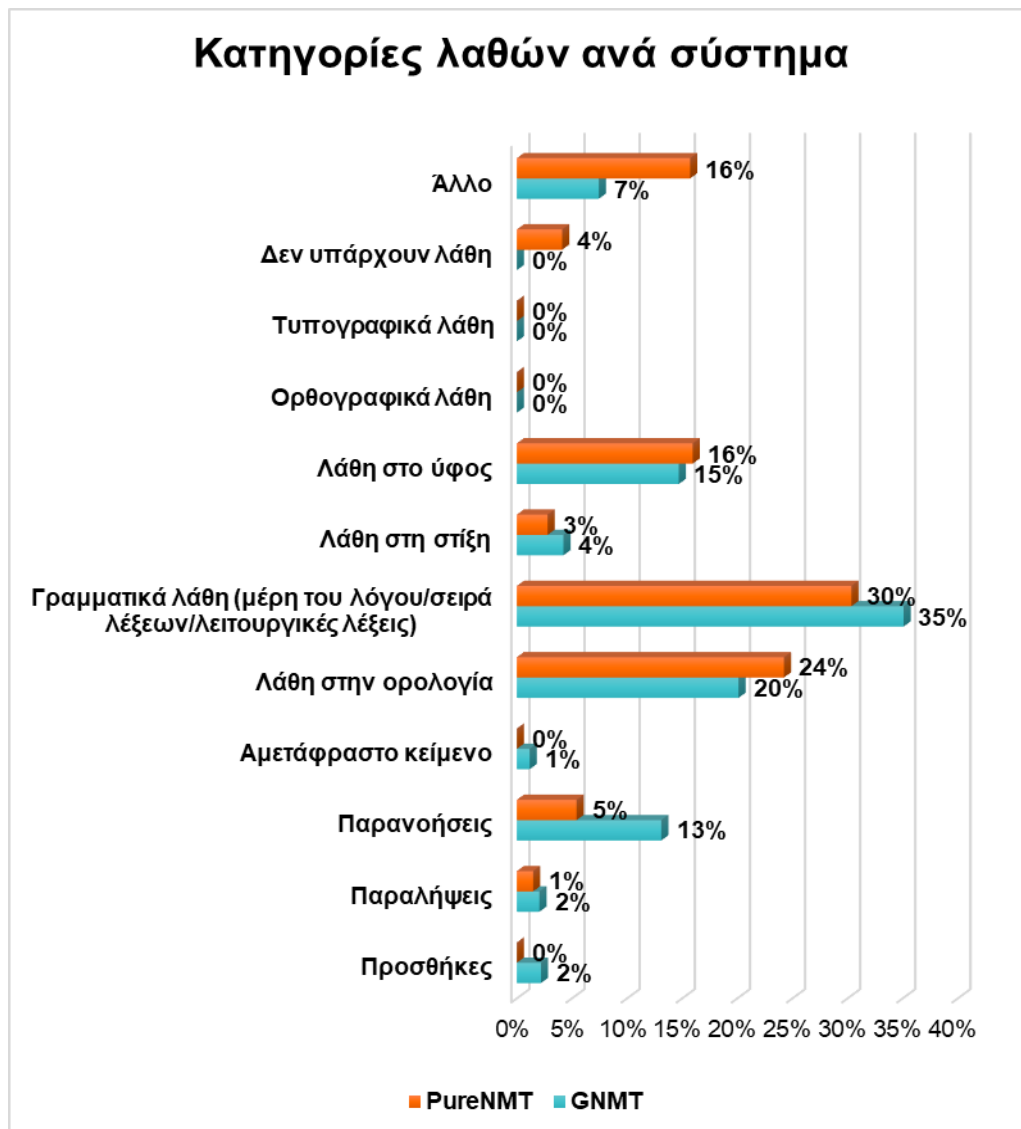


Γράφημα 2. Ποσοστιαία αξιολόγηση της ακρίβειας του νοήματος των προτάσεων, που έχουν προκύψει από δυο διαφορετικά συστήματα ΜΜ.

Στο επόμενο βήμα, οι επισημειωτές κλήθηκαν να αξιολογήσουν την ευχέρεια των δύο μεταφράσεων, απαντώντας στην ερώτηση «Σύμφωνα μόνο με τους γραμματικούς και συντακτικούς κανόνες η μεταφρασμένη πρόταση είναι:». Τα ποσοστά μας δείχνουν ότi η απόδοση του GNMT στην προκειμένη είναι κακή, καθώς ένα μεγάλο ποσοστό απαντήσεων (38%), μας δείχνει ότi η γραμματικότητα του αποτελέσματος αυτού του συστήματος είναι «Κακή». Το αμέσως μεγαλύτερο ποσοστό αποδίδεται στην κατηγορία «Καλή» και είναι το 25%. Ένα ελάχιστο ποσοστό κρίνει την ευχέρεια του GNMT ως «Πολύ Καλή», «Εξαιρετική» και αυτό είναι το 4%. Αντιθέτως, η γραμματικότητα της απόδοσης του PureNMT κρίνεται ως «Καλή» στο 29% των απαντήσεων και σε ένα ικανοποιητικό ποσοστό (17%) ως «Εξαιρετική». Άξιο αναφοράς είναι, όμως, και το ποσοστό που κρίνει την ευχέρεια ως «Κακή», καθώς αποτελεί ένα αρκετά υπολογίσιμο ποσοστό, στο 25%. Παρόμοιος αριθμός συμφωνίας με την αξιολόγηση της ακρίβειας υπάρχει και σ' αυτό το σημείο, με το IAA να είναι 0,42 [μέτρια συμφωνία (moderate agreement)], γεγονός που δημιουργεί ανασφάλεια για την αντικειμενικότητα των αποτελεσμάτων.



Γράφημα 3. Ποσοστιαία αξιολόγηση της ευχέρειας των προτάσεων, που έχουν προκύψει από δυο διαφορετικά συστήματα ΜΜ.



Γράφημα 4. Ποσοστιαία κατηγοριοποίηση των λαθών των προτάσεων, που έχουν προκύψει από δυο διαφορετικά συστήματα ΜΜ.

