ΕΘΝΙΚΟ ΚΑΙ ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ΦΙΛΟΣΟΦΙΚΗ ΣΧΟΛΗ
ΤΜΗΜΑ ΦΙΛΟΛΟΓΙΑΣ
ΤΟΜΕΑΣ ΓΛΩΣΣΟΛΟΓΙΑΣ

ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ
ΤΟΜΕΑΣ ΣΗΜΑΤΩΝ ΕΛΕΓΧΟΥ ΚΑΙ ΡΟΜΠΟΤΙΚΗΣ

ΙΝΣΤΙΤΟΥΤΟ ΕΠΕΞΕΡΓΑΣΙΑΣ ΤΟΥ ΛΟΓΟΥ
ΕΡΕΥΝΗΤΙΚΟ ΚΕΝΤΡΟ «ΑΘΗΝΑ»

ΔΙΕΠΙΣΤΗΜΟΝΙΚΟ ΔΙΑΠΑΝΕΠΙΣΤΗΜΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ

ΤΕΧΝΟΓΛΩΣΣΙΑ VIII

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Αξιολόγηση ποιότητας μηχανικής μετάφρασης

Τέσσερα συστήματα στο μικροσκόπιο της μη αυτοματοποιημένης αξιολόγησης και της αυτόματης μετρικής Bleu

Ειρήνη Χατζηκουμή

Επιβλέπων: Στέλιος Πιπερίδης (ΙΕΛ)
Συνδιορθωτές: Διονύσης Γούτσος (ΕΚΠΑ)
Αναστάσιος Πατρικάκος (ΙΕΛ)

ΑΘΗΝΑ
ΙΟΥΝΙΟΣ 2015

Ευχαριστίες

Θα ήθελα κατ' αρχάς να ευχαριστήσω για τις μεταφράσεις-αναφορά που χρησιμοποιήθηκαν στην παρούσα εργασία τους συνήθεις αφανείς ήρωες, τους μεταφραστές. Επίσης θα ήθελα να ευχαριστήσω τους συμφοιτητές της Τεχνογλωσσίας VIII Έλπη Μαργαρίτη, Άρη Φεργάδη, Γιώργο Φραγκάκη και Μίλτο Δεληγιάννη για τις γνώσεις και τις συμβουλές που μοιράστηκαν απλόχερα σε όλη τη διάρκεια του μεταπτυχιακού προγράμματος και κατά την εκπόνηση της παρούσας διπλωματικής εργασίας, την υποτιτλίστρια Ρένια Γκούση για τις γλωσσικές της παρατηρήσεις, και τον Σωκράτη Σοφιανόπουλο, επιστημονικό συνεργάτη του ΙΕΛ στο τμήμα Μηχανικής Μετάφρασης, για τις πολύτιμες συμβουλές και τεχνικές γνώσεις με τις οποίες συνέβαλε στην εκπόνηση της παρούσας. Τέλος, ευχαριστώ τον επιβλέποντα, Κο Στέλιο Πιπερίδη, του οποίου την εξαντλητική επιμονή και προσήλωση, αλλά και τον ενθουσιασμό για τη γλώσσα και τη μετάφραση εκτίμησα στο διάστημα αυτό.

Περιεχόμενα

Δεῦτε καὶ καταβάντες συγχέωμεν αὐτῶν ἐκεῖ τὴν γλῶσσαν, ἵνα μὴ ἀκούσωσιν ἕκαστος τὴν φωνὴν τοῦ πλησίον. Καὶ διέσπειρεν αὐτοὺς Κύριος ἐκεῖθεν ἐπὶ πρόσωπον πάσης τῆς γῆς, καὶ ἐπαύσαντο οἰκοδομοῦντες τὴν πόλιν καὶ τὸν πύργον. Διὰ τοῦτο ἐκλήθη τὸ ὄνομα αὐτῆς Σύγχυσις, ὅτι ἐκεῖ συνέχεε Κύριος τὰ χείλη πάσης τῆς γῆς, καὶ ἐκεῖθεν διέσπειρεν αὐτοὺς Κύριος ἐπὶ πρόσωπον πάσης τῆς γῆς.

Γένεση, 11:7-9

1 Εισαγωγή

Η πρόσφατη άνθιση των υπολογιστικών γλωσσικών εφαρμογών καθιστά επιτακτική την ανάγκη να ανταποκριθούν στις εξελίξεις οι γλώσσες με μικρότερη παρουσία στον παγκόσμιο χάρτη, όπως είναι η ελληνική, υπό την απειλή της ψηφιακής εξαφάνισης.

Σπουδαία θέση μεταξύ των εφαρμογών αυτών κατέχει η μηχανική μετάφραση, η οποία χρησιμοποιείται είτε ως μεμονωμένη εφαρμογή είτε ως μέρος άλλων εφαρμογών. Ο μέσος άνθρωπος έχει καθημερινά πρόσβαση σε πλήθος αυτόματων μεταφραστών σε διάφορες συσκευές. Η ποιότητα των μεταφράσεων αυτών μπορεί να ποικίλλει πολύ ανάλογα με το ζεύγος γλωσσών και την κατεύθυνση της μετάφρασης, το πεδίο στο οποίο ανήκει το κείμενο προς μετάφραση, το μεταφραστικό σύστημα που χρησιμοποιείται, τον στόχο για τον οποίο προορίζει ο χρήστης τη μετάφραση, ακόμη και από το αν το σύστημα μηχανικής μετάφρασης χρησιμοποιείται για την εύρεση λέξεων, φράσεων ή ολόκληρων προτάσεων. Αποτελεί βέβαια κοινή παραδοχή ότι υπάρχουν σημαντικά περιθώρια βελτίωσης αυτής της ποιότητας.

Στον χώρο της μετάφρασης, η μηχανική μετάφραση έχει εγείρει σημαντικά ερωτήματα σχετικά με θέματα όπως η χρήση της ως μεταφραστικό εργαλείο, η αποδαιμονοποίησή της, η επίδρασή της στις αμοιβές των μεταφραστών, η τοξικότητα των μηχανικά μεταφρασμένων κειμένων, τα πνευματικά δικαιώματα των κειμένων που μεταφράζονται με ελεύθερες μηχανές μετάφρασης στο διαδίκτυο κ.ά. Ένα θέμα το οποίο δεν έχει ακόμη τεθεί είναι η συνειδητοποίηση του μέσου χρήστη ως προς τον σκοπό για τον οποίο μπορεί να χρησιμοποιήσει τη μηχανική μετάφραση. Δεν είναι αυτονόητο για όλους τους χρήστες π.χ. ότι για να χρησιμοποιήσουν μηχανική μετάφραση για διαφήμιση εταιρείας, πρέπει να προηγηθεί επιμέλεια του μεταφράσματος από άνθρωπο.

Στόχος της παρούσας εργασίας είναι, μέσω της αξιολόγησης ενός σώματος μηχανικών μεταφρασμάτων τεσσάρων διαφορετικών συστημάτων στο ζεύγος αγγλικά-ελληνικά και στις δύο κατευθύνσεις, να αποτιμήσουμε την ποιότητα της μετάφρασης τριών ευρέως χρησιμοποιούμενων συστημάτων και ενός συστήματος υπό ανάπτυξη, να εξαγάγουμε συμπεράσματα σχετικά με αυτή την ποιότητα και να προτείνουμε τρόπους αξιοποίησης των συμπερασμάτων. Για την αξιολόγηση αυτή χρησιμοποιείται η αυτόματη μετρική Bleu και μη αυτοματοποιημένες μέθοδοι. Καθώς τα τρία από τα συστήματα είναι εμπορικά ενώ το τέταρτο είναι υπό ανάπτυξη, η μεθοδολογία μη αυτοματοποιημένης αξιολόγησής τους διαφοροποιείται και στα μεν εμπορικά συστήματα η αξιολόγηση εστιάζεται στην κατάταξή τους ενώ στο υπό ανάπτυξη σύστημα στη βελτίωσή του. Παράλληλα με την ποιότητα των μεταφρασμάτων, αξιολογείται έμμεσα και η αυτόματη μετρική Bleu, καθώς μελετάται η συσχέτισή της με τη μη αυτοματοποιημένη αξιολόγηση.

Στο Κεφάλαιο 2 παρουσιάζονται συνοπτικά οι υπάρχουσες προσεγγίσεις της μηχανικής μετάφρασης με ιδιαίτερη έμφαση στα στατιστικά συστήματα. Στο Κεφάλαιο 3 αναλύονται οι κυριότερες μέθοδοι αξιολόγησης μηχανικής μετάφρασης, οι οποίες κατατάσσονται σε μία από τις δύο κατηγορίες "αυτόματη" και "μη αυτοματοποιημένη". Στο Κεφάλαιο 3.1 περιγράφεται ο τρόπος λειτουργίας των αυτόματων μετρικών αξιολόγησης, με ιδιαίτερη έμφαση στη μετρική Bleu. Στο Κεφάλαιο 3.2 παρουσιάζεται αρχικά η διάκριση μεταξύ μεθόδων μη αυτοματοποιημένης αξιολόγησης με άμεσα εκπεφρασμένη αξιολογική κρίση και μεθόδων μη αυτοματοποιημένης αξιολόγησης χωρίς άμεσα εκπεφρασμένη αξιολογική κρίση, και στη συνέχεια περιγράφονται οι υπάρχουσες τεχνικές από την κάθε κατηγορία, με εκτενή αναφορά στις υφιστάμενες τυπολογίες λαθών. Στο Κεφάλαιο 4 παρουσιάζεται η μεθοδολογία και τα αποτελέσματα της αξιολόγησής μας. Αναλύεται ο τρόπος κατάρτισης του σώματος κειμένων για τη διεξαγωγή των πειραμάτων, γίνεται σύγκριση των υπαρχουσών τυπολογιών λαθών και εκτίθενται τα κριτήρια επιλογής της κατάλληλης τυπολογίας, καθώς και οι οδηγίες επισημείωσης που ακολουθήθηκαν κατά τη μη αυτοματοποιημένη αξιολόγηση. Στη συνέχεια παρουσιάζονται τα αποτελέσματα της αυτόματης και της μη αυτοματοποιημένης αξιολόγησης, καθώς και τα συμπεράσματα που εξάγονται από αυτά. Στο Κεφάλαιο 5 συνοψίζονται τα συμπεράσματά μας και στο Κεφάλαιο 6 προτείνεται μελλοντική εργασία σε συνέχεια της παρούσας.

Η παρούσα εργασία άπτεται του τομέα της μηχανικής μετάφρασης, της θεωρίας της μετάφρασης και της γλωσσολογίας, στους οποίους μεγάλο μέρος των δημοσιεύσεων είναι στην αγγλική γλώσσα. Κυρίως στον τομέα της μηχανικής μετάφρασης δημιουργούνται συνεχώς νέοι όροι η απόδοση των οποίων δεν έχει παγιωθεί στην ελληνική. Γι' αυτόν τον λόγο περιλαμβάνεται στο Παράρτημα γλωσσάρι με τους όρους που έχουν χρησιμοποιηθεί στην παρούσα εργασία. Στο γλωσσάρι αυτό έχουν αποδελτιωθεί οι ελληνικές αποδόσεις που επιλέχθηκαν για το συγκεκριμένο συγκείμενο[1] και παρατίθενται με τους αντίστοιχους αγγλικούς όρους, τόσο σε ευθεία όσο και σε αντίστροφη κατεύθυνση.

2 Μηχανική μετάφραση

Ως μηχανική ή αυτόματη μετάφραση νοείται η μεταφραστική διαδικασία στην οποία δεν συμμετέχει ο άνθρωπος. Ο χρήστης τροφοδοτεί ένα σύστημα με ένα κείμενο-πηγή και λαμβάνει ως εξαγόμενο το κείμενο-στόχος. Κείμενο-πηγή είναι το πρωτότυπο κείμενο προς μετάφραση και κείμενο-στόχος το μετάφρασμα αυτού. Αντίστοιχα γλώσσα-πηγή είναι η γλώσσα του πρωτότυπου κειμένου και γλώσσα-στόχος η γλώσσα στην οποία μεταφράζεται το πρωτότυπο.

[1] Ως συγκείμενο νοείται το άμεσο κειμενικό περιβάλλον, βλ. Γεωργακοπούλου, Α. και Γούτσος, Δ. [1999].

Δύο ακόμη βασικές έννοιες στον χώρο της μετάφρασης που θα μας απασχολήσουν στην παρούσα εργασία αποτελούν οι έννοιες της πιστότητας (accuracy) και της γραμματικότητας (fluency). Για την απόδοση των αγγλικών όρων επιλέξαμε τις προαναφερθείσες αποδόσεις, οι οποίες και εννοούνται ως εξής: Η πιστότητα είναι η ανταπόκριση του μεταφράσματος στο πρωτότυπο, ο βαθμός στον οποίο μεταδίδονται όλα τα μηνύματα αυτού και μόνο αυτά. Στα αγγλικά συναντώνται επίσης οι όροι “fidelity” και “adequacy” συνήθως με την ίδια έννοια, αλλά για τις ανάγκες της παρούσας εργασίας δεν χρειάζεται περαιτέρω διάκριση και μας καλύπτει ο όρος “πιστότητα”. Ο όρος “fluency” εμπεριέχει τις έννοιες της γραμματικής ορθότητας, της ορθής γραφής αλλά και της αποδοχής από το γλωσσικό αισθητήριο του φυσικού ομιλητή. Επιλέγουμε τη μονολεκτική απόδοσή της ως “γραμματικότητα”, παρόλο που υπάρχει και ο όρος “grammaticality” στα αγγλικά, αφενός μεν για να αποφευχθούν περιγραφικές αποδόσεις σε ένα τόσο συχνά χρησιμοποιούμενο όρο, αφετέρου δε διότι η απόδοση αυτή συνάδει με την έννοια των γραμματικώς ορθών προτάσεων κατά Chomsky, οι οποίες είναι αυτές που “γίνονται αποδεκτές ως ορθές από τα μέλη της γλωσσικής κοινότητας” [Μπαμπινιώτης, Γ., 1998] και συνεπώς δεν προσκρούουν στο γλωσσικό τους αισθητήριο[2].