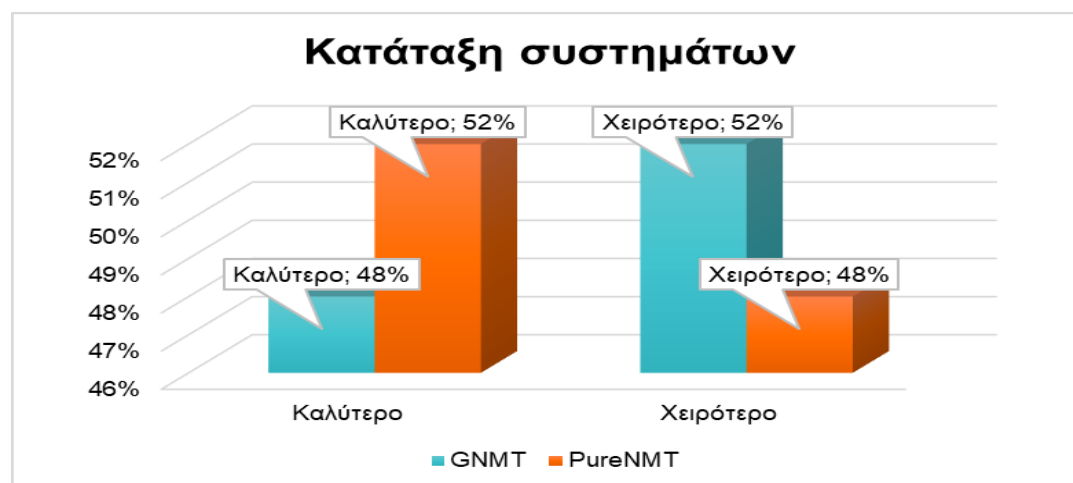
Τέλος, ακολουθεί η κατηγοριοποίηση των λαθών, κατά την οποία τα λάθη που παρατηρήθηκαν ήταν κυρίως λάθη ορολογίας, γραμματικά λάθη, λάθη στο ύφος και κάποιες παρανοήσεις. Τα περισσότερα λάθη και για τα δυο συστήματα ήταν στην κατηγορία των γραμματικών λαθών, που για το PureNMT είναι το 30% και για το GNMT το 35%. Η αμέσως επόμενη κατηγορία με τα περισσότερα λάθη και για τα δυο συστήματα είναι τα «Λάθη στην ορολογία» (το PureNMT με 24% ενώ το GNMT 20%) και μετά τα «Λάθη στο ύφος», με τα δυο συστήματα να έχουν 1% διαφορά στη συγκεκριμένη κατηγορία, ενώ το PureNMT είναι αυτό με το μεγαλύτερο ποσοστό. Επίσης, το GNMT παρουσιάζει αρκετές αδυναμίες στο κομμάτι των παρανοήσεων με το 13% των λαθών στις προτάσεις του να εμπίπτουν σ' αυτήν την κατηγορία. Στις υπόλοιπες κατηγορίες, τα ποσοστά των λαθών και των δυο συστημάτων είναι πολύ μικρά έως μηδενικά. Τα ποσοστά παρουσιάζονται και αναλυτικά στο γράφημα 4. Οι κατηγορίες που εντοπίζονται, τα λάθη και τα ποσοστά αυτών για κάθε σύστημα δικαιολογούν και τα αποτελέσματα των παραπάνω ενοτήτων, καθώς βλέπουμε, ότι κατά 3% το GNMT έχει παρουσιάσει περισσότερα λάθη από το PureNMT, οπότε γίνονται και πλέον αντιληπτά τα αποτελέσματα της κατάταξης των συστημάτων. Επίσης, οι κατηγορίες, στις οποίες εμφανίζονται τα περισσότερα λάθη για το PureNMT, είναι στις κατηγορίες που, σύμφωνα με το πρότυπο MQM-DQF, εμπίπτουν στην ευρύτερη κατηγορία της ευχέρειας. Αυτό δικαιολογεί την μικρότερη διαφορά ανάμεσα στα δυο συστήματα, όπως βλέπουμε στο κομμάτι της αξιολόγησης, όμως με το GNMT να έχει μεγαλύτερα ποσοστά στις κατηγορίες λαθών της ευχέρειας (γραμματικά λάθη, λάθη στη στίξη) από το PureNMT. Έτσι καταλαβαίνουμε, γιατί προέκυψαν και τα συγκεκριμένα αποτελέσματα. Επίσης, άλλο ένα κριτήριο που οδήγησε τους επισημειωτές να κρίνουν ότι το GNMT αποδίδει χειρότερα την ευχέρεια από το PureNMT είναι η βαρύτητα των λαθών στις συγκεκριμένες κατηγορίες, που γίνονται αντιληπτά, αν εξεταστούν τα επιμέρους λάθη ξεχωριστά. Μερικά παραδείγματα αυτών είναι, π.χ. η πρόταση του αγγλικού κειμένου «CADB jurisdiction is limited to matters referred to it by ASIC and APRA, and to matters involving RCAs.», της οποίας η μετάφραση από το GNMT είναι «Η αρμοδιότητα της CADB περιορίζεται στα θέματα που της υποβάλλονται από την ASIC και την APRA και σε θέματα που αφορούν RCA.» και από το PureNMT αντίστοιχα «Η δικαιοδοσία του CADB περιορίζεται σε θέματα που του παραπέμπονται από την ASIC και την APRA, καθώς και σε θέματα που αφορούν τις ΑΠΑ.». Γί αυτήν την πρόταση συμφώνησαν και οι δυο επισημειωτές, ότι το PureNMT δεν παρουσιάζει γραμματικά λάθη, εν αντιθέσει με το GNMT. Άλλο ένα παράδειγμα, για να ενισχυθεί η υπόθεση μας, είναι οι μεταφράσεις της αγγλικής πρότασης «CADB has received six applications from ASIC on RCA-related matters since 2010.» από το GNMT ως «Το CADB έλαβε από την ASIC έξι αιτήσεις από το ASIC σχετικά με θέματα που σχετίζονται με την RCA από το 2010.» και από το PureNMT ως «Το CADB έχει λάβει έξι αιτήσεις από το ASIC σχετικά με θέματα που αφορούν την RCA από το 2010.», για την οποία πρόταση υπάρχει η ίδια συμφωνία μεταξύ των επισημειωτών, όπως στην προηγούμενη. Αυτά τα ενδεικτικά παραδείγματα σε σχέση με τα ποσοστά, συμπληρώνουν την ολιστική μας έρευνα με μια αναλυτική και αιτιολογούν τα ολιστικά δεδομένα που προέκυψαν από την κατάταξη και την αξιολόγηση της ακρίβειας και της ευχέρειας.

Επιπλέον, τα δεύτερα μεγαλύτερα ποσοστά και των δυο συστημάτων στην κατηγορία λαθών της ορολογίας μας αποδεικνύουν, ότι και τα δυο συστήματα δεν απέδωσαν πλήρως το νόημα του αγγλικού κειμένου στη ΓΣ. Παραδείγματα που ενισχύουν αυτήν την υπόθεση είναι, ενδεικτικά, η αγγλική πρόταση «Members of CADB are appointed by the Treasurer and the ASIC Act requires members to have a breadth of law, accounting, and business expertise.», της οποίας η μετάφραση του GNMT είναι «Τα μέλη του CADB διορίζονται από τον Ταμία και ο νόμος ASIC απαιτεί από τα μέλη να έχουν ένα εύρος δικαίου, λογιστικής και επιχειρηματικής εμπειρογνωμοσύνης.» και η μετάφραση του PureNMT είναι «Τα μέλη του CADB διορίζονται από το Δημόσιο Ταμείο και ο νόμος ASIC απαιτεί από τα μέλη να έχουν ένα εύρος νομικής, λογιστικής και επιχειρηματικής εμπειρογνωμοσύνης.». Στο συγκεκριμένο παράδειγμα και στις δύο προτάσεις έχουν εντοπιστεί από τους επισημειωτές λάθη στην

ορολογία, πράγμα που διασαφηνίζει γιατί και τα δυο συστήματα είχαν τα μεγαλύτερα ποσοστά απαντήσεων στην κατηγορία «Αρκετά» της κλίμακας Likert, κατά την αξιολόγηση της ακρίβειας. Παρ' όλα αυτά, επειδή ένα μεγάλο ποσοστό του PureNMT αξιολογεί, την απόδοση της ακρίβειας ως «Καλή» και «Πολύ καλή», είναι αρκετά υψηλά και συγκρίσιμα με τα αντίστοιχα ποσοστά του GNMT στις συγκεκριμένες κατηγορίες, παρότι δεν είναι τα μεγαλύτερα. Αυτό δικαιολογείται, αν παρατηρήσει κανείς τα ποσοστά στις κατηγορίες λαθών και συγκεκριμένα στις κατηγορίες αυτές που εμπίπτουν στην ευρύτερη της ακρίβειας. Το GNMT έχει περισσότερα λάθη στις κατηγορίες «Προσθήκες», «Παραλείψεις» και ειδικά στην κατηγορία «Παρανοήσεις». Στην τελευταία, η διαφορά είναι 8%, μια διαφορά που είναι άξια αναφοράς. Ας πούμε, στην πρόταση «• to ensure that the proceedings at a hearing are conducted with as little formality and technicality, and with as much expedition, as the requirements of the Corporations legislation and a proper consideration of the matters before CADB permit; and», η οποία μεταφράστηκε από το GNMT ως «• να διασφαλιστεί ότι η διαδικασία σε ακρόαση διεξάγεται με ελάχιστη διατύπωση και τεχνική και με όσο το δυνατόν μεγαλύτερη αποστολή, όπως οι απαιτήσεις της νομοθεσίας των εταιρειών και η σωστή εξέταση των ζητημάτων πριν από το CADB. και» και από το PureNMT ως «• να διασφαλίσει ότι οι διαδικασίες σε μια ακρόαση διεξάγονται με ελάχιστη τυπικότητα και τεχνική λεπτομέρεια, και με την ίδια ταχύτητα όπως οι απαιτήσεις της νομοθεσίας των Εταιρειών και η σωστή εξέταση των θεμάτων πριν από την άδεια του CADB, και», σύμφωνα και με τους δυο επισημειωτές η μετάφραση του GNMT έχει παρανοήσεις, και σύμφωνα με τον έναν έχει και λάθη στην ορολογία. Στην κατηγοριοποίηση των λαθών, η οποία έγινε ξεχωριστά για κάθε σύστημα, η συμφωνία των επισημειωτών ήταν πολύ καλή συμφωνία (near perfect agreement) και για τα δυο συστήματα, με το IAA να υπολογίζεται στο 0,83 για το PureNMT και στο 0,82 για το GNMT.