Ακολουθεί στο Κεφάλαιο 2.1 συνοπτική παρουσίαση των βασικών προσεγγίσεων μηχανικής μετάφρασης.

2.1 Συστήματα άμεσης μετάφρασης

Τα πρώτα συστήματα μηχανικής μετάφρασης, τα οποία και υπάγονται στην κατηγορία των άμεσων προσεγγίσεων (direct translation approaches), λειτουργούν εκτελώντας τα εξής βήματα: μορφολογική ανάλυση του κειμένου-πηγή, αναζήτηση σε ένα δίγλωσσο λεξικό και, τέλος, εφαρμογή τοπικών κανόνων αναδιοργάνωσης για την τοποθέτηση των λέξεων στη σωστή σειρά [Σοφιανόπουλος, Σ., 2009].

2.2 Συστήματα βασιζόμενα σε κανόνες (rule-based)

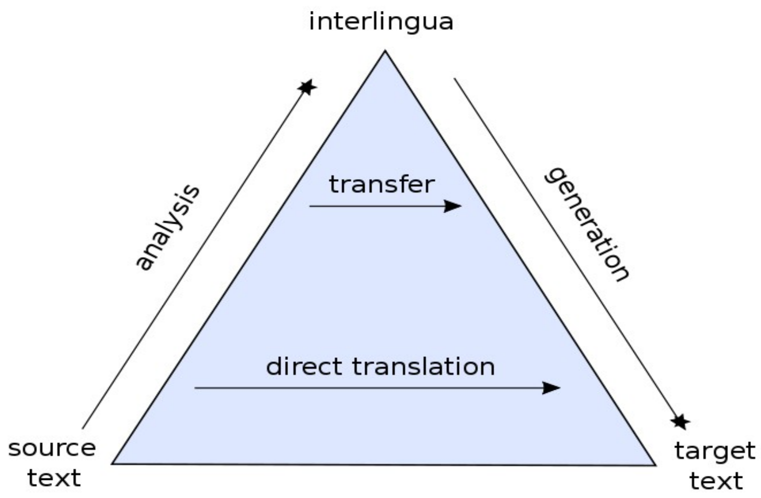
Τα άμεσα συστήματα ακολούθησαν τα συστήματα που βασίζονται σε κανόνες. Για τη λειτουργία αυτών των συστημάτων απαιτούνται γλωσσικές πληροφορίες, οι οποίες περιέχονται σε λεξικά και γραμματικές. Οι γραμματικές, οι οποίες καταρτίζονται συνήθως από υπολογιστικούς γλωσσολόγους σε διάφορους γραμματικούς φορμαλισμούς, χρησιμεύουν στην ανάλυση και στην παραγωγή γραμματικών δομών [Fordyce, C. S., 2007]. Παράδειγμα τέτοιου συστήματος που χρησιμοποιείται ως τις μέρες μας αποτελεί το Systran.

[2] Οι δύο αυτές έννοιες αναλύονται με περισσότερη λεπτομέρεια στο Κεφάλαιο 3.2.

Ένα είδος συστήματος βασιζόμενου σε κανόνες είναι τα λεγόμενα συστήματα μεταφοράς (transfer), στα οποία η μεταφραστική διαδικασία πραγματοποιείται σε τρία στάδια: το στάδιο της συντακτικής και σημασιολογικής ανάλυσης, από την οποία δημιουργείται μια αφηρημένη αναπαράσταση της πρότασης-πηγή, το στάδιο της μεταφοράς, κατά το οποίο η αφηρημένη αναπαράσταση μετατρέπεται σε μια αναπαράσταση στη γλώσσα-στόχος, και το στάδιο της παραγωγής, κατά το οποίο παράγεται η τελική πρόταση στη γλώσσα-στόχος [Fordyce, C. S., 2007].

Άλλο είδος συστήματος βασιζόμενου σε κανόνες αποτελεί το διαγλωσσικό σύστημα (διαγλώσσα [interlingua] είναι μια ενδιάμεση τεχνητή γλώσσα), στο οποίο εμπεριέχονται μόνο το στάδιο της μετατροπής από την πρόταση-πηγή σε μια πρόταση στη διαγλώσσα και το στάδιο της μετατροπής από την πρόταση στη διαγλώσσα σε μια πρόταση στη γλώσσα-στόχος [Fordyce, C. S., 2007].

Τα παραπάνω αναπαρίστανται στο ακόλουθο τρίγωνο του Vauqois[3].



Εικόνα 1: Το τρίγωνο του Vauqois

2.3 Προσεγγίσεις βασιζόμενες σε εμπειρική γνώση (data-driven)

Οι νεότερες προσεγγίσεις της μηχανικής μετάφρασης, οι οποίες και χρησιμοποιούνται ως τις μέρες μας, δεν βασίζονται πλέον σε κανόνες αλλά σε μεγάλους όγκους δεδομένων, ήτοι κειμενικού υλικού. Τέτοιου είδους προσεγγίσεις είναι τα συστήματα με χρήση παραδειγμάτων (example-based) και τα στατιστικά συστήματα [Σοφιανόπουλος, Σ., 2009]. Στην παρούσα εργασία θα μας απασχολήσουν κυρίως τα στατιστικά συστήματα, καθώς το μεγαλύτερο μέρος

3 https://kuiwon.wordpress.com/2013/05/24/an-overview-of-automated-machine-translation-the-ruled-based-approach/ (τελευταία προσπέλαση 11/06/2015)

της αξιολόγησης αφορά την ποιότητα μεταφράσεων στατιστικών συστημάτων, τα οποία είναι και τα πιο διαδεδομένα.

2.3.1 Συστήματα με χρήση παραδειγμάτων (example-based)

Τα συστήματα αυτά λειτουργούν βάσει του εντοπισμού ομοιότητας μεταξύ του κειμένου προς μετάφραση και ενός κειμένου-παράδειγμα, όπου ως “παράδειγμα” νοείται ένα ζεύγος κειμένων, από τα οποία το ένα είναι μετάφραση του άλλου, και “κείμενο” μπορεί να είναι ένα κείμενο αποτελούμενο από μία λέξη μέχρι μία παράγραφο [Fordyce, C. S., 2007]. Το παράδειγμα (ζεύγος κειμένων από τα οποία το ένα είναι μετάφραση του άλλου) αποτελεί ένα παράλληλο κείμενο, έννοια η οποία θα μας απασχολήσει στην παρούσα εργασία, όπως και η έννοια του σώματος κειμένων. Ως σώμα κειμένων ορίζεται μια συλλογή κειμενικού υλικού με στόχο τον χαρακτηρισμό μιας κατάστασης ή μιας ποικιλίας της γλώσσας [Sinclair, J., 1991] ή μια συλλογή τμημάτων λόγου δομημένη βάσει συγκεκριμένων κριτηρίων [Sinclair, J., 1996]. Ένας από τους διαχωρισμούς που γίνονται ως προς τα σώματα κειμένων είναι η διάκριση μεταξύ μονόγλωσσου και παράλληλου σώματος. Μονόγλωσσο είναι το σώμα το οποίο αποτελείται από κείμενα σε μία γλώσσα, ενώ παράλληλο είναι το σώμα που απαρτίζεται από ένα σώμα πρωτότυπων κειμένων σε μία γλώσσα και ένα σώμα αποτελούμενο από τις μεταφράσεις αυτών των πρωτότυπων κειμένων σε άλλη γλώσσα[4].

Η μεταφραστική διαδικασία στα συστήματα με χρήση παραδειγμάτων περιλαμβάνει 4 στάδια: α) ανάκτηση παραδειγμάτων από δίγλωσσα παράλληλα σώματα, β) διαχείριση της βάσης παραδειγμάτων (αποθήκευση και συντήρηση παραδειγμάτων), γ) εφαρμογή παραδειγμάτων (αποσύνθεση του εισαγόμενου κειμένου σε παραδείγματα και μετατροπή των κειμένων-πηγή σε κείμενα-στόχος), και δ) σύνθεση της πρότασης-στόχος (τοποθέτηση μετατραπέντων παραδειγμάτων σε σειρά, ώστε να βελτιωθεί η αναγνωσιμότητα του κειμένου-στόχος μετά τη μετατροπή) [Fordyce, C. S., 2007].

2.3.2 Στατιστική μηχανική μετάφραση

Ένα σύστημα στατιστικής μηχανικής μετάφρασης απαρτίζεται από τρία μέρη: το μεταφραστικό μοντέλο, το γλωσσικό μοντέλο και τον αποκωδικοποιητή (decoder). Το μεταφραστικό και το γλωσσικό μοντέλο παράγονται από τα σώματα κειμένων με τα οποία εκπαιδεύεται το σύστημα, ενώ ο αποκωδικοποιητής αποτελεί τον αλγόριθμο της μεταφραστικής διαδικασίας.

---

4 Εκτενέστερη ανάλυση σχετικά με τα είδη σωμάτων κειμένων γίνεται στο Κεφάλαιο 4.1.

Το γλωσσικό μοντέλο ενός στατιστικού συστήματος αποτελείται από ένα μονόγλωσσο σώμα κειμένων στη γλώσσα-στόχος και στόχος του είναι η βελτιστοποίηση της γραμματικότητας του μεταφράσματος, κάτι το οποίο επιτυγχάνεται μέσω του υπολογισμού της πιθανότητας μιας πρότασης βάσει των μετρήσεων συχνότητας ν-γραμμάτων στο μονόγλωσσο σώμα. Καθώς υπάρχουν μη συχνά ν-γράμματα τα οποία θα έπρεπε να συμπεριληφθούν στη μετάφραση, χρησιμοποιούνται τεχνικές με τις οποίες αποφεύγεται η προκαταβολική απόρριψη μη συχνών ν-γραμμάτων [España i Bonet, C., 2010].

Το μεταφραστικό μοντέλο αποτελείται από στοιχισμένα παράλληλα σώματα κειμένων στη γλώσσα-πηγή και στη γλώσσα-στόχος. Στοίχιση (alignment) είναι η διαδικασία αντιστοίχισης των προτάσεων του πρωτότυπου κειμένου με τις προτάσεις του μεταφράσματος, η οποία στην ανάπτυξη στατιστικών συστημάτων επιτυγχάνεται μέσω μη επιβλεπόμενης μηχανικής μάθησης. Στόχος του μεταφραστικού μοντέλου είναι η βελτιστοποίηση του μεταφράσματος σε επίπεδο πιστότητας, η οποία επιτυγχάνεται μέσω του υπολογισμού της πιθανότητας μιας μετάφρασης από το παράλληλο σώμα κειμένων, αρχικά της μετάφρασης μιας λέξης και πλέον στα περισσότερα στατιστικά συστήματα της μετάφρασης μιας φράσης, φράσης όχι με τη γλωσσολογική έννοια του όρου αλλά νοούμενης ως μια ακολουθία λέξεων. Η ποιότητά του εξαρτάται από την ποιότητα της στοίχισης [España i Bonet, C., 2010] και του μεταφρασμένου μέρους του παράλληλου σώματος.

Ο αποκωδικοποιητής είναι ο αλγόριθμος με τον οποίο επιτελείται η αναζήτηση στον χώρο των πιθανών μεταφράσεων και η επιλογή της υπόθεσης με το χαμηλότερο κόστος [Koehn, Ph., 2006]. Όπως και στις προηγούμενες προσεγγίσεις, η μετάφραση εξακολουθεί να γίνεται σε επίπεδο πρότασης.

Παραδείγματα στατιστικών συστημάτων μετάφρασης είναι τα ευρέως χρησιμοποιούμενα Google Translate και Bing Translator.

2.4 Υβριδικά συστήματα

Τα υβριδικά συστήματα μηχανικής μετάφρασης άρχισαν να αναπτύσσονται με στόχο την αντιμετώπιση των αδυναμιών των διαφόρων υπαρχουσών τεχνικών και την αξιοποίηση των καλύτερων χαρακτηριστικών τους [Fordyce, C. S., 2007]. Μεγάλο μέρος της έρευνας σήμερα περιλαμβάνει κάποιο βαθμό υβριδοποίησης [Fordyce, C. S., 2007], δηλαδή συνδυασμό των ανωτέρω προσεγγίσεων μηχανικής μετάφρασης.

3 Αξιολόγηση ποιότητας μηχανικής μετάφρασης

Όπως κάθε σύστημα, τα συστήματα μηχανικής μετάφρασης χρειάζεται να αξιολογούνται, προκειμένου να γνωρίζουμε τον βαθμό αξιοπιστίας των αποτελεσμάτων τους αλλά και να μπορέσουμε να τα βελτιώσουμε [Bonnie, D. κ.ά., 2011]. Στη συγκεκριμένη εργασία ασχολούμαστε αποκλειστικά με την αξιολόγηση της ποιότητας των εξαγόμενων των συστημάτων μηχανικής μετάφρασης, των μεταφρασμάτων δηλαδή, και όχι με άλλες παραμέτρους αυτών, όπως η ταχύτητα, η χρηστικότητα (φιλικότητα προς τον χρήστη) ή η αποδοτικότητα (απαιτούμενη ποσότητα υπολογιστικών πόρων).