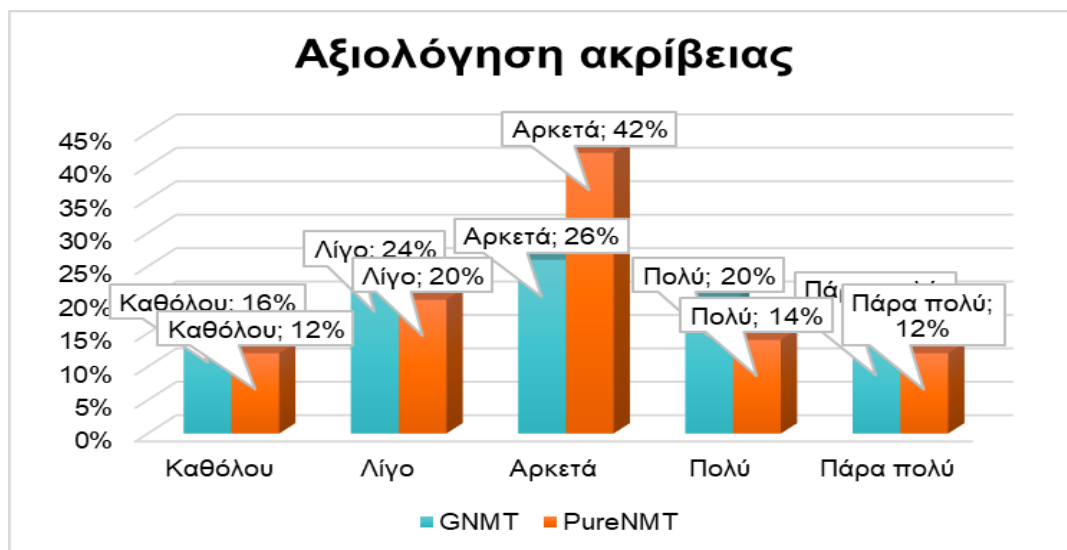
Ιατρικό κείμενο

Η ίδια διαδικασία αξιολόγησης έγινε και για το ιατρικό κείμενο, με το PureNMT να υπερτερεί του GNMT κι εδώ, με μικρότερη, όμως, διαφορά. Το 48% των απαντήσεων καταχωρούν το GNMT ως το καλύτερο ενώ 52% το PureNMT. Η διαφορά σε αυτό το είδος κειμένου είναι μόλις 4%, πράγμα που μας δείχνει ότι η απόδοση των δυο συστημάτων ήταν στα ίδια, κυρίως, επίπεδα με μια μικρή υπεροχή του PureNMT με το 52% των καταχωρήσεων. Το γράφημα 4 μας δείχνει τη συνολική εικόνα της κατάταξης των συστημάτων, αντιπαραβολικά, για το ιατρικό κείμενο. Το IAA, εδώ, δείχνει ότι υπάρχει μέτρια συμφωνία (moderate agreement) με 0,52. Είναι μια ικανοποιητική συμφωνία, όπως αναφέρθηκε και παραπάνω, που όμως εφιστά την προσοχή για μια πιο λεπτομερή εξέταση των δεδομένων.

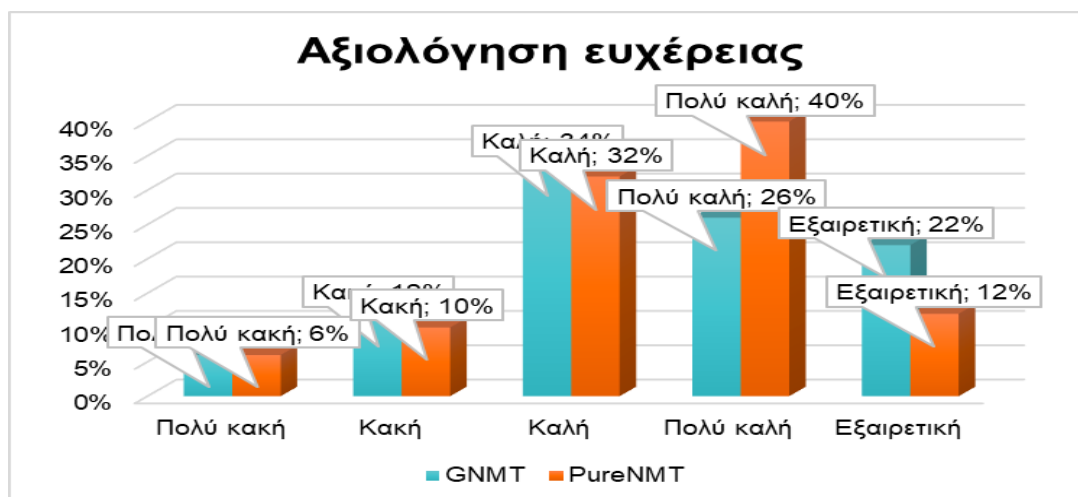


Γράφημα 5. Ποσοστιαία κατάταξη από το καλύτερο προς το χειρότερο των προτάσεων, που έχουν προκύψει από δυο διαφορετικά συστήματα MM.

Στην αξιολόγηση της μεταφοράς του νοήματος, το PureNMT έχει μεταφέρει το νόημα «Αρκετά» κατά 42% και το GNMT κατά 26%, ποσοστά που αποτελούν τα μεγαλύτερα και για τα δυο συστήματα. Το GNMT έχει μεγαλύτερα ποσοστά απαντήσεων (συνολικά 40%) στο «Λίγο» και το «Καθόλου», ενώ το PureNMT τα αμέσως επόμενα ποσοστά των απαντήσεων στο «Λίγο» και το «Πολύ». Ενδιαφέρουσα είναι και η ισοβαθμία των ποσοστών του PureNMT στο «Καθόλου» και το «Πάρα πολύ» με 12%, καθώς αυτά είναι τα δυο άκρα. Αυτό δείχνει, ότι ένας αριθμός προτάσεων αυτού του συστήματος δεν μετέφερε καθόλου το νόημα και ο ίδιος αριθμός μετέφερε όλο το νόημα της αγγλικής πρότασης. Συγκριτικά με τα δύο συστήματα, επίσης, παρατηρούμε ότι ένα 20% των προτάσεων του GNMT μετέφερε το νόημα «Πολύ» ενώ του PureNMT «Λίγο». Άρα, σε αυτό το σημείο μπορούμε να συμπεράνουμε ότι το νόημα έχει μεταφερθεί περίπου στα ίδια επίπεδα, αν κοιτάξουμε τα αποτελέσματα συνολικά, καθώς τα ποσοστά και των δυο είναι σχετικά ισόποσα μοιρασμένα σε όλες τις κατηγορίες, πέραν της κατηγορίας «Αρκετά» για το PureNMT. Στο γράφημα 5 παρουσιάζονται τα ποσοστά αναλυτικά. Για την αξιολόγηση της ακρίβειας του κείμενου, το IAA είναι 0,50 και αυτό σημαίνει, ότι και εδώ υπάρχει μέτρια συμφωνία (moderate agreement).



Γράφημα 6. Ποσοστιαία αξιολόγηση της ακρίβειας του νοήματος των προτάσεων, που έχουν προκύψει από δυο διαφορετικά συστήματα MM.



Γράφημα 7. Ποσοστιαία αξιολόγηση της ευχέρειας των προτάσεων, που έχουν προκύψει από δυο διαφορετικά συστήματα MM.

Από το γράφημα 6 για την αξιολόγηση της ευχέρειας, φαίνεται ότι για το ιατρικό κείμενο, το PureNMT έχει με διαφορά καλύτερη απόδοση της γραμματικότητας του κειμένου. Αν και το μεγαλύτερο ποσοστό του GNMT είναι το 34% στο επίπεδο «Καλή» της κλίμακας Likert, το μεγαλύτερο ποσοστό (40%) του PureNMT βρίσκεται στο επίπεδο 4 της κλίμακας Likert (Πολύ καλή). Αξιοσημείωτο είναι και το ποσοστό των προτάσεων του GNMT που καταχωρήθηκαν ως «Εξαιρετικές» (22% ενώ το PureNMT μόλις έχει 12%). Επίσης, σύμφωνα με το παραπάνω γράφημα τα δύο συστήματα έχουν το ίδιο ποσοστό απαντήσεων ως «Πολύ κακή», και παρόμοιο ποσοστό απαντήσεων στο επίπεδο «Κακή». Η απόδοση και τον δύο συστημάτων φαίνεται να είναι σχετικά ίδια και σε αυτήν την ενότητα της αξιολόγησης, όμως υπάρχουν στοιχειώδεις διαφορές που καθιστούν το PureNMT καλύτερο στην απόδοση της ευχέρειας. Οι επισημειωτές είχαν, επίσης, μέτρια συμφωνία (moderate agreement) με 0,50, ακριβώς όση συμφωνία είχαν και στην αξιολόγηση της ακρίβειας.



Γράφημα 8. Ποσοστιαία κατηγοριοποίηση των λαθών των προτάσεων, που έχουν προκύψει από δυο διαφορετικά συστήματα MM.