Παρόλο που, σύμφωνα με τον Ricoeur [2010], “δεν υφίσταται ένα απόλυτο κριτήριο το οποίο θα μας επέτρεπε να ορίσουμε την καλή μετάφραση”, έχουν διατυπωθεί διάφοροι ορισμοί της καλής ή της ποιοτικής μετάφρασης. Ένας πρόσφατος συνοπτικός ορισμός που χρησιμοποιείται στον χώρο είναι ο ορισμός των Koby και Melby [2013]: “Η ποιοτική μετάφραση (1) χαρακτηρίζεται από την απαιτούμενη πιστότητα και γραμματικότητα (2) για το κοινό και τον σκοπό της και (3) συμμορφώνεται με όλες τις άλλες συμφωνημένες προδιαγραφές λαμβάνοντας υπόψη τις ανάγκες του τελικού χρήστη”[5]. Η αναφορά στον σκοπό της μετάφρασης δεν είναι τυχαία. Σύμφωνα με τη θεωρία του Σκοπού [Vermeer, H. J., 2000], η μετάφραση είναι μια πράξη με ένα συγκεκριμένο σκοπό, αποτέλεσμα της οποίας είναι ένα κείμενο-στόχος. Ο σκοπός που διέπει τη διαδικασία της μετάφρασης καθορίζεται από μια οντότητα, τον εντολέα της, και η μετάφραση πρέπει να λειτουργήσει με τέτοιο τρόπο, ώστε να επιτευχθεί ο επιδιωκόμενος στόχος. Τα ίδια στοιχεία όμως που απαρτίζουν την ποιοτική μετάφραση σύμφωνα με τα παραπάνω εμφανίζονται ήδη από το πρώτο σημαντικό βιβλίο που γράφτηκε για τη μετάφραση, σύμφωνα με τον Newmark [1988], από τον Tytler το 1790, όπου όριζε την καλή μετάφραση ως εξής: μια μετάφραση στην οποία η αξία του πρωτότυπου έργου μεταδίδεται τόσο πλήρως σε μια άλλη γλώσσα ώστε ο φυσικός ομιλητής της γλώσσας αυτής να την αντιλαμβάνεται τόσο ξεκάθαρα και να τη νιώθει τόσο έντονα όσο ο φυσικός ομιλητής της γλώσσας του πρωτότυπου έργου. Και το 1982 ο Newmark δηλώνει, αναφερόμενος στον Rieu [1953], ότι στην εποχή του είναι ευρέως αν και όχι απόλυτα αποδεκτό ότι ο βασικός στόχος του μεταφραστή είναι να προκαλέσει όσο το δυνατόν πλησιέστερο αντίκτυπο στους αναγνώστες του με αυτόν που προκλήθηκε στους αναγνώστες του πρωτότυπου. Πρόκειται για την αρχή της ισοδυναμίας[6]. Όλα τα παραπάνω στοιχεία της πιστότητας, της γραμματικότητας, του κοινού στο οποίο απευθύνεται η μετάφραση, του σκοπού, αλλά και της ισοδυναμίας είναι χαρακτηριστικά που λαμβάνονται υπόψη στην αξιολόγηση ποιότητας της παρούσας εργασίας.

---

5 Σε δική μας μετάφραση.

6 Για μια συνοπτική παρουσίαση της αρχής της ισοδυναμίας και το σχετικό έργο των Roman Jacobson, Eugene A. Nida και John Cunnison Catford, βλ. Kanelli, O., 2010.

Για τη βελτίωση της ποιότητας των μηχανικών μεταφράσεων, η οποία βρίσκεται ακόμη σε αρκετά χαμηλά επίπεδα, κρίνεται επιτακτική η ύπαρξη ποιοτικής και αξιόπιστης αξιολόγησης, και πράγματι τα τελευταία χρόνια αυξάνονται ραγδαία οι προτεινόμενες μέθοδοι. Στη διεθνή βιβλιογραφία, όπου επικρατούν οι δημοσιεύσεις στην αγγλική, κάθε νέα μέθοδος έρχεται να ενταχθεί σε μία από τις δύο κατηγορίες με τις αγγλικές ονομασίες “automatic” και “human” ή “manual”. Για αυτές τις έννοιες θα χρησιμοποιήσουμε τις αποδόσεις “αυτόματη” και “μη αυτοματοποιημένη” αξιολόγηση. Εκ πρώτης όψεως, φαίνονται δύο αρκετά διακριτές κατηγορίες, οι οποίες σχεδόν αυτοπροσδιορίζονται, και κάθε αναγνώστης εύκολα αντιλαμβάνεται ότι στη μία περίπτωση πρόκειται για αξιολόγηση η οποία πραγματοποιείται αυτόματα, από κάποιο μηχάνημα δηλαδή, και στη δεύτερη για αξιολόγηση η οποία πραγματοποιείται από έναν ή περισσότερους ανθρώπους. Στην πράξη όμως αυτές οι δύο κατηγορίες δεν είναι τόσο διακριτές. Αφενός μεν, η “αυτόματη” αξιολόγηση μέχρι στιγμής χρησιμοποιεί είτε μεταφράσεις που έχουν γίνει από ανθρώπους είτε επισημειώσεις που έχουν επίσης πραγματοποιηθεί από ανθρώπους και αυτό που γίνεται αυτόματα είναι ένας υπολογισμός, αφετέρου δε, στη μη αυτοματοποιημένη αξιολόγηση χρησιμοποιούνται εκτεταμένα ποικίλα υπολογιστικά εργαλεία και αυτοματοποιημένες διαδικασίες. Ανάμεσα σε αυτές τις δύο κατηγορίες βρίσκονται οι αποκαλούμενες “semi-automated” (“ημι-αυτόματες”) μέθοδοι αξιολόγησης, οι οποίες στη δική μας ταξινόμηση υπάγονται στη μη αυτοματοποιημένη αξιολόγηση (Κεφάλαιο 3.2).

Η παρούσα εργασία διαρθρώνεται γύρω από τις δύο προαναφερθείσες κατηγορίες, την αυτόματη και τη μη αυτοματοποιημένη. Στην αυτόματη υπάγονται αυτοματοποιημένες τεχνικές στις οποίες η ανθρώπινη επέμβαση περιορίζεται στην ανάπτυξη του συστήματος και δεν λαμβάνει χώρα στο στάδιο της εκτέλεσης της αξιολόγησης. Αντίθετα, στη μη αυτοματοποιημένη υπάγονται οι τεχνικές αυτές στις οποίες ο άνθρωπος επεμβαίνει με μη αυτοματοποιημένη εργασία (και) στην εκτέλεση της αξιολόγησης, γι' αυτό και συγκαταλέγονται σε αυτές και οι ημι-αυτόματες μετρικές.

Η κατηγορία της μη αυτοματοποιημένης αξιολόγησης χωρίζεται στην παρούσα εργασία σε δύο περαιτέρω υποκατηγορίες με βάση το εξής κριτήριο: την ύπαρξη άμεσα εκπεφρασμένης αξιολογικής κρίσης[7] ή μη. Η διάκριση αυτή προκύπτει από την ουσιαστική διαφορά που υπάρχει μεταξύ μιας μεθόδου στην οποία ένας ή περισσότεροι άνθρωποι δηλώνουν ότι μια μετάφραση είναι πολύ καλή, μέτρια ή κακή, ή ότι η γλώσσα σε μια μετάφραση είναι πολύ καλή, μέτρια ή κακή, και μιας μεθόδου στην οποία ένας ή περισσότεροι άνθρωποι καλούνται να ταξινομήσουν ή/και να διορθώσουν λάθη σε μια μετάφραση. Η ανάλυση και καταμέτρηση λαθών δεν ενέχει κάποια άμεση αξιολογική κρίση εκ μέρους των διορθωτών, το μοναδικό αξιολογικό στοιχείο στη διόρθωση, στην ίδια την ανάθεση της διόρθωσης δηλαδή, είναι η υπόνοια ότι αυτή η μετάφραση

[7] Στην αξιολογική κρίση θεωρούμε ότι υπάρχει υποκειμενικός χαρακτηρισμός ενός φαινομένου και έκφραση γνώμης [Βώρος, Φ. Κ., 2004].

δεν είναι τέλεια, κάτι το οποίο θεωρείται δεδομένο στον χώρο της μετάφρασης. Ο Ricoeur [2010] μιλάει για την "απάρνηση του ιδανικού της τέλειας μετάφρασης", για το "πένθος για την ανυπαρξία της απόλυτης μετάφρασης" και για "αντιστοιχία δίχως πλήρη αναλογία". Ο Newmark [1988], ορίζοντας τη μετάφραση, αναφέρει: "Η μετάφραση είναι μια τέχνη[8] που συνίσταται στην *απόπειρα* αντικατάστασης ενός γραπτού μηνύματος ή/και δήλωσης σε μια γλώσσα με το ίδιο μήνυμα ή/και δήλωση σε μια άλλη γλώσσα. Κάθε άσκηση περιλαμβάνει κάποιου είδους απώλεια σε σημασία, η οποία οφείλεται σε διάφορους παράγοντες [...]. Η βασική απώλεια βρίσκεται σε ένα συνεχές μεταξύ της υπερμετάφρασης (υπερβολική λεπτομέρεια) και της υπομετάφρασης (υπερβολική γενίκευση)"[9]. Δεν είναι λίγες οι κριτικές κατά της μετάφρασης και των μεταφραστών και άλλωστε ο ίδιος ο Newmark [1988] αναφέρει τα λεγόμενα του Widmer [1959]: "Ο βασικός όμως λόγος για τη διατύπωση μιας θεωρίας της μετάφρασης, για την πρόταση μεθόδων μετάφρασης που να σχετίζονται με αυτήν και να απορρέουν από αυτήν, για τη διδασκαλία μετάφρασης, για τα μαθήματα μετάφρασης, είναι η αποκρουστικά κακή ποιότητα τόσων πολλών δημοσιευμένων μεταφράσεων"[10]. Και ο Newmark συμπληρώνει: "Είτε πρόκειται για λογοτεχνική μετάφραση είτε όχι, σπανίζουν οι μεταφράσεις χωρίς λάθη"[11]. Βέβαια, αυτή η αποδεδειγμένη "αδυναμία" ή "αναπηρία" του μεταφράσματος δεν έχει ποτέ αποτρέψει τον άνθρωπο από την πράξη της μετάφρασης. Όπως καταλήγει και ο Ricoeur [2010], "θέλοντας και μη μεταφερόμαστε στην αντίπερα όχθη: δεδομένου ότι η μετάφραση υφίσταται ως πρακτική, πρέπει ως εκ τούτου να υπάρχει η δυνατότητα ύπαρξής της".

Η διαφορά μεταξύ της ύπαρξης άμεσα εκπεφρασμένης αξιολογικής κρίσης και μη είναι πολύ σημαντική, γιατί έχει επιπτώσεις στον βαθμό αντικειμενικότητας της αξιολόγησης, γιατί προϋποθέτει διαφορετικές νοητικές διαδικασίες αλλά και γιατί η καθεμία από αυτές τις μεθόδους επιτελεί διαφορετικούς σκοπούς.

3.1 Αυτόματη αξιολόγηση ποιότητας μηχανικής μετάφρασης

Ως αυτόματη αξιολόγηση μηχανικής μετάφρασης θεωρούμε την αξιολόγηση που επιτελείται μέσω συστημάτων στα οποία ο ανθρώπινος παράγοντας δεν επεμβαίνει στο στάδιο της αξιολόγησης αυτής καθαυτήν αλλά η όποια ανθρώπινη επέμβαση λαμβάνει χώρα κατά την ανάπτυξη του συστήματος. Σε όλα τα υπάρχοντα μέχρι στιγμής συστήματα γίνεται χρήση ανθρώπινης εργασίας, όταν όμως αυτή αφορά τη συλλογή δεδομένων ή επισημείωση που θα υποστεί επεξεργασία από

8 Η τέχνη νοούμενη ως "ανθρώπινη δραστηριότητα που στηρίζεται σε ορισμένες γνώσεις και εμπειρίες και που έχει ως σκοπό τη δημιουργία ενός πνευματικού ή τεχνικού έργου" (Λεξικό της Κοινής Νεοελληνικής Τριανταφυλλίδη).

9 Δική μας η πλάγια γραφή και η μετάφραση.

10 Σε δική μας μετάφραση από την αγγλική.

11 Σε δική μας μετάφραση.

το σύστημα, τότε αυτό το σύστημα κατατάσσεται στα αυτόματα συστήματα αξιολόγησης.

Χρησιμοποιούνται δύο ειδών συστήματα αυτόματης αξιολόγησης: α) μετρικές που αποδίδουν βαθμολογία σε μηχανικά μεταφράσματα με βάση τον βαθμό ομοιότητάς τους με μεταφράσεις-αναφορά, και β) μετρικές εκτίμησης ποιότητας (Confidence ή Quality Estimation), συστήματα δηλαδή ταξινόμησης των μεταφρασμάτων σε επίπεδα ποιότητας.

Όπως κάθε άλλο σύστημα, έτσι και τα συστήματα αυτόματης αξιολόγησης πρέπει να μπορούν να αξιολογηθούν με συγκεκριμένα κριτήρια, τα οποία αρχικά διατυπώθηκαν για τις μετρικές με μεταφράσεις-αναφορά και ακολουθήθηκαν έπειτα στην πράξη και στις μετρικές εκτίμησης ποιότητας. Σύμφωνα με τους Banerjee, S. & Lavie, A. [2005], ένα επιθυμητό σύστημα αυτόματης αξιολόγησης πρέπει να πληροί τις ακόλουθες προϋποθεσεις: α) υψηλή συσχέτιση με ποσοτικοποιημένες ανθρώπινες κρίσεις σχετικά με την ποιότητα της μετάφρασης, β) ευαισθησία σε μικρές διαφορές ποιότητας μεταξύ συστημάτων και του ίδιου συστήματος σε διαφορετικά στάδια ανάπτυξης, γ) συνέπεια στα αποτελέσματα (παρεμφερή αποτελέσματα για παρεμφερή κείμενα στο ίδιο μεταφραστικό σύστημα), δ) αξιοπιστία (βεβαιότητα ανταπόκρισης μεταξύ αποτελεσμάτων αξιολόγησης και επιδόσεων), ε) ευρύτητα φάσματος πεδίων, και στ) ταχύτητα και ευκολία στη χρήση. Οι προϋποθέσεις αυτές συμπληρώνονται από αυτές που προτείνει ο Koehn, Ph. [2010] με κάποια μικρή επικάλυψη: α) χαμηλό κόστος, β) δυνατότητα άμεσης βελτιστοποίησης επιδόσεων συστήματος προς τη μετρική, γ) διαισθητικά εφικτή ερμηνεία της βαθμολογίας, δ) συνέπεια, και ε) ορθότητα. Από όλες τις παραπάνω προϋποθέσεις η πιο βασική θεωρείται η συσχέτιση με την ανθρώπινη κρίση [Specia, L. κ.ά., 2010].

3.1.1 Μετρικές με χρήση μεταφράσεων-αναφορά

Αυτές οι μετρικές αποδίδουν βαθμολογία σε μηχανικά μεταφράσματα με βάση τον βαθμό ομοιότητάς τους με μεταφράσεις-αναφορά, ποιοτικά μεταφράσματα δηλαδή που έχουν παραχθεί από ανθρώπους [Papineni, K. κ.ά., 2002]. Για τον υπολογισμό αυτού του βαθμού ομοιότητας χρησιμοποιούνται διάφορες τεχνικές και συνδυασμοί αυτών. Βάσει της κάθε τεχνικής έχουν αναπτυχθεί αρκετές υλοποιήσεις. Η πληθώρα των μετρικών αλλά και των συνδυασμών τεχνικών που έχουν προταθεί προσφέρονται για πολλών ειδών ταξινομήσεις. Για τους σκοπούς της παρούσας εργασίας προβήκαμε σε μια γενική ταξινόμηση και στην αναφορά των υλοποιήσεων με τη μεγαλύτερη παρουσία στη διεθνή βιβλιογραφία.

3.1.1.1 Απόσταση διόρθωσης (edit distance)

Μια από τις τεχνικές που χρησιμοποιούνται σε μετρικές αυτόματης αξιολόγησης είναι η μέθοδος που βασίζεται στην απόσταση διόρθωσης και συγκεκριμένα την απόσταση κατά Levenshtein [Levenshtein, V., 1966]. Ως απόσταση διόρθωσης μεταξύ δύο συμβολοσειρών α και β νοείται ο ελάχιστος αριθμός ενεργειών που χρειάζονται ώστε να μετατραπεί η συμβολοσειρά α σ ε β [Navarro, G., 2001]. Ως ενέργειες διόρθωσης θεωρούνται η εισαγωγή, η απαλοιφή και η αντικατάσταση ενός χαρακτήρα με κάποιον άλλο [Levenshtein, V., 1966]. Στις υλοποιήσεις για αξιολόγηση μετάφρασης οι ενέργειες διόρθωσης είναι τα βήματα που πρέπει να γίνουν ώστε από το μηχανικό μετάφρασμα να καταλήξουμε στη μετάφραση-αναφορά, και δεν αφορούν χαρακτήρες αλλά λέξεις [Fordyce, C. S., 2007].

Υλοποιήσεις αυτής της μετρικής στην αξιολόγηση μηχανικής μετάφρασης είναι ο WER (Word Error Rate-Δείκτης Σφαλμάτων Λέξης) [Niessen, S. κ.ά., 2000] και πολλές παραλλαγές καθώς και επεκτάσεις του. Ο δείκτης σφαλμάτων λέξης είναι ο λόγος του αθροίσματος των ενεργειών διόρθωσης σε ένα μηχανικό μετάφρασμα προς τον αριθμό λέξεων της μετάφρασης-αναφορά [Niessen, S. κ.ά., 2000].

Κάποιες από τις παραλλαγές και επεκτάσεις του WER είναι οι εξής: WERg (για τη χρήση του βλ. Blatz, J. κ.ά., 2004), TER (Translation Edit Rate) [Snover, M. κ.ά., 2006], MWER (Multiple Reference WER) [Niessen, S. κ.ά., 2000], invWER (Inversion WER) [Leusch κ.ά., 2003], SER (Sentence Error Rate) (για τη χρήση του βλ. Tomás κ.ά., 2003), CDER (Cover Disjoint Error Rate) [Leusch κ.ά., 2006], HyTER (Hybrid TER) [Dreyer, M. και Marcu, D., 2012] και PER (Position-Independent WER), ο οποίος καταχρηστικά συγκαταλέγεται εδώ, διότι ενώ είναι παραλλαγή του WER, δεν μπορεί να θεωρηθεί ότι κάνει χρήση απόστασης [Leusch, G. κ.ά., 2003]. Τέλος, έχουν χρησιμοποιηθεί οι WER και PER για διάφορες κατηγορίες μερών του λόγου [Popović, M. και Ney, H., 2007]. Σε αυτές τις παραλλαγές και επεκτάσεις οι διαφορές έγκεινται κυρίως στη χρήση μίας ή περισσότερων μεταφράσεων-αναφορά και στη χρήση ή μη της αντιστροφής λεξικών μονάδων/συνόλων ως επιπλέον ενέργεια διόρθωσης. Συγκαταλέγεται σε αυτή την κατηγορία και η μετρική RED, η οποία χρησιμοποιεί την απόσταση διόρθωσης, όχι όμως για να υπολογίσει βαθμολογία (score) αλλά για να κατατάξει τις μεταφράσεις (ranking) [Akiba κ.ά., 2001].

3.1.1.2 Ακρίβεια και ανάκληση

Άλλες τεχνικές που χρησιμοποιούνται σε μετρικές αυτόματης αξιολόγησης είναι η ακρίβεια (precision) και η ανάκληση (recall). Η ακρίβεια και η ανάκληση είναι μετρικές που χρησιμοποιούνται ευρέως, μεταξύ άλλων, στην ανάκτηση πληροφοριών, και έχουν αξιοποιηθεί και στην αξιολόγηση μηχανικής μετάφρασης [Lavie, A., 2011, Koehn, P., 2010]. Στην περίπτωση της μετάφρασης, η ακρίβεια είναι ο λόγος των αποδεκτών ν-γραμμάτων της μηχανικής μετάφρασης (αυτών δηλαδή που έχουν εντοπιστεί και σε κάποια μετάφραση-αναφορά) προς τον αριθμό των ν-γραμμάτων της μηχανικής μετάφρασης, δηλαδή σε μια πρόταση με 10 1-γράμματα, εκ των οποίων είναι αποδεκτά τα 6, η ακρίβεια είναι 6/10. Η ανάκληση είναι ο λόγος των αποδεκτών ν-γραμμάτων της μηχανικής μετάφρασης (αυτών δηλαδή που έχουν εντοπιστεί και σε κάποια μετάφραση-αναφορά) προς τον αριθμό των ν-γραμμάτων της μετάφρασης-αναφορά (του ιδανικού δηλαδή αριθμού ν-γραμμάτων). Στη μία περίπτωση δηλαδή μετράμε τι ποσοστό από τις παραχθείσες λέξεις είναι σωστό και στην άλλη πόσες από τις επιθυμητές λέξεις παράχθηκαν. Στο παραπάνω παράδειγμα ακρίβειας 6/10, εάν στη μετάφραση-αναφορά έχουμε 15 1-γράμματα (η μηχανική μετάφραση δηλαδή είναι αρκετά μικρότερη από την ιδανική), η ανάκληση θα είναι 6/15.

Οι ευρύτερα χρησιμοποιούμενες υλοποιήσεις για την αξιολόγηση μηχανικής μετάφρασης με χρήση ακρίβειας είναι το BLEU (Bilingual Evaluation Understudy) [Papineni, K. κ.ά., 2002] και η παραλλαγή του που υλοποιήθηκε από το NIST[12] [Doddington, G., 2002].

Συνδυασμός ακρίβειας και ανάκλησης χρησιμοποιείται από τη μέθοδο υπολογισμού με χρήση του αρμονικού μέσου όρου ακρίβειας και ανάκλησης (F-measure) [Melamed κ.ά., 2003], από τις διάφορες υλοποιήσεις ROUGE (Recall-Oriented Understudy for Gisting Evaluation) [Lin, Ch. Y. και Och, F. J., 2004], οι οποίες αποτελούν προσαρμογές του BLEU για αξιολόγηση αυτόματων περιλήψεων, καθώς και από τις μετρικές GTM (General Text Matcher) [Melamed κ.ά., 2003], WNM (Weighted N-gram Model) [Babych και Hartley, 2004], METEOR (Metric for Evaluation of Translation with Explicit Ordering) [Banerjee, S. και Lavie, A., 2005], ParaEval [Zhou κ.ά., 2006], καθώς και από την οικογένεια μετρικών BLANC (Broad Learning and Adaptation for Numeric Criteria) [Lita L.V. κ.ά., 2005] και την πιο πρόσφατη μετρική LEPOR [Han, A.L.F. κ.ά., 2012].

3.1.1.3 Άλλες μετρικές

Αξίζει να αναφερθεί ότι έχουν προταθεί και μετρικές που δεν μένουν στην ομοιότητα σε επιφανειακό επίπεδο αλλά κάνουν χρήση και συντακτικών δέντρων εξάρτησης [Amigo κ.ά., 2006, Koehn και Monz, 2006] αλλά και άλλων στοιχείων γλωσσικής ανάλυσης, όπως σημασιολογικοί ρόλοι, σχέσεις εξάρτησης, συντακτικές δομές, ονοματισμένες οντότητες (named entities), κ.λπ. [Giménez και Màrquez, 2007].

[12] National Institute of Standards and Technology – Εθνικό Ίδρυμα Προτύπων και Τεχνολογίας του Υπουργείου Εμπορίου των Ηνωμένων Πολιτειών Αμερικής

3.1.1.4 BLEU

Στην παρούσα εργασία χρησιμοποιείται η μετρική BLEU [Papineni, K. κ.ά., 2002], η οποία παρουσιάστηκε για πρώτη φορά το 2001 [Fordyce, C. S., 2007] και προτάθηκε για να καλυφθούν οι ανάγκες γρήγορης και οικονομικής αξιολόγησης συστημάτων μηχανικής μετάφρασης για οποιοδήποτε ζεύγος γλωσσών τόσο για γενικότερη χρήση όσο και για το στάδιο ανάπτυξης αυτών των συστημάτων, κατά το οποίο χρειάζεται συνεχής ανατροφοδότηση για τη βελτίωσή τους.

Με τη μέτρηση του BLEU, το μετάφρασμα αξιολογείται ως προς την πιστότητα και ως προς γραμματικότητα στη φυσική γλώσσα-στόχος μέσω της σύγκρισής του με μεταφράσεις-αναφορά [Papineni, K. κ.ά., 2002]. Καθώς θεωρείται ότι η ποιότητα της μηχανικής μετάφρασης κρίνεται με γνώμονα την εγγύτητά της στη μετάφραση που παράγεται από ανθρώπους, υπολογίζεται ο βαθμός ομοιότητας του μεταφράσματος προς αξιολόγηση με τις μεταφράσεις-αναφορά και θεωρείται ότι όσο μεγαλύτερη είναι αυτή τόσο καλύτερη είναι και η μηχανική μετάφραση, κάτι το οποίο εκφράζεται με μια βαθμολογία στην κλίμακα από το 0 μέχρι το 1, όπου το 0 είναι η χαμηλότερη δυνατή βαθμολογία [Papineni, K. κ.ά., 2002].

Για να επιτευχθεί αυτή η σύγκριση, χρειάζεται αφενός ο αλγόριθμος με τον οποίο υπολογίζεται ο βαθμός ομοιότητας και αφετέρου οι μεταφράσεις-αναφορά με τις οποίες θα συγκριθεί η μηχανική μετάφραση προς αξιολόγηση.

Για να υπολογιστεί ο βαθμός ομοιότητας συγκρίνονται ν-γράμματα, από 1- έως 4-γράμματα[13]. Ο υπολογισμός του βαθμού ομοιότητας βασίζεται, όπως προαναφέρθηκε, στη μετρική της ακρίβειας, η οποία στην περίπτωση του BLEU ορίζεται ως εξής: ο λόγος του αριθμού των ν-γραμμάτων της μηχανικής μετάφρασης που συναντώνται σε οποιαδήποτε από τις μεταφράσεις-αναφορά προς τον συνολικό αριθμό των ν-γραμμάτων της μηχανικής μετάφρασης, δηλαδή ο λόγος των “σωστών” ν-γραμμάτων προς τον συνολικό αριθμό ν-γραμμάτων (πόσες λέξεις σε μία μεταφρασμένη πρόταση είναι σωστές)[14]. Αυτή είναι η βασική ιδέα στην οποία βασίζεται η μετρική. Αυτό όμως που χρησιμοποιείται στην πραγματικότητα δεν είναι η ακρίβεια αλλά η τροποποιημένη ακρίβεια, η οποία εξασφαλίζει την αποφυγή υψηλής βαθμολογίας σε ένα μετάφρασμα με υπερβολικά μεγάλο αριθμό των ίδιων λέξεων οι οποίες υπάρχουν στις μεταφράσεις-αναφορά, φαινόμενο που παρατηρείται στις μηχανικές μεταφράσεις[15]. Η

[13] Ο αριθμός 4 θεωρείται ότι αποδίδει τη μεγαλύτερη συσχέτιση με την αξιολόγηση από μονόγλωσσους κριτές [Papineni, K. κ.ά., 2002].

[14] Η αρχική μέτρηση γίνεται ανά πρόταση και έπειτα υπολογίζεται για όλο το σύνολο προτάσεων.

[15] Στο παράδειγμα του μεταφράσματος “the the the the the the the”, εάν υπάρχει μια μετάφραση-αναφορά “the cat is on the mat”, εάν υπολογίσουμε την ακρίβεια, το 1-γραμμα “the” υπάρχει στη

τροποποιημένη ακρίβεια είναι ο λόγος του μέγιστου αριθμού των ν-γραμμάτων της μηχανικής μετάφρασης που συναντάται σε μία από τις μεταφράσεις-αναφορά προς τον συνολικό αριθμό των ν-γραμμάτων της μηχανικής μετάφρασης. Αντί δηλαδή να μετρήσουμε τον αριθμό των "σωστών" ν-γραμμάτων, μετράμε τον μέγιστο αριθμό που τα εν λόγω ν-γράμματα συναντώνται σε κάποια από τις μεταφράσεις-αναφορά[16].

Ο υπολογισμός της τροποποιημένης ακρίβειας εκτελείται για τα 1-γράμματα, τα 2-γράμματα, τα 3-γράμματα και τα 4-γράμματα, και με αυτόν τον τρόπο θεωρείται ότι αξιολογείται η πιστότητα (σε επίπεδο λέξης) και η γραμματικότητα (σε επίπεδο μεγαλύτερων ν-γραμμάτων) [Papineni, K. κ.ά., 2002].