Και σ' αυτό το κείμενο, οι επισημειωτές προχώρησαν στην κατηγοριοποίηση των λαθών που εντοπίστηκαν στα μεταφράσματα των δυο συστημάτων. Συνολικά, το GNMT είχε 1% περισσότερα λάθη από το PureNMT. Τα ποσοστά των λαθών έχουν μικρή διαφορά εδώ, όπως και στα υπόλοιπα κομμάτια της αξιολόγησης γι' αυτό το κείμενο. Τα περισσότερα λάθη, βάσει ποσοστών, βρίσκονται στην ορολογία και για τα δυο συστήματα με το PureNMT να έχει

ελαφρώς περισσότερα (34% εν αντιθέσει με το GNMT, που έχει 30%). Την ίδια ακριβώς διαφορά έχουν τα ποσοστά και στην κατηγορία γραμματικών λαθών, με το GNMT να βρίσκεται στο 24% και το PureNMT στο 28%. Το ίδιο ποσοστό παρανοήσεων έχουν και τα δυο συστήματα, στο 14%. Η μεγαλύτερη διαφορά είναι στην κατηγορία των λαθών στο ύφος, με το GNMT να έχει 10% περισσότερα λάθη από το PureNMT, γεγονός που επιβεβαιώνει και την διαφορά που υπάρχει στην ευχέρεια. Για παράδειγμα, στις μεταφράσεις της αγγλικής πρότασης «There are two subspecies of the parasite *Trypanosoma brucei* that cause disease in humans.», του GNMT «Υπάρχουν δύο υποείδη του παρασίτου *Trypanosoma brucei* που προκαλούν ασθένεια στους ανθρώπους.» και του PureNMT «Υπάρχουν δύο υποείδη του παρασίτου *Trypanosoma brucei* που προκαλούν ασθένειες στους ανθρώπους.», και οι δυο επισημειωτές κρίνουν ότι υπάρχουν λάθη στο ύφος της μετάφρασης του GNMT ενώ κανείς δεν βρίσκει τέτοιου είδους λάθος στην μετάφραση του PureNMT. Άλλο ένα ίδιο παράδειγμα, με το GNMT να παρουσιάζει λάθη στο ύφος, ενώ το PureNMT όχι, είναι η μετάφραση της πρότασης «Epidemics of sleeping sickness have been a significant public health problem in the past, but the disease is reasonably well-controlled at present, with 7,000-10,000 cases reported annually in recent years.» από το PureNMT ως «Οι επιδημίες της ασθένειας του ύπνου έχουν αποτελέσει σημαντικό πρόβλημα για τη δημόσια υγεία στο παρελθόν, αλλά η ασθένεια ελέγχεται επαρκώς επί του παρόντος, με 7.000-10.000 κρούσματα να αναφέρονται ετησίως τα τελευταία χρόνια.» και από το GNMT ως «Οι επιδημίες της νόσου του ύπνου υπήρξαν ένα σημαντικό πρόβλημα δημόσιας υγείας στο παρελθόν, αλλά η ασθένεια αυτή τη στιγμή είναι αρκετά καλά ελεγχόμενη, ενώ 7,000-10,000 περιπτώσεις αναφέρθηκαν ετησίως τα τελευταία χρόνια.». Επιπλέον, παραδείγματα που επιβεβαιώνουν την ύπαρξη γραμματικών λαθών και στα δυο συστήματα είναι η αγγλική πρόταση «*T. b. rhodesiense* (East African sleeping sickness) is found in focal areas of eastern and southeastern Africa.», η οποία μεταφράστηκε από το GNMT ως «*T. B. rhodesiense* (ανατολικός αφρικανικός ύπνος) βρίσκεται σε περιοχές εστίασης της ανατολικής και νοτιοανατολικής Αφρικής.» και από το PureNMT ως «*T. β. η ροδεσιένζη* (νόσος του ύπνου στην Ανατολική Αφρική) απαντάται σε εστιακές περιοχές της ανατολικής και νοτιοανατολικής Αφρικής.», όπως επίσης και η αγγλική πρόταση «*Tsetse flies inhabit rural areas, living in the woodlands and thickets that dot the East African savannah.*», της οποίας η μετάφραση από το GNMT είναι «*Tsetse* πετούν κατοίκους αγροτικών περιοχών, που ζουν στα δάση και τα παχιά που τεμαχίζουν την ανατολική αφρικανική σαβάνα.» και από το PureNMT είναι «Οι μύγες της *Tsetse* κατοικούν σε αγροτικές περιοχές, ζουν στα δάση και στα μεγάλα καλάθια που περιβάλλουν την Ανατολική Αφρική σαβάνα.». Εντύπωση προκαλεί και μια πρόταση του κειμένου, της οποίας η απόδοση του PureNMT παρουσιάζει πλήθος τυπογραφικών λαθών και είναι η μετάφραση της αγγλικής πρότασης «Humans are the important reservoir of infection, although the parasite can sometimes be found in domestic animals (e.g., pigs, dogs, goats).» ως «Οι άνθρωποι αριστολούν τη σημαντική δεξαμενή όλυνση, αν και το ριαρασίτο βρίσκεται ενίοτε σε οικόσιτα ζώα (ρί.χ. χοίροι, σκύλοι, αίγες).». Αντίθετα το GNMT δεν παρουσιάζει τέτοιου είδους λάθος στη δική του απόδοση της πρότασης, η οποία είναι «Οι άνθρωποι είναι η σημαντική δεξαμενή μόλυνσης, αν και το παράσιτο μπορεί μερικές φορές να βρεθεί σε κατοικίδια ζώα (π.χ. χοίροι, σκύλοι, κατσίκες).». Αυτό αποτελεί το μοναδικό παράδειγμα τυπογραφικών λαθών σε ολόκληρο το κείμενο. Τα δυο συστήματα ενέχουν λάθη, που ανήκουν στην ευρύτερη κατηγορία της ακρίβειας, με το GNMT να έχει ορισμένα λάθη παραπάνω στις κατηγορίες «Προσθήκες», «Παραλήψεις», ίσα ποσοστά στην κατηγορία των παρανοήσεων και λιγότερο ποσοστό λαθών από το PureNMT στην κατηγορία «Λάθη στην ορολογία». Ένα παράδειγμα, που αποδεικνύει την παρουσία παρανοήσεων και στα δύο συστήματα είναι η απόδοση της αγγλικής πρότασης «Imported infection in the U.S. is

extremely rare, and most cases have occurred in African nationals who have immigrated rather than in returning U.S. travelers.», η οποία αποδόθηκε από το GNMT ως «Η εισαγόμενη λοίμωξη στις Η.Π.Α. είναι εξαιρετικά σπάνια και οι περισσότερες περιπτώσεις έχουν συμβεί σε αφρικανικούς υπηκόους που έχουν μεταναστεύσει και όχι σε ταξιδιώτες που επιστρέφουν στις ΗΠΑ.» και από το PureNMT ως «Η εισαγόμενη μόλυνση στις Η.Π.Α. είναι εξαιρετικά σπάνια, και οι περισσότερες περιπτώσεις έχουν συμβεί σε Αφρικανούς υπηκόους που έχουν μεταναστεύσει αντί να επιστρέφουν σε Αμερικάνους ταξιδιώτες.». Επιπλέον, άλλο ένα παράδειγμα με λάθη ορολογίας είναι η αγγλική πρόταση «A wild animal reservoir is thought to be responsible for sporadic transmission to hunters and visitors to game parks.», της οποίας η απόδοση από το GNMT είναι «Μια δεξαμενή άγριων ζώων θεωρείται υπεύθυνη για σποραδική μετάδοση σε κυνηγούς και επισκέπτες σε πάρκα παιχνιδιών.» και από το PureNMT είναι «Δεξαμενή άγριων ζώων θεωρείται υπεύθυνη για τη σποραδική μετάδοση σε κυνηγούς και επισκέπτες σε πάρκα θηραμάτων.». Σ' αυτό το παράδειγμα και οι δυο επισημειωτές συμφωνούν ότι υπάρχουν λάθη ορολογίας και στα δυο συστήματα. Γι' αυτό παρατηρούμε, ότι στην αξιολόγηση της ακρίβειας υπάρχει διαφορά ανάμεσα στα δυο συστήματα, όμως κανένα από τα δυο δεν έχει αξιολογηθεί υψηλά για την μεταφορά του νοήματος του συγκεκριμένου κειμένου από την αγγλική στην ελληνική. Τέλος, για την κατηγοριοποίηση των λαθών στο ιατρικό κείμενο, υπάρχει πολύ καλή συμφωνία (near perfect agreement), με το IAA να υπολογίζεται 0,87 για την κατηγοριοποίηση του αποτελέσματος του PureNMT και 0,83 για το αποτέλεσμα του GNMT.