Αυτό που δεν καλύπτεται με την περιγραφείσα μέχρι στιγμής μέθοδο είναι η ανάκληση, η οποία είναι ιδιαίτερα περίπλοκο να υπολογιστεί, εφόσον υπάρχουν πολλές μεταφράσεις-αναφορά (η "σωστή" απάντηση δηλαδή δεν είναι μία) [Papineni, K. κ.ά., 2002]. Αυτό έχει ως αποτέλεσμα να μη λαμβάνεται υπόψη αρνητικά η περίπτωση το μετάφρασμα να είναι υπερβολικά μικρό σε σχέση με τα μεταφράσματα των ανθρώπων (όπως στο προαναφερθέν παράδειγμα, όπου η ακρίβεια ήταν 6/10 αλλά η ανάκληση 6/15, γιατί το μηχανικό μετάφρασμα ήταν αρκετά μικρότερο από τη μετάφραση-αναφορά). Γι' αυτόν τον λόγο τα αποτελέσματα πολλαπλασιάζονται με μια ποινή για μικρές προτάσεις, τη λεγόμενη "ποινή συντομίας" (brevity penalty). Για να υπολογιστεί αυτή η ποινή, εντοπίζεται το πιο κοντινό σε αριθμό ν-γραμμάτων μήκος πρότασης από τις μεταφράσεις-αναφορά, το "μήκος βέλτιστης σύμπτωσης" (best match length), προστίθενται τα μήκη βέλτιστης σύμπτωσης για κάθε πρόταση της μηχανικής μετάφρασης (από όπου προκύπτει το μήκος αναφοράς) και έπειτα ορίζεται η ποινή συντομίας. Όταν το συνολικό μήκος του μηχανικά μεταφρασμένου κειμένου είναι μεγαλύτερο από το μήκος αναφοράς, η ποινή ορίζεται ότι ισούται με 1, ενώ όταν είναι μικρότερο ή ίσο του μήκους αναφοράς, υπολογίζεται ποινή με εκθετική μείωση, η οποία πολλαπλασιάζεται με τον γεωμετρικό μέσο[17] των επιμέρους βαθμολογιών των ν-γραμμάτων και το γινόμενο αυτών αποτελεί την τελική βαθμολογία [Papineni, K. κ.ά., 2002].

Όσο αφορά τη βαθμολογία του BLEU, όπως προαναφέρθηκε, κυμαίνεται από το 0 εώς το 1 και όσο καλύτερη είναι μια μετάφραση τόσο μεγαλύτερος είναι και ο βαθμός της. Συγκεκριμένα, μια

μετάφραση-αναφορά, συνεπώς είναι σωστό, άρα ο αριθμός των σωστών 1-γραμμάτων είναι 7 και ο συνολικός αριθμός των 1-γραμμάτων είναι 7, με αποτέλεσμα να έχουμε ακρίβεια 7/7, πράγμα το οποίο διαισθητικά και μόνο αντιλαμβανόμαστε ότι δεν μας αρκεί για να αξιολογήσουμε τη μετάφραση [Papineni, K. κ.ά., 2002].

16 Με αυτόν τον τρόπο, στο παράδειγμα του μεταφράσματος "the the the the the the the", εάν υπάρχει μια μετάφραση-αναφορά "The cat is on the mat", ο μέγιστος αριθμός που συναντάται σε μετάφραση-αναφορά το 1-γραμμα "the" είναι 2, οπότε η τροποποιημένη ακρίβεια είναι 2/7 [Papineni, K. κ.ά., 2002].

17 Ο γεωμετρικός μέσος είναι η νιοστή ρίζα του γινομένου ν αριθμών. Π.χ. ο γεωμετρικός μέσος των αριθμών 2 και 8 είναι η τετραγωνική ρίζα (γιατί έχουμε 2 αριθμούς) του γινομένου 2 x 8 ($\sqrt{2 \times 8} = 4$).

μετάφραση με βαθμό τουλάχιστον 0,3 θεωρείται κατανοητή και με βαθμό τουλάχιστον 0,5 θεωρείται καλή και ο λόγος της φυσικός [Lavie, A., 2011]. Η ανώτατη δυνατή βαθμολογία, το 1, δηλώνει πλήρη ομοιότητα με τις μεταφράσεις-αναφορά, πράγμα το οποίο σημαίνει ότι ούτε μια μετάφραση από άνθρωπο δεν θα έφτανε αυτόν τον βαθμό. Επίσης όσο περισσότερες μεταφράσεις-αναφορά υπάρχουν, τόσο υψηλότερη είναι και η βαθμολογία, γι' αυτό και δεν συνιστάται η σύγκριση μετρήσεων με διαφορετικό αριθμό μεταφράσεων-αναφορά [Papineni, K. κ.ά., 2002]. Τέλος, έχει παρατηρηθεί ότι εάν το σώμα κειμένων είναι μεγάλο, μπορεί να χρησιμοποιηθεί μόνο μία μετάφραση-αναφορά για κάθε κείμενο, εφόσον δεν έχουν παραχθεί όλες από τον ίδιο μεταφραστή και έχει έτσι εξασφαλιστεί η ποικιλία ύφους [Papineni, K. κ.ά., 2002].

Το βασικό κριτήριο αξιολόγησης ενός τέτοιου συστήματος μέτρησης, όπως προαναφέρθηκε, είναι ο βαθμός συσχέτισης των αποτελεσμάτων του με τα αποτελέσματα από κριτές [Doddington, G., 2002, Koehn, P., 2010, Banerjee, S. και Lavie, A., 2005]. Το BLEU, μια από τις “de facto” μετρικές [Specia, L. κ.ά., 2010], έχει επιτύχει καλά αποτελέσματα σε τέτοιου είδους συσχετίσεις [Papineni, K. κ.ά., 2002, Coughlin, D., 2003, Doddington, G., 2002], είναι ευρέως αποδεκτό και χρησιμοποιούμενο [Fordyce, C. S., 2007, Snover, M. κ.ά., 2006] και τα πλεονεκτήματά του αναγνωρίζονται ακόμη και από όσους ασκούν κριτική εναντίον του [Callison-Burch, C. κ.ά., 2006]. Πέρα από τον υψηλό βαθμό συσχέτισης με αξιολογήσεις από κριτές, ακόμη και σε περιπτώσεις όπου χρησιμοποιήθηκε μόνο μία μετάφραση-αναφορά [Coughlin, D., 2003], παρουσιάζει τα εξής πλεονεκτήματα: δεν επηρεάζεται η απόδοσή του από το μέγεθος του σώματος κειμένων [Coughlin, D., 2003], μπορεί να χρησιμοποιηθεί για ρύθμιση των παραμέτρων του συστήματος κατά την ανάπτυξή του (tuning) [Lavie, A., 2011], αλλά και για την αξιολόγηση της ποιότητας φυσικής γλώσσας σε άλλα συστήματα παραγωγής φυσικής γλώσσας [Papineni, K. κ.ά., 2002].

Ως προς τις αδυναμίες του συστήματος, από τις πιο συχνά αναφερόμενες είναι οι χαμηλότερες επιδόσεις σε επίπεδο πρότασης [Specia, L. κ.ά., 2010][18], ο εντοπισμός και η σύγκριση της ίδιας ακριβώς μορφής της κάθε λέξης χωρίς στελέχωση/λημματοποίηση[19] ή χρήση συνωνύμων και

[18] Ωστόσο, σύμφωνα με τους Papineni, K. κ.ά. [2002], αυτό γίνεται σκόπιμα για να αποφευχθούν πιθανά λάθη σε βαθμολογίες επιμέρους προτάσεων, και συγκεκριμένα αναφέρουν: “ποιότητα χάρις στην ποσότητα” (“quantity leads to quality”).

[19] Η στελέχωση (stemming) είναι η αποκοπή μερών λέξεων, συνήθως επιθημάτων. Με την πρακτική της στελέχωσης το εναπομείναν μέρος της λέξης δεν συμπίπτει απαραίτητα με το θέμα ή τη ρίζα της λέξης. Η λημματοποίηση (lemmatisation) είναι η διαδικασία ομαδοποίησης λέξεων σε λήμματα (το λήμμα είναι “κάθε λέξη που περιλαμβάνεται και σχολιάζεται μέσα σ' ένα λεξικό· εμφανίζεται με τον βασικό της τύπο (ονομαστική ενικού για ουσιαστικά, α΄ πρόσ. ενικού οριστικής ενεργητικού ενεστ. για ρήματα κ.ο.κ.) και εξαίρεται συνήθως με έντονα (μαύρα) γράμματα”, σύμφωνα με το Λεξικό της νέας ελληνικής γλώσσας του Γ. Μπαμπινιώτη). Η λημματοποίηση συνήθως χειρίζεται τις κλιτικές μορφές και όχι τα παράγωγα [Manning, C. D. κ.ά., 2008].

παραφράσεων [Callison-Burch, C. κ.ά., 2006, Lavie, A., 2011] πέρα από αυτές που υπάρχουν στις μεταφράσεις-αναφορά [Callison-Burch, C. κ.ά., 2006][20], καθώς και το γεγονός ότι στον υπολογισμό της βαθμολογίας χρησιμοποιείται το ίδιο βάρος σε κάθε ν-γραμμα, με αποτέλεσμα να αντιμετωπίζονται με τον ίδιο τρόπο τόσο οι λεξικές μονάδες που φέρουν υψηλού βαθμού σημασία όσο και οι υπόλοιπες [Callison-Burch, C. κ.ά., 2006, Doddington, G., 2002, Lavie, A., 2011]. Επίσης το γεγονός ότι σε πολύ χαμηλού επιπέδου ή πολύ ελεύθερες μεταφράσεις, όπου δεν συμπίπτουν 3-γράμματα και 4-γράμματα με τις μεταφράσεις-αναφορά, δεν γίνεται διάκριση μεταξύ των πολύ χαμηλών βαθμολογιών [Coughlin, D., 2003]. Μειονέκτημα θεωρείται και το ότι η γραμματικότητα μετράται με τα μεγάλα ν-γράμματα αντί για κάποιον άλλον πιο άμεσο τρόπο [Banerjee, S. και Lavie, A., 2005]. Άλλες κριτικές που έχουν αναφερθεί είναι η χρήση του γεωμετρικού μέσου [Doddington, G., 2002], διότι με αυτήν, καθώς πρόκειται για γινόμενο, εάν υπάρχει μηδενική τιμή, μηδενίζεται και το γινόμενο, δηλαδή ο τελικός βαθμός [Lavie, A., 2011, Banerjee, S. και Lavie, A., 2005, Specia, L. κ.ά., 2010], αλλά και οι χαμηλές επιδόσεις σε μικρά κείμενα [Fordyce, C. S., 2007] και σε συγκρίσεις μεταξύ μεταφράσεων από ανθρώπους και από μηχανικά συστήματα [Fordyce, C. S., 2007][21]. Από τις σοβαρότερες όμως κριτικές είναι η έλλειψη χρήσης της ανάκλησης, η οποία θεωρείται σημαντική [Lavie, A., 2011, Callison-Burch, C. κ.ά., 2006, S. Banerjee και A. Lavie, 2005]. Τίθεται υπό αμφισβήτηση το κατά πόσο αυτή η έλλειψη αντισταθμίζεται κατάλληλα από την ποινή συντομίας [Lavie, A., 2011] και, σύμφωνα με τους Banerjee, S. και Lavie, A. [2005], αποδεικνύεται ότι δεν αντισταθμίζεται. Η ανάκληση θεωρείται σημαντικότερη από την ακρίβεια για την επίτευξη υψηλότερης συσχέτισης με κρίσεις ανθρώπων, καθώς εκφράζει τον βαθμό στον οποίο μεταφέρεται το περιεχόμενο του πρωτότυπου στο μετάφρασμα [Banerjee, S. και Lavie, A., 2005]. Σύμφωνα με άλλες κριτικές, το BLEU παρουσιάζει χαμηλές επιδόσεις σε συγκρίσεις διαφορετικών ειδών συστημάτων μηχανικής μετάφρασης, ιδίως αν πρόκειται για σύγκριση μεταξύ στατιστικών συστημάτων με συστήματα που κάνουν χρήση κανόνων [Fordyce, C. S., 2007, Coughlin, D., 2003]. Επιπλέον, σε έρευνα που πραγματοποιήθηκε με στόχο την αξιολόγηση του BLEU [Callison-Burch, C. κ.ά., 2006] αναδεικνύονται κάποιες από τις αδυναμίες αυτής της μετρικής και σε θεωρητική βάση, σύμφωνα με την οποία ο επιτρεπόμενος αριθμός παραλλαγών στη σειρά των ν-γραμμάτων είναι υπερβολικά μεγάλος, με αποτέλεσμα να θεωρείται απίθανο ένας τόσο μεγάλος αριθμός επιλογών με την ίδια βαθμολογία στο BLEU να έχει εξίσου καλή βαθμολογία και από κριτές. Όπως και οι ίδιοι οι συντάκτες

[20] Ας σημειωθεί ότι για τη χρήση τέτοιου είδους γλωσσολογικών πληροφοριών, η οποία έχει όντως εφαρμοστεί [Banerjee, S. και Lavie, A., 2005], είναι απαραίτητη η ύπαρξη συγκεκριμένων γλωσσικών πόρων (λημματοποιητές, ηλεκτρονικά σημασιολογικά λεξικά, κ.λπ.), οι οποίοι δεν είναι διαθέσιμοι ή επαρκείς για όλες τις φυσικές γλώσσες.

[21] Βέβαια είναι αμφίβολο αν μπορεί να χρησιμοποιηθεί αυτό το κριτήριο στην αξιολόγηση ενός συστήματος αξιολόγησης μηχανικής μετάφρασης, κυρίως γιατί εξ ορισμού πρόκειται για αξιολόγηση μηχανικής μετάφρασης και όχι μετάφρασης από ανθρώπους. Άλλωστε, αυτό το επιχείρημα απλώς αποδεικνύει την ποικιλότητα μεταξύ μεταφράσεων από διαφορετικούς ανθρώπους (δηλαδή η υπό αξιολόγηση μετάφραση δεν συμπίπτει με άλλες μεταφράσεις από ανθρώπους), πράγμα το οποίο αποτελεί ακριβώς τον λόγο για τον οποίο χρησιμοποιούνται πολλές μεταφράσεις-αναφορά.

αναφέρουν, πρόκειται για μια θεωρητική αιτιολόγηση με επινοημένα παραδείγματα, από την οποία θεωρούμε ότι δεν αποδεικνύεται το κατά πόσο έχει αντίκρυσμα στην πράξη (κατά πόσο δηλαδή τα συστήματα μηχανικής μετάφρασης δίνουν κείμενα με τόσο ανακατεμένα ν-γράμματα). Τέλος, η υψηλή συσχέτιση του BLEU με τις κρίσεις ανθρώπων, η οποία αποτελεί από τα βασικά του προτερήματα, καταρρέει στον διαγωνισμό NIST για συστήματα μηχανικής μετάφρασης του 2005 [Callison-Burch, C. κ.ά., 2006].

Ωστόσο σημαντικές ενστάσεις προκύπτουν και από την ίδια την αρχική αξιολόγηση του συστήματος που παρουσιάζεται από τους ερευνητές που το ανέπτυξαν [Papineni, K. κ.ά., 2002]. Σε αυτή την αξιολόγηση κρίνονται μέσω BLEU με χρήση 2 μεταφράσεων-αναφορά οι μεταφράσεις από τα κινέζικα στα αγγλικά 3 συστημάτων μηχανικής μετάφρασης και 2 ανθρώπων, εκ των οποίων ο ένας είναι φυσικός ομιλητής της αγγλικής και ο άλλος δεν είναι φυσικός ομιλητής ούτε της κινεζικής ούτε της αγγλικής[22]. Για να διαπιστωθεί ο βαθμός συσχέτισης, οι ίδιες μεταφράσεις κρίνονται και από 2 ομάδες αξιολογητών. Η μία ομάδα αποτελείται από 10 φυσικούς ομιλητές της αγγλικής και η άλλη από 10 φυσικούς ομιλητές της κινεζικής που έχουν ζήσει αρκετά χρόνια στις ΗΠΑ, κανένας από τους οποίους δεν είναι επαγγελματίας μεταφραστής. Δεν παρέχεται κάποια εξήγηση σχετικά με τον λόγο επιλογής μη μεταφραστών, όπως και δεν παρέχεται το προφίλ των κριτών με στοιχεία όπως το επίπεδό τους στις γλώσσες εργασίας, μορφωτικό επίπεδο, εξοικείωση με τομείς κειμένων των μεταφράσεων, κ.λπ. Αν σε αυτό προσθέσουμε το γεγονός ότι οι δίγλωσσοι κριτές αξιολογούν αντίστροφη για αυτούς μετάφραση, μετάφραση δηλαδή από τη μητρική τους γλώσσα προς κάποια άλλη, τίθεται θέμα αξιοπιστίας της αξιολόγησης από τα εν λόγω υποκείμενα αλλά και της σύγκρισης των αποτελεσμάτων μεταξύ των δύο ομάδων. Τέλος, οι συγγραφείς καταλήγουν σε συμπεράσματα συγκρίνοντας ανόμοια δεδομένα. Συναγάγουν το συμπέρασμα ότι οι δίγλωσσοι κριτές σε σχέση με τους μονόγλωσσους επικεντρώνονται περισσότερο στην πιστότητα από ό,τι στη γραμματικότητα του μεταφράσματος. Επίσης συμπεραίνουν ότι οι δίγλωσσοι κριτές σε σχέση με τους μονόγλωσσους είναι πολύ επιεικείς με τον μεταφραστή που δεν είναι φυσικός ομιλητής της κινεζικής και της αγγλικής. Ωστόσο οι συγκρίσεις μεταξύ των δύο ομάδων κριτών είναι άστοχη, διότι οι δίγλωσσοι καλούνται να κρίνουν και τα δύο προαναφερθέντα επίπεδα (πιστότητα και γραμματικότητα) ενώ οι μονόγλωσσοι μόνο το δεύτερο, εφόσον δεν έχουν γλωσσική πρόσβαση στο πρωτότυπο (προφανώς δεν γνωρίζουν την κινεζική και ελέγχουν μόνο το μετάφρασμα).

Παρά τις προαναφερθείσες κριτικές και αδυναμίες, το BLEU παραμένει από τις πιο χρήσιμες μετρικές αξιολόγησης της ποιότητας μηχανικής μετάφρασης, ακόμη κι αν αυτό είναι μέχρι να προταθεί κάποια καλύτερη [Callison-Burch, C. κ.ά., 2006].

[22] Η χρήση μη φυσικού ομιλητή γίνεται, για να διαπιστωθεί αν θα φανεί η ποιοτική διαφορά μεταξύ των μεταφράσεων των δύο ανθρώπων.

3.1.1.5 Πλεονεκτήματα μετρικών με μεταφράσεις-αναφορά

Συνοψίζοντας, τα πλεονεκτήματα που παρουσιάζουν αυτές οι μετρικές στο σύνολό τους, σε σχέση πάντα με τη μη αυτοματοποιημένη ποιοτική αξιολόγηση, είναι η ταχύτητα [Lavie, A., 2011, Banerjee, S. και Lavie, A., 2005] και το χαμηλό κόστος [Lavie, A., 2011, Banerjee, S. και Lavie, A., 2005, Koehn, P., 2010], η μειωμένη ανάγκη ανθρώπινων πόρων [Lavie, A., 2011, Banerjee, S. και Lavie, A., 2005] και η δυνατότητα επαναχρησιμοποίησής τους [Lavie, A., 2011, Banerjee, S. και Lavie, A., 2005]. Αναφέρεται επίσης ως πλεονέκτημα το ότι δεν απαιτούν την ύπαρξη δίγλωσσων κριτών [Lavie, A., 2011, Banerjee, S. και Lavie, A., 2005]. Είναι όμως σαφές ότι απαιτείται η ύπαρξη μεταφραστών και η παραγωγή μεταφράσεων, πράγμα το οποίο, όσο και αν "αποσιωπάται", δεν επιτρέπει την πλήρη αυτοματοποίηση της διαδικασίας και προφανώς την ελαχιστοποίηση κόστους και χρόνου. Το βασικό πλεονέκτημα είναι η δυνατότητα χρήσης τους κατά την ανάπτυξη συστημάτων μηχανικής μετάφρασης, πράγμα το οποίο επιτρέπει αλλαγές και βελτιώσεις σε αυτά και συνεχή επαναξιολόγηση [Lavie, A., 2011, Banerjee, S. και Lavie, A., 2005, Koehn, P., 2010]. Τέλος, γίνεται λόγος για αντικειμενικότητα αυτών των μετρικών, σε βαθμό μάλιστα που υπάρχουν και αναφορές σε αυτές με τον όρο "αντικειμενική αξιολόγηση" σε αντιπαράθεση με τη μη αυτοματοποιημένη αξιολόγηση (υποκειμενική) [Fordyce, C. S., 2007]. Ωστόσο σε αυτό το σημείο θεωρούμε ότι υπάρχει σύγχυση της αντικειμενικότητας με τη συνέπεια που επιδεικνύει ένα μηχανικό σύστημα σε σχέση με έναν άνθρωπο. Δεν υπάρχει καμία βεβαιότητα ότι ένας άνθρωπος θα δώσει την ίδια βαθμολογία στο ίδιο υπό αξιολόγηση κείμενο δύο φορές και ούτε βέβαια ότι δύο οι περισσότεροι άνθρωποι θα συμφωνήσουν στην ίδια βαθμολογία. Ενώ μια μετρική θα αποδώσει ακριβώς την ίδια βαθμολογία για το ίδιο κείμενο, εφόσον διατηρούνται όλες οι παράμετροι αξιολόγησης. Από αυτή την άποψη, είναι σαφές πλεονέκτημα η συνέπεια των μηχανικών συστημάτων [Fordyce, C. S., 2007, Koehn, P., 2010]. Ωστόσο δεν μπορούμε να μιλάμε για αντικειμενικότητα τη στιγμή που γίνεται χρήση μεταφράσεων-αναφορά, οι οποίες αποτελούν προϊόντα της ανθρώπινης διάνοιας, εφόσον οι άνθρωποι δεν είναι άλλο παρά υποκείμενα, και τη στιγμή μάλιστα που το επιθυμητό αποτέλεσμα είναι η εγγύτητα σε αυτό το υποκειμενικό προϊόν. Και φυσικά είναι ιδιαίτερα αμφισβητήσιμο κατά πόσο μπορεί οτιδήποτε άπτεται των γλωσσικών φαινομένων να αντιμετωπιστεί και να μετρηθεί αντικειμενικά.

3.1.1.6 Μειονεκτήματα μετρικών με μεταφράσεις-αναφορά

Εν συντομία, τα μειονεκτήματα που έχουν αναφερθεί σχετικά με αυτές τις μετρικές αυτόματης αξιολόγησης περιλαμβάνουν το γεγονός ότι η ανάγκη ύπαρξης μεταφράσεων-αναφορά περιορίζει τον αριθμό των δεδομένων που μπορούν να αξιολογηθούν [Specia, L. κ.ά., 2010], την αδυναμία διάκρισης μεταξύ λεπτών διαφορών [Lavie, A., 2011], την αναξιοπιστία των βαθμολογιών ανά πρόταση [Lavie, A., 2011], καθώς και τη δυσκολία ερμηνείας των εκάστοτε βαθμολογιών [Koehn, P., 2010]. Τέλος, σε γενικές γραμμές, οι περισσότερες από τις προαναφερθείσες αδυναμίες του BLEU θεωρείται ότι ισχύουν και για άλλες μετρικές [Callison-Burch, C. κ.ά., 2006].

3.1.2 Εκτίμηση Ποιότητας (Quality Estimation)

Οι αυτόματες μετρικές εκτίμησης ποιότητας είναι συστήματα ταξινόμησης, τα οποία έχουν περάσει από τον χώρο της αναγνώρισης φωνής στον χώρο της μηχανικής μετάφρασης από τις αρχές της δεκαετίας του 2000 [Specia, L. κ.ά., 2013] και στην περίπτωση της μηχανικής μετάφρασης ταξινομούν τα μεταφράσματα σε επίπεδα ποιότητας.

Σε αυτή την κατηγορία μετρικών, η εκτίμηση της ποιότητας πραγματοποιείται όχι πλέον μέσω της σύγκρισης του μηχανικού μεταφράσματος με μεταφράσεις-αναφορά αλλά μέσω της ταξινόμησης των μηχανικών μεταφρασμάτων σε δύο ή περισσότερα επίπεδα ποιότητας. Η ταξινόμηση επιτελείται με τη χρήση γλωσσικών και μη χαρακτηριστικών, καθώς και μεταφράσεων επισημειωμένων ως προς αυτά τα επίπεδα, με τα οποία εκπαιδεύονται τα συστήματα (επιβλεπόμενη μηχανική μάθηση). Οι εν λόγω μετρικές προτείνονται τόσο για αξιολόγηση της ποιότητας, όσο και για την επιλογή μεταξύ διαφορετικών συστημάτων μηχανικής μετάφρασης μεταξύ άλλων [Specia, L. κ.ά., 2013], ενέργεια για την οποία δεν προσφέρονται οι παραδοσιακές μετρικές, όπως έχουμε δει. Οι μετρικές εκτίμησης ποιότητας δεν έχουν τον ίδιο στόχο με τις παραδοσιακές μετρικές και δεν προορίζονται να τις αντικαταστήσουν, αλλά να καλύψουν το κενό που δημιουργείται στις περιπτώσεις όπου δεν υπάρχουν μεταφράσεις-αναφορά αλλά και τις ανάγκες εκτίμησης ποιότητας σε επίπεδο προτάσεων, όπου οι άλλες μετρικές υστερούν [Specia, L. κ.ά., 2010].

Στις πρώτες προσπάθειες που πραγματοποιήθηκαν για τη χρήση μετρικών εκτίμησης ποιότητας στη μηχανική μετάφραση εξακολουθούσε να υπάρχει η ανάγκη ύπαρξης μεταφράσεων-αναφορά. Ωστόσο, έχουν ήδη αναπτυχθεί συστήματα για τα οποία απαιτείται μόνο χειρωνακτική επισημείωση μεταφράσεων ως προς τα καθορισμένα επίπεδα ποιότητας και έχουν προταθεί πολλά χαρακτηριστικά που θεωρείται ότι συμβάλλουν σε καλά αποτελέσματα [Specia, L. κ.ά., 2013].

Συγκεκριμένα, η πλατφόρμα που παρουσιάζουν οι Specia κ.ά. [2013] απαρτίζεται από δύο μέρη. Στο ένα μέρος υπάρχουν επιμέρους συστήματα για την ανάκτηση χαρακτηριστικών από τα κείμενα-πηγή και τα κείμενα-στόχος. Το άλλο μέρος είναι το μέρος της μηχανικής μάθησης και είναι τελείως ανεξάρτητο από το πρώτο. Αυτό χρησιμοποιεί τα χαρακτηριστικά που έχει ανακτήσει το πρώτο μέρος, για να δημιουργήσει μοντέλα εκτίμησης ποιότητας με τη βοήθεια αλγόριθμων παλινδρόμησης[23] και ταξινόμησης.

Τα χαρακτηριστικά που χρησιμοποιούνται στην εν λόγω πλατφόρμα χωρίζονται σε τρεις κατηγορίες: α) πολυπλοκότητας, β) γραμματικότητας και γ) πιστότητας. Τα χαρακτηριστικά πολυπλοκότητας αφορούν τον υπολογισμό της πολυπλοκότητας της μετάφρασης των συγκεκριμένων προτάσεων γενικά ή για το συγκεκριμένο σύστημα μηχανικής μετάφρασης, τα χαρακτηριστικά γραμματικότητας ανακτώνται από τις μεταφράσεις και αφορούν την πρόταση-στόχος, και τα χαρακτηριστικά πιστότητας αφορούν τον υπολογισμό του βαθμού στον οποίο διατηρούνται η δομή και η σημασία του πρωτότυπου στη μετάφραση.