10. Συζήτηση αποτελεσμάτων

Σύμφωνα με τα παραπάνω αποτελέσματα, το PureNMT φάνηκε να έχει την καλύτερη απόδοση συνολικά και σε όλα τα επίπεδα αξιολόγησης, σε σχέση με το GNMT και στους δυο κειμενικούς τύπους. Η καλύτερη απόδοση με διαφορά παρατηρήθηκε στο νομικό κείμενο, συγκριτικά με το GNMT, ενώ στο ιατρικό, η απόδοση και των δυο συστημάτων ήταν σχετικά ίδια. Έτσι, συμπεραίνουμε ότι το PureNMT έχει καλύτερη απόδοση στα νομικά κείμενα, ενώ στα ιατρικά αποδίδει ικανοποιητικά, όπως και το GNMT. Μπορούμε να υποθέσουμε, ότι το PureNMT ίσως είναι καλύτερα εκπαιδευμένο για μετάφραση νομικών κειμένων, ενώ η απόδοση του στα ιατρικά κείμενα είναι χαμηλότερη, όπως επιβεβαιώνουν και τα αποτελέσματα της έρευνας. Παρ' όλα αυτά, η επίδοση του είναι καλύτερη αντιπαραβολικά με το GNMT και σ' αυτόν τον κειμενικό τύπο. Μια υπόθεση που αιτιολογεί αυτό το αποτέλεσμα είναι, ότι τα δεδομένα εκπαίδευσης του PureNMT είναι πιο καθαρά και από συγκεκριμένες πηγές, ενώ του GNMT, παρόλο που τα περνά από μια επεξεργασία "καθαρισμού", δεν ελέγχονται αναλυτικά και έτσι η απόδοση του συστήματος είναι συγκριτικά χειρότερη από το PureNMT.

Σύμφωνα με το IAA, τα αποτελέσματα των επισημειωτών είναι, ως επί το πλείστον, αξιόπιστα και αρκετά αντικειμενικά, με μερικά, όμως, σημεία στην αξιολόγηση να θέτουν υπό αμφισβήτηση την αντικειμενική κρίση των επισημειωτών, όπως στην αξιολόγηση της ακρίβειας και της ευχέρειας και για τα δυο κείμενα και στην κατάταξη για το ιατρικό κείμενο. Παρ' όλα αυτά, τα αποτελέσματα, που προέκυψαν σ' αυτές τις ενότητες της αξιολόγησης, επιβεβαιώνονται από τα αποτελέσματα της κατηγοριοποίησης των λαθών και ενισχύεται η ορθότητά τους με την υψηλή συμφωνία των επισημειωτών, κατά την κατηγοριοποίηση των λαθών και στα δυο κείμενα.

Επίσης, σε παρόμοια έρευνα που έχει γίνει από τους Stasimioti et al. (2020), για τον ίδιο γλωσσικό συνδυασμό, προκύπτει, ότι το σύστημα που έχει εκπαιδευτεί για ένα συγκεκριμένο

κειμενικό είδος, δίνει καλύτερα αποτελέσματα από το GNMT, με το οποίο γίνεται συγκριτική αξιολόγηση. Αυτό αποδεικνύει ότι η απόδοση του εκάστοτε συστήματος επηρεάζεται σημαντικά από τα δεδομένα εκπαίδευσής του. Επιπλέον, για τα νευρωνικά συστήματα MM έχουν εντοπιστεί συγκεκριμένες κατηγορίες λαθών, που εμφανίζονται συχνά στα αποτελέσματα της MM και αυτές οι κατηγορίες εμφανίστηκαν και στη παρούσα έρευνα με μεγαλύτερα ποσοστά σε σχέση με άλλες κατηγορίες. Αυτές ήταν οι «Παρανοήσεις», «Λάθη στην ορολογία», «Γραμματικά λάθη» και «Λάθη στο ύφος». Τα αποτελέσματα αυτά για τις συνήθεις κατηγορίες λαθών που εντοπίζονται σε κείμενα που έχουν μεταφραστεί αυτόματα από συστήματα MM συμφωνούν με τα αποτελέσματα της προαναφερθείσας έρευνας για τον ίδιο γλωσσικό συνδυασμό Αγγλικά – Ελληνικά (Stasimioti et al., 2020). Η συγκεκριμένη έρευνα έγινε με δημοσιογραφικά κείμενα, μη ειδικά, με γενικό λεξιλόγιο, χωρίς κάποια συγκεκριμένη ορολογία. Παρ' όλα αυτά, οι κατηγορίες λαθών είναι παρόμοιες ανεξαρτήτου είδους κειμένων.

11. 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 γενικής χρήσης (generic), που μεταφράζουν δυο διαφορετικούς τύπους κειμένων, έτσι ώστε να αποδειχθεί, ποιο σύστημα έχει την καλύτερη απόδοση στο σύνολο του κάθε κειμένου και αξιολογούνται αναλυτικά ανά πρόταση, ως προς την ακρίβεια, την ευχέρεια και την κατηγοριοποίηση των λαθών που εντοπίζονται στα αποτελέσματα του κάθε συστήματος. Στόχος της έρευνας, όμως, δεν είναι μόνο η ανάδειξη των αποτελεσμάτων, αλλά και η παρουσίαση της μεθοδολογίας της ανθρώπινης αξιολόγησης της MM ως διαδικασία. Για την υλοποίηση της συγκεκριμένης έρευνας, έλαβαν μέρος δυο επαγγελματίες μεταφραστές ως επισημειωτές, έτσι ώστε να αξιολογήσουν την ποιότητα. Τα αποτελέσματα της έρευνας αυτής είναι αντιπροσωπευτικά, όμως, όπως εξηγήθηκε και παραπάνω, για να διαπιστωθεί η αντικειμενικότητά τους, είναι απαραίτητο να γίνουν παρόμοιες μελέτες με τους ίδιους τύπους κειμένων. Για πλήρως αντιπροσωπευτικά αποτελέσματα, μια πιθανή λύση θα ήταν η μελέτη πολλαπλών κειμένων ενός κειμενικού είδους, μαζί με τη χρήση αυτόματων μετρικών για την ενίσχυση των αποτελεσμάτων της αξιολόγησης, που γίνεται από τον άνθρωπο. Όπως εξηγήθηκε και στη θεωρία, ναι μεν, 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. Με την ανθρώπινη αξιολόγηση είναι ευκολότερο να εντοπιστούν οι αδυναμίες του εκάστοτε συστήματος και με τις αυτόματες μετρικές μπορεί να εξεταστεί αντικειμενικά η απόδοση του εκάστοτε συστήματος στο σύνολο του κειμένου. Έτσι, ο συνδυασμός των δυο τεχνικών αξιολόγησης μπορεί να καταστήσει τα δεδομένα αντικειμενικά και να δημιουργηθούν συμπεράσματα, τα οποία θα βοηθήσουν, τόσο την ακαδημαϊκή κοινότητα, όσο και τους επαγγελματίες μεταφραστές, που αναπτύσσουν πλέον τα δικά τους νευρωνικά συστήματα MM και παράλληλα αξιολογούν την απόδοσή τους. Η μεθοδολογία της συγκεκριμένης έρευνας ήταν μια χρονοβόρα και απαιτητική διαδικασία, πράγμα που επιβεβαιώνει ότι η ανθρώπινη αξιολόγηση απαιτεί χρόνο και είναι πολυέξοδη. Τα κείμενα, που επιλέχθηκαν γι' αυτήν την εργασία, 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 δύο συστημάτων για την αυτόματη μετάφραση ενός νομικού και ενός ιατρικού κειμένου από MM, δίνοντάς μας μια ιδέα για την απόδοση των

συγκεκριμένων σε τέτοιου είδους κείμενα. Εξ όσων γνωρίζω, η παρούσα μελέτη εξετάζει για πρώτη φορά τους δυο συγκριμένους τύπους συστημάτων σ' αυτόν τον γλωσσικό συνδυασμό και στο συγκεκριμένο κειμενικό είδος, όπως αναφέρεται και στην «Εισαγωγή». Εφόσον, λοιπόν, δεν υπάρχει παρόμοια έρευνα, που να εξετάζει την αποτελεσματικότητα του GNMT και του PureNMT στους συγκεκριμένους κειμενικούς τύπους, τα αποτελέσματα της παρούσας δεν μπορούν να επιβεβαιωθούν ή να διαψευσθούν μέχρι ώρας.

12. Παράρτημα

Νομικό Κείμενο – Πηγή

The Companies Auditors Disciplinary Board (CADB) is an independent statutory body established under Part 11 of the ASIC Act. It is responsible for determining whether a RCA's registration should be cancelled or suspended. This includes determining whether a RCA has:

- breached certain requirements under the Corporations Act,
- failed to comply with a condition of their registration,
- failed to adequately and properly perform the duties of a RCA, or
- is otherwise not a fit and proper person.