Παρόλο που εξακολουθεί να απαιτείται μη αυτοματοποιημένη εργασία για την υλοποίηση αυτών των νεότερων τεχνικών, παρουσιάζουν ιδιαίτερο ενδιαφέρον και ανοίγουν δρόμους πιθανών συνδυασμών με τις παραδοσιακές [Specia, L. κ.ά., 2010].

3.2 Μη αυτοματοποιημένη αξιολόγηση ποιότητας μηχανικής μετάφρασης

Ως μη αυτοματοποιημένη αξιολόγηση ποιότητας της μηχανικής μετάφρασης ορίζεται στην παρούσα εργασία η αξιολόγηση στην οποία επεμβαίνει ο άνθρωπος στην εκτέλεση της αξιολόγησης, σε αντίθεση με την αυτόματη αξιολόγηση, όπου επεμβαίνει μόνο σε προγενέστερα στάδια. Στην παρουσίασή μας γίνεται η περαιτέρω διάκριση μεταξύ μεθόδων αξιολόγησης με άμεσα εκπεφρασμένη αξιολογική κρίση και μεθόδων χωρίς άμεσα εκπεφρασμένη αξιολογική κρίση.

Ουσιαστικό ρόλο στη μη αυτοματοποιημένη αξιολόγηση διαδραματίζουν οι ίδιοι οι κριτές ή επισημειωτές. Οι κριτές πρέπει να πληρούν συγκεκριμένα κριτήρια, προκειμένου να διασφαλιστεί η αξιοπιστία της διαδικασίας και των αποτελεσμάτων. Ανάλογα με το είδος της αξιολόγησης, οι κριτές μπορούν να είναι μονόγλωσσοι ή δίγλωσσοι, δηλαδή άτομα που είναι φυσικοί ή σχεδόν φυσικοί ομιλητές της γλώσσας-στόχος ή της γλώσσας-στόχος και της γλώσσας-πηγή. Απαραίτητη είναι η εκπαίδευσή τους και η ύπαρξη οδηγιών επισημείωσης με παραδείγματα, καθώς και η εξοικείωσή τους με το πεδίο των κειμένων. Ως προς τη μητρική γλώσσα των δίγλωσσων κριτών, σε ό,τι αφορά τους μεταφραστές, για ένα μεγάλο διάστημα είχε υιοθετηθεί η αρχή της μετάφρασης προς τη μητρική γλώσσα του μεταφραστή που προτεινόταν στη Σύσταση της

[23] Αλγόριθμοι με τους οποίους υπολογίζονται οι σχέσεις μεταξύ μίας εξαρτημένης μεταβλητής και μίας ή περισσότερων ανεξάρτητων μεταβλητών.

22/11/1976 της UNESCO[24], η οποία όμως έχει αμφισβητηθεί τις τελευταίες δεκαετίες [Sánchez-Gijón, P. και Torres-Hostench, O., 2014]. Καθώς υπάρχει μια τάση η μετεπεξεργασία (post-editing)[25] να αντιμετωπίζεται όπως η μετάφραση, έχει ήδη αρχίσει να διεξάγεται σχετική έρευνα σε αυτό τον τομέα [Lacruz, I. κ.ά., 2014]. Τέλος, ιδανικά χρειάζονται περισσότεροι από ένας κριτές, το επίπεδο συμφωνίας των οποίων πρέπει να υπολογίζεται. Στην πράξη το θέμα της εύρεσης κριτών είναι ακανθώδες, διότι σε πολλές περιπτώσεις συνεπάγεται οικονομική επιβάρυνση. Πέρα από τα χαρακτηριστικά των κριτών, είναι σημαντικές και οι συνθήκες στις οποίες εκτελείται η αξιολόγηση, όπως ο λογικός όγκος μετάφρασης προς επισημείωση και η διεξαγωγή της χωρίς διακοπές [Przybocki, M., κ.ά., 2011].

3.2.1 Μέθοδοι αξιολόγησης με άμεσα εκπεφρασμένη αξιολογική κρίση

Σε αυτή την κατηγορία μεθόδων οι κριτές εκφέρουν άμεσα αξιολογική κρίση για την ποιότητα του μεταφράσματος. Αυτή η αξιολογική κρίση εκφράζεται συνήθως για την πιστότητα και τη γραμματικότητα. Αυτό επιτυγχάνεται είτε με την αντιπαράθεση πρωτότυπου με μηχανικό μετάφρασμα είτε με την αντιπαράθεση μηχανικού μεταφράσματος με μετάφραση-αναφορά. Επειδή ακριβώς πρόκειται για την πιο διαδεδομένη αξιολόγηση τέτοιου τύπου, υπάρχει πληθώρα εργαλείων που χρησιμοποιούνται για την εκτέλεσή της. Μάλιστα συνηθίζεται η αξιολόγηση αυτή να γίνεται σε κλίμακα πέντε επιπέδων και για τις δύο προαναφερθείσες πτυχές [Callison-Burch, Ch. κ.ά., 2007], αν και χρησιμοποιούνται και κλίμακες με επτά επίπεδα [Przybocki, M. κ.ά., 2009]. Στη μεν πιστότητα δηλώνεται αν εκφράζεται στη μετάφραση όλο (5), το μεγαλύτερο μέρος (4), μεγάλο μέρος (3), μικρό μέρος (2) ή τίποτα (1) από το περιεχόμενο στη μετάφραση-αναφορά [Callison-Burch, Ch. κ.ά., 2007] ή στο πρωτότυπο. Στη δε γραμματικότητα δηλώνεται αν η γλώσσα της μετάφρασης είναι άψογη (5), καλή (4), όχι φυσική (όχι όπως ενός φυσικού ομιλητή δηλαδή) (3), μη γραμματική (2) ή μη κατανοητή (1) [Callison-Burch, Ch. κ.ά., 2007]. Άλλες μέθοδοι συνίστανται στην κατάταξη διαφόρων συστημάτων, στη σύγκριση μεταξύ δύο συστημάτων ή σε μια γενικότερη κρίση για το μετάφρασμα. Ακολουθούν οι περιγραφές κάποιων από τα πιο διαδεδομένα εργαλεία που χρησιμοποιούνται για τέτοιου είδους αξιολόγηση.

[24] Οι Συστάσεις της UNESCO διατίθενται στην ιστοσελίδα της (http://portal.unesco.org/en/ev.php-URL_ID=12026&URL_DO=DO_TOPIC&URL_SECTION=-471.html, τελευταία προσπέλαση 19/06/2015).

[25] Ο αγγλικός όρος “post-editing” σημαίνει “επιμέλεια μηχανικού μεταφράσματος” και αποδίδεται στα ελληνικά ως “μετεπεξεργασία”, “μετεπιμέλεια” ή “επιμέλεια αυτόματης μετάφρασης”. Η πολυλεκτική απόδοση, αν και πιο περιεκτική, υστερεί σε επίπεδο οικονομίας λόγου και ευχρηστίας. Μεταξύ των αποδόσεων “μετεπεξεργασία” και “μετεπιμέλεια” θεωρούμε ότι η “μετεπεξεργασία” είναι πιο εύστοχη, διότι με τον όρο “μετεπιμέλεια” δημιουργείται το παράδοξο να εκτελείται πρώτα μηχανική μετάφραση, μετά μετεπιμέλεια και στη συνέχεια επιμέλεια, ο απαραίτητος σε κάθε μετάφραση τελικός έλεγχος.

3.2.1.1 DQF: Εργασίες επισημείωσης σε πιστότητα, γραμματικότητα και κατάταξη

Το DQF (Dynamic Quality Framework)[26] είναι μια πλατφόρμα την οποία ανέπτυξε η εταιρεία TAUS (Translation Automation User Society)[27] το 2011 με στόχο την τυποποίηση της αξιολόγησης της ποιότητας της μετάφρασης (μηχανικής και μη) και λειτουργεί από το 2014 με ελεύθερη πρόσβαση για άτομα από τον ακαδημαϊκό χώρο. Σε αυτήν παρέχονται εργαλεία, καλές πρακτικές, μετρικές, αναφορές και δεδομένα που χρησιμεύουν στην αξιολόγηση της ποιότητας της μετάφρασης. Ο χρήστης συμπληρώνει στοιχεία σχετικά με την αξιολόγηση που επιθυμεί να πραγματοποιήσει (περιεχόμενο, σκοπός, είδος επικοινωνίας) και επιλέγει εργασία αξιολόγησης. Τα κείμενα φορτώνονται σε μορφή υπολογιστικών φύλλων που περιέχουν συγκεκριμένες πληροφορίες σχετικά με αυτά. Τα αποτελέσματα εξάγονται στην ίδια μορφή. Στον πίνακα με τα αποτελέσματα εμφανίζονται και οι προτάσεις (πρωτότυπη και μεταφρασμένη). Ως προς τις μεθόδους αξιολόγησης με άμεσα εκπεφρασμένη αξιολογική κρίση, παρέχει περιβάλλοντα επισημείωσης για αξιολόγηση πιστότητας και γραμματικότητας αλλά και για κατάταξη.

Πιστότητα και γραμματικότητα

Η πιστότητα (“adequacy” στο DQF) ορίζεται ως ο βαθμός στον οποίο το νόημα που εκφράζεται στη μετάφραση-αναφορά ή στο κείμενο-πηγή εκφράζεται και στη μετάφραση[28]. Η γραμματικότητα ορίζεται ως ο βαθμός στον οποίο η μετάφραση είναι γραμματική, ορθογραφημένη, ανταποκρίνεται στην κοινή χρήση των όρων, των τίτλων και των ονομάτων, είναι διαισθητικά αποδεκτή και λογικά ερμηνεύσιμη από έναν φυσικό ομιλητή[29].

Η πιστότητα είναι ένα πεδίο που αφορά τόσο το κείμενο-πηγή όσο και το κείμενο-στόχος και τη σχέση μεταξύ τους. Η γραμματικότητα αφορά αποκλειστικά το κείμενο-στόχος και αξιολογείται χωρίς να λαμβάνεται υπόψη το πρωτότυπο.

Στο εργαλείο που παρέχεται σε αυτή την πλατφόρμα ο χρήστης επιλέγει αν θα χρησιμοποιήσει πιστότητα ή γραμματικότητα ή και τα δύο. Το σύνηθες είναι και τα δύο. Σε αυτό το εργαλείο η κλίμακα έχει τέσσερα και όχι πέντε επίπεδα. Για την εκτέλεση της αξιολόγησης ο επισημειωτής καλείται να απαντήσει στις εξής ερωτήσεις:

---

[26] https://evaluate.taus.net/evaluate/dqf/dynamic-quality-framework (τελευταία προσπέλαση 19/06/2015)

[27] https://www.taus.net (τελευταία προσπέλαση 19/06/2015)

[28] Οδηγίες TAUS Μάιος 2013 (https://www.taus.net/think-tank/best-practices/evaluate-best-practices/adequacy-fluency-guidelines, τελευταία προσπέλαση 04/05/2015)

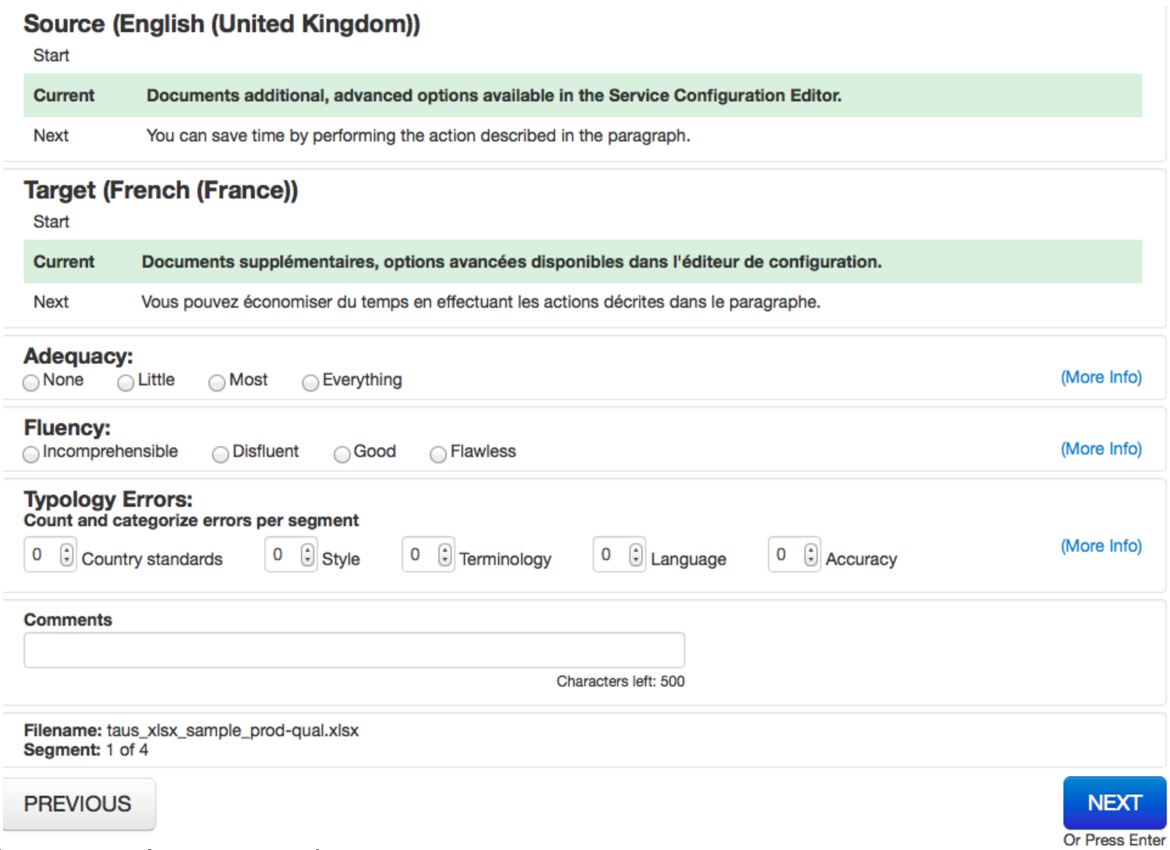
[29] Οδηγίες TAUS Μάιος 2013 (https://www.taus.net/think-tank/best-practices/evaluate-best-practices/adequacy-fluency-guidelines, τελευταία προσπέλαση 04/05/2015)