CADB jurisdiction is limited to matters referred to it by ASIC and APRA, and to matters involving RCAs. Its functions and powers are conferred to it by particular sections of the ASIC and Corporations Acts. Members of CADB are appointed by the Treasurer and the ASIC Act requires members to have a breadth of law, accounting, and business expertise. CADB currently has seven board members including a chairperson and a deputy chairperson both of whom must hold legal qualifications. Legislation permits the appointment of up to 12 members. CADB does not have any permanent administrative staff, with resources allocated from ASIC when a panel is convened. A three- or five-member panel is convened from the members of CADB on an ad hoc basis when the chairperson receives an application. The composition of the panel for a matter will depend on the requirements of that particular matter and member availability, but must include at least one business member and one accounting member and be chaired by the CADB chairperson or deputy chairperson. From 1 March 2017, the power to discipline registered liquidators shifted from the CALDB to committees constituted under Division 40 of the Insolvency Practice Schedule (Corporations) to the Corporations Act. The CALDB was renamed the CADB, and retained disciplinary powers with respect to RCAs. As a result, the volume of matters being heard by CADB is less than when it was in the form of CALDB. As part of these reforms, the discipline of registered liquidators shifted to a different disciplinary process. CADB has received six applications from ASIC on RCA-related matters since 2010. Prior to the responsibility for disciplining liquidators being removed, it received substantially more applications on liquidator-related matters. To

facilitate compliance with CADB's key statutory requirements, CADB has designed its practices and procedures:

- to ensure that the proceedings at a hearing are conducted with as little formality and technicality, and with as much expedition, as the requirements of the Corporations legislation and a proper consideration of the matters before CADB permit; and
- to ensure that the rules of natural justice are observed at and in connection with a hearing.

Αποτέλεσμα αυτόματης μετάφρασης από το Google Translate

Το Πειθαρχικό Συμβούλιο των Ελεγκτών Εταιρειών (CADB) είναι ανεξάρτητο θεσμικό όργανο που ιδρύθηκε σύμφωνα με το Μέρος 11 του Νόμου ASIC. Είναι υπεύθυνη για τον προσδιορισμό του αν θα πρέπει να ακυρωθεί ή να ανασταλεί η εγγραφή του RCA. Αυτό περιλαμβάνει τον προσδιορισμό εάν ένα RCA έχει:

- παραβίασε ορισμένες απαιτήσεις βάσει του νόμου περί εταιρειών,
- δεν συμμορφώθηκε με την προϋπόθεση της εγγραφής τους,
- δεν κατάφερε να εκτελέσει κατάλληλα και σωστά τα καθήκοντα ενός RCA, ή
- διαφορετικά δεν είναι κατάλληλο και κατάλληλο πρόσωπο.

Η αρμοδιότητα της CADB περιορίζεται στα θέματα που της υποβάλλονται από την ASIC και την APRA και σε θέματα που αφορούν RCA. Οι αρμοδιότητες και οι αρμοδιότητές του ανατίθενται σε συγκεκριμένα τμήματα της Πράξης ASIC και των εταιρειών. Τα μέλη του CADB διορίζονται από τον Ταμία και ο νόμος ASIC απαιτεί από τα μέλη να έχουν ένα εύρος δικαίου, λογιστικής και επιχειρηματικής εμπειρογνωμοσύνης. Το CADB έχει σήμερα επτά μέλη του διοικητικού συμβουλίου, μεταξύ των οποίων ένας πρόεδρος και ένας αντιπρόεδρος, οι οποίοι και οι δύο πρέπει να διαθέτουν νομικά προσόντα. Η νομοθεσία επιτρέπει τον διορισμό μέχρι 12 μελών. Το CADB δεν διαθέτει μόνιμο διοικητικό προσωπικό, με πόρους που διατίθενται από την ASIC όταν συγκαλείται ομάδα. Μια επιτροπή τριών ή πενταμελών συγκαλείται από τα μέλη του CADB σε ad hoc βάση όταν ο πρόεδρος λαμβάνει μια αίτηση. Η σύνθεση της επιτροπής για ένα θέμα θα εξαρτηθεί από τις απαιτήσεις της συγκεκριμένης ύλης και τη διαθεσιμότητα των μελών, αλλά θα πρέπει να περιλαμβάνει τουλάχιστον ένα μέλος της επιχείρησης και ένα μέλος της λογιστικής και να προεδρεύεται από τον πρόεδρο ή τον αντιπρόεδρο του CADB. Από την 1η Μαρτίου 2017, η εξουσία πειθαρχίας των εγγεγραμμένων εκκαθαριστών μετατοπίστηκε από το CALDB στις επιτροπές που συγκροτήθηκαν στο τμήμα 40 του Πίνακα Πρακτικών Αφερεγγυότητας (Corporations) στο νόμο περί εταιρειών. Το CALDB μετονομάστηκε σε CADB και διατήρησε πειθαρχικές εξουσίες σε σχέση με τους RCAs. Ως αποτέλεσμα, ο όγκος των ζητημάτων που ακούει η CADB είναι μικρότερος από ό, τι όταν ήταν υπό μορφή CALDB. Στο πλαίσιο αυτών των μεταρρυθμίσεων, η πειθαρχία των εγγεγραμμένων εκκαθαριστών μετατοπίστηκε σε μια διαφορετική πειθαρχική διαδικασία. Το CADB έλαβε από την ASIC έξι αιτήσεις από το ASIC σχετικά με θέματα που σχετίζονται με την RCA από το 2010. Πριν από την ευθύνη για την κατάργηση της πειθαρχίας των εκκαθαριστών, έλαβε ουσιαστικά περισσότερες αιτήσεις σε θέματα που σχετίζονται με εκκαθαριστές. Προκειμένου να διευκολυνθεί η συμμόρφωση με τις βασικές κανονιστικές απαιτήσεις της CADB, η CADB σχεδίασε τις πρακτικές και τις διαδικασίες της:

- να διασφαλιστεί ότι η διαδικασία σε ακρόαση διεξάγεται με ελάχιστη διατύπωση και τεχνική και με όσο το δυνατόν μεγαλύτερη αποστολή, όπως οι απαιτήσεις της νομοθεσίας των εταιρειών και η σωστή εξέταση των ζητημάτων πριν από το CADB. και
- να διασφαλιστεί ότι τηρούνται οι κανόνες της φυσικής δικαιοσύνης κατά την ακρόαση και σε σχέση με αυτήν.

Αποτέλεσμα αυτόματης μετάφρασης από το Pure NMT

Το Πειθαρχικό Συμβούλιο των Ελεγκτών Εταιρειών (CADB) είναι ένας ανεξάρτητος θεσμοθετημένος οργανισμός που συστάθηκε δυνάμει του Μέρους 11 του νόμου ASIC. Είναι υπεύθυνο για να καθορίσει εάν η καταχώριση μιας RCA πρέπει να ακυρωθεί ή να ανασταλεί.

Αυτό περιλαμβάνει τον προσδιορισμό του εάν μια RCA έχει:

- παραβίασε ορισμένες απαιτήσεις του νόμου περί εταιρειών,
- δεν τήρησαν την προϋπόθεση της καταχώρισής τους,
- απέτυχε να εκτελέσει επαρκώς και ορθά τα καθήκοντα της RCA, ή
- κατά τα άλλα δεν είναι κατάλληλο και κατάλληλο άτομο.

Η δικαιοδοσία του CADB περιορίζεται σε θέματα που του παραπέμπονται από την ASIC και την APRA, καθώς και σε θέματα που αφορούν τις ΑΠΑ. Τα καθήκοντα και οι εξουσίες του ανατίθενται σε αυτήν από συγκεκριμένα τμήματα του ASIC και του νόμου περί εταιρειών. Τα μέλη του CADB διορίζονται από το Δημόσιο Ταμείο και ο νόμος ASIC απαιτεί από τα μέλη να έχουν ένα εύρος νομικής, λογιστικής και επιχειρηματικής εμπειρογνομosύνης. Το CADB διαθέτει επί του παρόντος επτά μέλη του συμβουλίου, μεταξύ των οποίων έναν πρόεδρο και έναν αντιπρόεδρο, αμφότεροι εκ των οποίων πρέπει να έχουν νομικά προσόντα. Η νομοθεσία επιτρέπει το διορισμό μέχρι 12 μελών. Το CADB δεν διαθέτει μόνιμο διοικητικό προσωπικό, με τους πόρους να διατίθενται από το ASIC κατά τη σύγκληση μιας επιτροπής. Μια τριμελής ή πενταμελής επιτροπή συγκαλείται από τα μέλη του CADB σε ad hoc βάση όταν ο πρόεδρος λαμβάνει αίτηση. Η σύνθεση της επιτροπής για ένα θέμα θα εξαρτηθεί από τις απαιτήσεις του συγκεκριμένου θέματος και τη διαθεσιμότητα των μελών, αλλά πρέπει να περιλαμβάνει τουλάχιστον ένα μέλος της επιχείρησης και ένα λογιστικό μέλος και να προεδρεύεται από τον πρόεδρο ή τον αντιπρόεδρο του CADB. Από την 1η Μαρτίου 2017, η εξουσία πειθαρχίας των εγγεγραμμένων εκκαθαριστών μετατέθηκε από την CALDB σε επιτροπές που συνιστώνται στο πλαίσιο του τμήματος 40 του χρονοδιαγράμματος για τις πρακτικές αφερεγγυότητας (εταιρείες) στον νόμο περί εταιρειών. Το CALDB μετονομάστηκε σε CADB και διατήρησε πειθαρχικές εξουσίες όσον αφορά τις ΑΠΑ. Κατά συνέπεια, ο όγκος των θεμάτων που εξετάζονται από το CADB είναι μικρότερος από ό,τι όταν ήταν υπό μορφή CALDB. Στο πλαίσιο αυτών των μεταρρυθμίσεων, η πειθαρχία των εγγεγραμμένων εκκαθαριστών μεταφέρθηκε σε διαφορετική πειθαρχική διαδικασία. Το CADB έχει λάβει έξι αιτήσεις από το ASIC σχετικά με θέματα που αφορούν την RCA από το 2010. Πριν από την απομάκρυνση της ευθύνης για την πειθάρχηση των εκκαθαριστών, έλαβε σημαντικά περισσότερες αιτήσεις για θέματα που σχετίζονται με τον εκκαθαριστή. Για να διευκολυνθεί η συμμόρφωση με τις βασικές κανονιστικές απαιτήσεις του CADB, το CADB έχει σχεδιάσει τις πρακτικές και τις διαδικασίες του:

- να διασφαλίσει ότι οι διαδικασίες σε μια ακρόαση διεξάγονται με ελάχιστη τυπικότητα και τεχνική λεπτομέρεια, και με την ίδια ταχύτητα όπως οι απαιτήσεις της νομοθεσίας των Εταιρειών και η σωστή εξέταση των θεμάτων πριν από την άδεια του CADB, και
- να εξασφαλίζει την τήρηση των κανόνων της φυσικής δικαιοσύνης κατά την ακρόαση και στο πλαίσιο αυτής.

Ιατρικό Κείμενο – Πηγή

Epidemiology & Risk Factors

There are two subspecies of the parasite *Trypanosoma brucei* that cause disease in humans. The clinical features of the infection depend on the subspecies involved. The two subspecies are found in different regions of Africa. At present, there is no overlap in their geographic distribution.

T. b. rhodesiense (East African sleeping sickness) is found in focal areas of eastern and southeastern Africa. Each year a few hundred cases are reported to the World Health Organization. Over 95% of the cases of human infection occur in Tanzania, Uganda, Malawi, and Zambia. Animals are the primary reservoir of infection. A wild animal reservoir is thought to be responsible for sporadic transmission to hunters and visitors to game parks. Infection of international travelers is rare, but it occasionally occurs. In the U.S., one case per year, on average, is diagnosed. Most cases of sleeping sickness imported into the U.S. have been in travelers who were on safari in East Africa.

T. b. gambiense (West African sleeping sickness) is found predominantly in central Africa and in limited areas of West Africa. Most of the sleeping sickness in Africa is caused by this form of the parasite. Epidemics of sleeping sickness have been a significant public health problem in the past, but the disease is reasonably well-controlled at present, with 7,000-10,000 cases reported annually in recent years. Over 95% of the cases of human infection are found in Democratic Republic of Congo, Angola, Sudan, Central African Republic, Chad, and northern Uganda. Humans are the important reservoir of infection, although the parasite can sometimes be found in domestic animals (e.g., pigs, dogs, goats). Imported infection in the U.S. is extremely rare, and most cases have occurred in African nationals who have immigrated rather than in returning U.S. travelers.

Both forms of sleeping sickness are transmitted by the bite of the tsetse fly (*Glossina* species). Tsetse flies inhabit rural areas, living in the woodlands and thickets that dot the East African savannah. In central and West Africa, they live in the forests and vegetation along streams. Tsetse flies bite during daylight hours. Both male and female flies can transmit the infection, but even in areas where the disease is endemic, only a very small percentage of flies are infected. Although the vast majority of infections are transmitted by the tsetse fly, other modes of transmission are possible.

Αποτέλεσμα αυτόματης μετάφρασης από το Google Translate

Επιδημιολογία & Παράγοντες Κινδύνου

Υπάρχουν δύο υποείδη του παρασίτου *Trypanosoma brucei* που προκαλούν ασθένεια στους ανθρώπους. Τα κλινικά χαρακτηριστικά της μόλυνσης εξαρτώνται από τα εμπλεκόμενα

υποείδη. Τα δύο υποείδη βρίσκονται σε διάφορες περιοχές της Αφρικής. Προς το παρόν, δεν υπάρχει επικάλυψη στη γεωγραφική τους κατανομή.

T. B. rhodesiense (ανατολικός αφρικανικός ύπνος) βρίσκεται σε περιοχές εστίασης της ανατολικής και νοτιοανατολικής Αφρικής. Κάθε χρόνο αναφέρονται μερικές εκατοντάδες περιπτώσεις στην Παγκόσμια Οργάνωση Υγείας. Πάνω από το 95% των περιπτώσεων ανθρώπινης μόλυνσης συμβαίνουν στην Τανζανία, την Ουγκάντα, το Μαλάουι και τη Ζάμπια. Τα ζώα είναι η κύρια δεξαμενή μόλυνσης. Μια δεξαμενή άγριων ζώων θεωρείται υπεύθυνη για σποραδική μετάδοση σε κυνηγούς και επισκέπτες σε πάρκα παιχνιδιών. Η μόλυνση των διεθνών ταξιδιωτών είναι σπάνια, αλλά συμβαίνει περιστασιακά. Στις Η.Π.Α. διαγιγνώσκεται μία περίπτωση κατά μέσο όρο, κατά μέσο όρο. Οι περισσότερες περιπτώσεις νοσημάτων ύπνου που εισάγονται στις Η.Π.Α. ήταν σε ταξιδιώτες που βρίσκονταν σε σαφάρι στην Ανατολική Αφρική.

T. B. gambiense (Δυτική αφρικανική υπνηλία) βρίσκεται κυρίως στην κεντρική Αφρική και σε περιορισμένες περιοχές της Δυτικής Αφρικής. Το μεγαλύτερο μέρος της νόσου του ύπνου στην Αφρική οφείλεται σε αυτή τη μορφή του παρασίτου. Οι επιδημίες της νόσου του ύπνου υπήρξαν ένα σημαντικό πρόβλημα δημόσιας υγείας στο παρελθόν, αλλά η ασθένεια αυτή τη στιγμή είναι αρκετά καλά ελεγχόμενη, ενώ 7,000-10,000 περιπτώσεις αναφέρθηκαν ετησίως τα τελευταία χρόνια. Πάνω από το 95% των περιπτώσεων ανθρώπινης μόλυνσης βρίσκονται στη Λαϊκή Δημοκρατία του Κονγκό, την Αγκόλα, το Σουδάν, την Κεντροαφρικανική Δημοκρατία, το Τσαντ και τη βόρεια Ουγκάντα. Οι άνθρωποι είναι η σημαντική δεξαμενή μόλυνσης, αν και το παράσιτο μπορεί μερικές φορές να βρεθεί σε κατοικίδια ζώα (π.χ. χοίροι, σκύλοι, κατσίκες). Η εισαγόμενη λοίμωξη στις Η.Π.Α. είναι εξαιρετικά σπάνια και οι περισσότερες περιπτώσεις έχουν συμβεί σε αφρικανικούς υπηκόους που έχουν μεταναστεύσει και όχι σε ταξιδιώτες που επιστρέφουν στις ΗΠΑ.

Και οι δύο μορφές νόσου του ύπνου μεταδίδονται από το δάγκωμα της μύγας tsetse (είδος *Glossina*). Tsetse πετούν κατοίκους αγροτικών περιοχών, που ζουν στα δάση και τα παχιά που τεμαχίζουν την ανατολική αφρικανική σαβάνα. Στην Κεντρική και Δυτική Αφρική ζουν στα δάση και στη βλάστηση κατά μήκος των ρέων. Tsetse μύγες κατά τη διάρκεια της ημέρας. Τόσο οι αρσενικές όσο και οι θηλυκές μύγες μπορούν να μεταδώσουν τη μόλυνση, αλλά ακόμη και σε περιοχές όπου η ασθένεια είναι ενδημική, μολύνεται μόνο ένα πολύ μικρό ποσοστό μυγών. Παρόλο που η μεγάλη πλειοψηφία των μολύνσεων μεταδίδονται από τη μύγα tsetse, είναι δυνατοί και άλλοι τρόποι μετάδοσης.

Αποτέλεσμα αυτόματης μετάφρασης από το Pure NMT

Επιδημιολογία και παράγοντες κινδύνου

Υπάρχουν δύο υποείδη του παρασίτου *Trypanosoma brucei* που προκαλούν ασθένειες στους ανθρώπους. Τα κλινικά χαρακτηριστικά της λοίμωξης εξαρτώνται από τα εμπλεκόμενα υποείδη. Τα δύο υποείδη απαντώνται σε διάφορες περιοχές της Αφρικής. Επί του παρόντος, δεν υπάρχει επικάλυψη στη γεωγραφική κατανομή τους.

Τ. β. η ροδεσιένζη (νόσος του ύπνου στην Ανατολική Αφρική) απαντάται σε εστιακές περιοχές της ανατολικής και νοτιοανατολικής Αφρικής. Κάθε χρόνο αναφέρονται στην Παγκόσμια Οργάνωση Υγείας μερικές εκατοντάδες περιπτώσεις. Πάνω από το 95% των περιπτώσεων μόλυνσης του ανθρώπου συμβαίνουν στην Τανζανία, την Ουγκάντα, το Μαλάουι και τη Ζάμπια. Τα ζώα αποτελούν την κύρια δεξαμενή μόλυνσης. Δεξαμενή άγριων ζώων θεωρείται

υπεύθυνη για τη σποραδική μετάδοση σε κυνηγούς και επισκέπτες σε πάρκα θηραμάτων. Η λοίμωξη των διεθνών ταξιδιωτών είναι σπάνια, αλλά συμβαίνει περιστασιακά. Στις Η.Π.Α., κατά μέσο όρο, έχει διαγνωστεί μία περίπτωση ετησίως. Οι περισσότερες περιπτώσεις ασθένειας ύπνου που εισήχθη στις ΗΠΑ ήταν σε ταξιδιώτες που ήταν σε σαφάρι στην Ανατολική Αφρική.

Τ. β. η γκαμπιζένζη (νόσος του ύπνου της Δυτικής Αφρικής) απαντάται κυρίως στην κεντρική Αφρική και σε περιορισμένες περιοχές της Δυτικής Αφρικής. Το μεγαλύτερο μέρος της ασθένειας του ύπνου στην Αφρική οφείλεται σε αυτή τη μορφή παρασίτου. Οι επιδημίες της ασθένειας του ύπνου έχουν αποτελέσει σημαντικό πρόβλημα για τη δημόσια υγεία στο παρελθόν, αλλά η ασθένεια ελέγχεται επαρκώς επί του παρόντος, με 7.000-10.000 κρούσματα να αναφέρονται ετησίως τα τελευταία χρόνια. Πάνω από το 95% των περιπτώσεων μόλυνσης του ανθρώπου εντοπίζονται στη Λαϊκή Δημοκρατία του Κονγκό, την Αγκόλα, το Σουδάν, την Κεντροαφρικανική Δημοκρατία, το Τσαντ και τη βόρεια Ουγκάντα. Οι άνθρωποι αριοτελούν τη σημαντική δεξαμενή όλυσης, αν και το ριαρασίτο βρίσκεται ενίοτε σε οικόσιτα ζώα (π.χ. χοίροι, σκύλοι, αίγες). Η εισαγόμενη μόλυνση στις Η.Π.Α. είναι εξαιρετικά σπάνια, και οι περισσότερες περιπτώσεις έχουν συμβεί σε Αφρικανούς υπηκόους που έχουν μεταναστεύσει αντί να επιστρέφουν σε Αμερικάνους ταξιδιώτες.

Και οι δύο μορφές ασθένειας του ύπνου μεταδίδονται με το δάγκωμα της μύγας του τσιγγάρου (είδος *Glossina*). Οι μύγες της *Tsetse* κατοικούν σε αγροτικές περιοχές, ζουν στα δάση και στα μεγάλα καλάθια που περιβάλλουν την Ανατολική Αφρική σαβάνα. Στην κεντρική και δυτική Αφρική, ζουν στα δάση και τη βλάστηση κατά μήκος των ποταμών. Οι μύγες της *Tsetse* δαγκώνουν κατά τη διάρκεια της ημέρας. Τόσο οι αρσενικές όσο και οι θηλυκές μύγες μπορούν να μεταδώσουν τη λοίμωξη, αλλά ακόμη και σε περιοχές όπου η ασθένεια είναι ενδημική, μόνο ένα πολύ μικρό ποσοστό των μύγων μολύνεται. Αν και η συντριπτική πλειονότητα των λοιμώξεων μεταδίδεται μέσω της μύγας του τσιμέντου, είναι δυνατοί άλλοι τρόποι μετάδοσης.

13. 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.
- Caswell, I. and Liang, B. (2020), Recent Advances in Google Translate. Google AI Blog. Retrieved 13 July 2020, from <https://ai.googleblog.com/2020/06/recent-advances-in-google-translate.html>
- 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.
- Common Sense Advisory (CSA), (2016): Global Market Research Firm Common Sense Advisory Finds Post-edited Machine Translation (PEMT) Among Fastest-growing Segments of the Language Industry

- Crego, J., Kim, J., Klein, G., Rebollo, A., Yang, K., Akhanov, J. S., et al. (2016). SYSTRAN's Pure Neural Machine Translation Systems. CoRR abs/1610.05540.
- 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. Retrieved 15 April 2020, from 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.
- Garvin, D. (1984). Product Quality: An Important Strategic Weapon. *Business Horizons*, 27, 40-43. [https://doi.org/10.1016/0007-6813\(84\)90024-7](https://doi.org/10.1016/0007-6813(84)90024-7)
- Glen, S. (2014). "Cohen's Kappa Statistic" From StatisticsHowTo.com: Elementary Statistics for the rest of us! <https://www.statisticshowto.com/cohens-kappa-statistic/>
- 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.
- House, J. (2001). Translation Quality Assessment: Linguistic Description versus Social Evaluation. *Meta*, 46 (2), 243–257. <https://doi.org/10.7202/003141ar>
- 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
- Hutchins, J. (1986). *Machine translation: past, present, future*. Chichester: Ellis Horwood.
- Hutchins, J. (2004). The Georgetown-IBM experiment demonstrated in January 1954. Machine translation: from real users to research: 6th conference of the Association for Machine Translation in the Americas (AMTA), Washington, DC. Proceedings of the conference, pp. 102-114.
- Johnson R., King M., des Tombe, L. (1985). EUROTRA: A Multilingual System under Development. In *Computational Linguistics* 11(2–3), pp. 155-169.
- 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. Retrieved 25 April 2020, from 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>. [Accessed 08 June 2020].
- Nagao, M. (1984). "A framework of a mechanical translation between Japanese and English by analogy principle". In A. Elithorn and R. Banerji (eds). *Artificial and Human Intelligence*. 173-180. Elsevier Science Publishers.
- Navarro, G. (2001). A Guided Tour to Approximate String Matching. In: *ACM Computing Surveys* 33 (1): 31–88.
- 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. (1991). *Textanalyse und Übersetzen: theoretische Grundlagen, Methode und didaktische Anwendung einer übersetzungsrelevanten Textanalyse*, Heidelberg: Groos.
- Nord, C. (1997). *Translating as a Purposeful Activity: Functionalist Approaches Explained*, Manchester: St. Jerome. [Google Scholar]

- Nord, C. (1997). *Translating as a Purposeful Activity: Functionalist Approaches Explained*, Manchester: St. Jerome. [Google Scholar]
- Och, F.J. (2003). Minimum Error Rate Training in Statistical Machine Translation. *Proceedings of the 41st Annual Meeting of the Association of Computational Linguistics*, pages 160–167, July.
- 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.
- 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/metricstr/2008/results/>, October.
- Reiss, K. (1971). *Möglichkeiten und Grenzen der Übersetzungskritik*, München: Hüber. [Google Scholar]
- Reiß, K. (1971). *Möglichkeiten und Grenzen der Übersetzungskritik*. München: Hueber
- Reiß, K. and Vermeer, H.J. (1984). *Grundlegung einer Allgemeinen Translationstheorie*, Tübingen: Niemeyer
- 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.
- 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)*.
- 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.

Vashee, K. (2017). Customization or ‘Specializing’ Neural Machine Translation in SYSTRAN. Medium, ITNEXT. Retrieved 15 April 2020, from 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.

Weaver, W. (1949). Translation. Machine translation of languages: fourteen essays, Cambridge, Mass.: Technology Press of the Massachusetts Institute of Technology, pp. 15-23.

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.

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>