- Πιστότητα: Τι μέρος του νοήματος εκφράζεται στη μετάφραση;

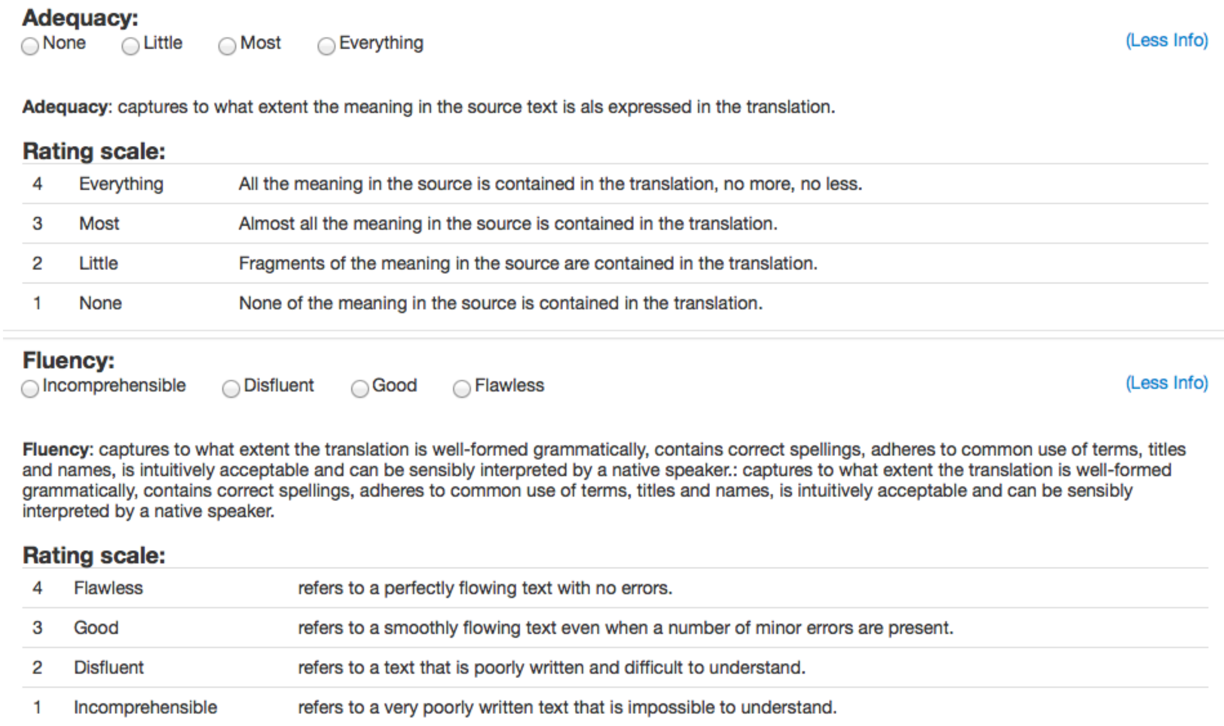
  Τίποτα / Μικρό μέρος / Το μεγαλύτερο μέρος / Όλο
- Γραμματικότητα: Πόσο γραμματική είναι η γλώσσα στη μετάφραση;

  Μη κατανοητή / Μη γραμματική / Καλή / Άψογη

Αυτές οι ερωτήσεις απαντώνται σε ένα περιβάλλον όπου εμφανίζεται η πρωτότυπη πρόταση (η αξιολόγηση πραγματοποιείται πάντα ανά πρόταση) με έντονους χαρακτήρες. Πριν από αυτήν εμφανίζεται η προηγούμενη πρόταση και ακολουθεί η επόμενη, αν υπάρχουν· αλλιώς δηλώνεται αρχή ή τέλος αρχείου. Από κάτω εμφανίζονται οι αντίστοιχες προτάσεις της μετάφρασης. Ακολουθεί η επιλογή μίας από τις 4 απαντήσεις για κάθε πτυχή (πιστότητα και γραμματικότητα). Όπως φαίνεται και στην Εικόνα 2, στα δεξιά υπάρχει η δυνατότητα εμφάνισης περισσότερων λεπτομερειών σχετικά με αυτές τις απαντήσεις. Εάν επιλεχθεί, εμφανίζονται συνοπτικές οδηγίες/επεξηγήσεις για την επιλογή της κατάλληλης απάντησης (Εικόνα 3). Ο χρήστης έχει τη δυνατότητα να επιστρέψει σε προηγούμενη πρόταση με την επιλογή “Previous”, και επιλέγοντας “Next” καταχωρίζονται οι απαντήσεις και περνάει στην επόμενη πρόταση.



Εικόνα 2: Περιβάλλον επισημείωσης DQF



Adequacy:
None Little Most Everything (Less Info)

Adequacy: captures to what extent the meaning in the source text is als expressed in the translation.

Rating scale:

4	Everything	All the meaning in the source is contained in the translation, no more, no less.
3	Most	Almost all the meaning in the source is contained in the translation.
2	Little	Fragments of the meaning in the source are contained in the translation.
1	None	None of the meaning in the source is contained in the translation.

Fluency:
Incomprehensible Disfluent Good Flawless (Less Info)

Fluency: captures to what extent the translation is well-formed grammatically, contains correct spellings, adheres to common use of terms, titles and names, is intuitively acceptable and can be sensibly interpreted by a native speaker.: captures to what extent the translation is well-formed grammatically, contains correct spellings, adheres to common use of terms, titles and names, is intuitively acceptable and can be sensibly interpreted by a native speaker.

Rating scale:

4	Flawless	refers to a perfectly flowing text with no errors.
3	Good	refers to a smoothly flowing text even when a number of minor errors are present.
2	Disfluent	refers to a text that is poorly written and difficult to understand.
1	Incomprehensible	refers to a very poorly written text that is impossible to understand.

Εικόνα 3: Οδηγίες επισημείωσης DQF

Η κλίμακα της πιστότητας είναι η εξής:

4) Όλο: Όλο το νόημα του κειμένου-πηγή περιέχεται στη μετάφραση, ούτε λιγότερο ούτε περισσότερο.

3) Το μεγαλύτερο μέρος: Το μεγαλύτερο μέρος του νοήματος του κειμένου-πηγή περιέχεται στη μετάφραση.

2) Μικρό μέρος: Στη μετάφραση περιέχονται αποσπάσματα του νοήματος του κειμένου-πηγή.

1) Τίποτα: Τίποτα από το νόημα του κειμένου-πηγή δεν περιέχεται στη μετάφραση.

Και η κλίμακα της γραμματικότητας:

4) Άψογη: Κείμενο στο οποίο ο λόγος ρέει άψογα χωρίς λάθη.

3) Καλή: Κείμενο στο οποίο ο λόγος ρέει ομαλά ακόμη και αν υπάρχουν μείζονος σημασίας λάθη.

2) Μη γραμματική: Κείμενο κακογραμμένο και δύσκολο στην κατανόηση.

1) Μη κατανοητή: Πολύ κακογραμμένο και ακατανόητο κείμενο.

Στην επιλογή της κατάλληλης απάντησης ο κριτής πρέπει να έχει πολύ ξεκάθαρα κριτήρια για τα όρια μεταξύ των επιπέδων και επαρκή εκπαίδευση, ώστε τα αποτελέσματα να είναι όσο το δυνατόν πιο αξιόπιστα. Στο πεδίο της γραμματικότητας σε δοκιμή που πραγματοποιήσαμε παρατηρήσαμε ότι οι περισσότερες προτάσεις λαμβάνουν τον χαρακτηρισμό “disfluent” (“μη γραμματική”) και “good” (“καλή”). Δύο παραδείγματα μετάφρασης που χαρακτηρίστηκαν ως “μη γραμματικά” είναι τα εξής[30]:

- “Στρες προκαλεί την παραγωγή κυτοκινών συγκεκριμένα στο κεντρικό νευρικό σύστημα ενεργεί στον τοίχο του παχέος εντέρου.”
- “Ειδικές εξετάσεις θα επιτρέψει σε μας για τον εντοπισμό μισαλλοδοξίας στα τρόφιμα, καθώς και την παροχή συμβουλών σχετικά με ειδική διατροφή σχέδια που θα ανακουφίσει τους ασθενείς συμπτώματα.”

Και δύο παραδείγματα προτάσεων που χαρακτηρίστηκαν ως “καλά”:

- “Οχι, Υποστήριξε ότι οι ασθενείς δεν τρέχει ένα μεγαλύτερο κίνδυνο ανάπτυξης καρκίνου του παχέος εντέρου σε σύγκριση με τον γενικό πληθυσμό.”
- “Για το λόγο αυτό είναι σημαντικό να επιδιωχθεί η βοήθεια του ψυχολόγου για τη θεραπεία των συμπτωμάτων του ευερέθιστου Εντέρου και η αιτία.”

Στο πεδίο της πιστότητας συναντήσαμε μεγαλύτερη ποικιλία μεταξύ των τεσσάρων επιπέδων. Οι παρεχόμενες κλίμακες αξιολόγησης κρίνονται επαρκείς τόσο σε επίπεδο επισημείωσης όσο και σε επίπεδο πιθανής αξιοποίησης αποτελεσμάτων, με μοναδική εξαίρεση τον προβληματισμό επιλογής μεταξύ μη γραμματικής και καλής πρότασης σε συγκεκριμένες περιπτώσεις, κάτι το οποίο όμως είναι γενικό γνώρισμα της γλώσσας και των εργαλείων επισημείωσης και αντιμετωπίζεται με αυστηρό ορισμό των δύο επιπέδων πριν την επισημείωση, καθώς και με επαρκή εκπαίδευση των επισημειωτών.

Κατάταξη

Πρόκειται για μια εργασία σύγκρισης για την επιλογή του καλύτερου συστήματος. Ο μέγιστος αριθμός των συστημάτων που μπορούν να συγκριθούν είναι τρία, διότι έχει αποδειχθεί ότι η ικανότητα ενός ανθρώπου για αξιόπιστη κρίση επηρεάζεται αρνητικά όταν βαθμολογεί περισσότερες από 3 προτάσεις [Görög, A., 2014]. Υπάρχουν δύο είδη εργασιων: η Quick Comparison (γρήγορη σύγκριση), όπου επιλέγεται η καλύτερη μετάφραση από ένα μέγιστο αριθμό τριών συστημάτων, και η Rank Translations (κατάταξη μεταφράσεων), όπου κατατάσσονται οι μεταφρασμένες προτάσεις από την καλύτερη (1) προς τη χειρότερη (3). Οι μεταφρασμένες προτάσεις από το κάθε σύστημα δεν εμφανίζονται πάντα με την ίδια σειρά.

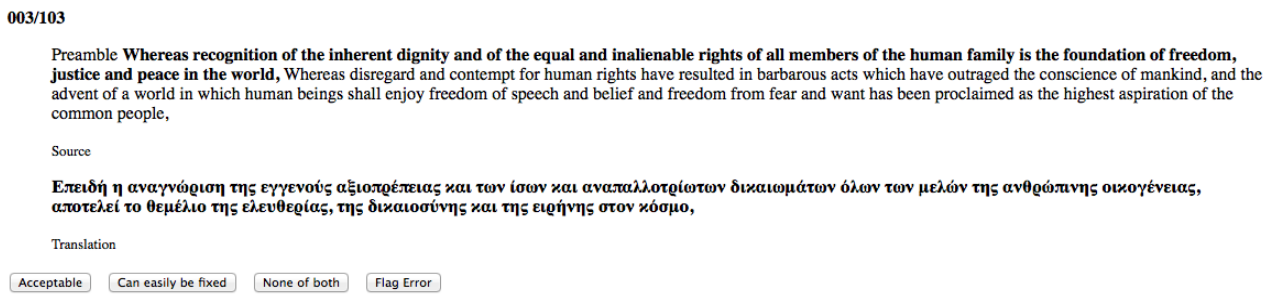
[30] Τα παραδείγματα είναι από το σώμα κειμένων μας, το οποίο παρουσιάζεται στο Κεφάλαιο 4.1.

3.2.1.2 Appraise: Εργασίες επισημείωσης σε έλεγχο ποιότητας και κατάταξη

Το Appraise [Federmann, Ch., 2010] είναι ένα εργαλείο ανοιχτού κώδικα με το οποίο μπορούν να επιτελεστούν εργασίες επισημείωσης για την αξιολόγηση μηχανικής μετάφρασης. Τροφοδοτείται με αρχεία σε μορφή xml και εξάγει αποτελέσματα στην ίδια μορφή. Σε αυτό το εργαλείο δεν δίνεται η δυνατότητα επιστροφής σε προηγούμενη πρόταση. Σε περίπτωση που ο χρήστης επιθυμεί να αλλάξει προηγούμενη επιλογή του, πρέπει να γίνει εκ νέου ανάθεση της εργασίας. Εκτός από την επιλογή της υποβολής της κάθε αξιολογηθείσας πρότασης (“submit”), παρέχεται και η επιλογή “reset”, με την οποία απαλείφεται ό,τι έχει επισημειώσει ο χρήστης στην τρέχουσα πρόταση. Σε αυτό το κεφάλαιο παρουσιάζουμε τρία είδη επισημείωσης, στα οποία ο επισημεωτής εκφράζει άμεσα αξιολογική κρίση για το μετάφρασμα ή το σύστημα υπό αξιολόγηση.

Quality Checking (Έλεγχος Ποιότητας)

Σε όλες τις εργασίες αξιολόγησης του Appraise εμφανίζεται το κείμενο ανά πρόταση υπό αξιολόγηση. Στο πάνω μέρος εμφανίζεται η πρωτότυπη πρόταση με έντονους χαρακτήρες, καθώς και η προηγούμενη και η επόμενη με κανονική γραφή. Από κάτω εμφανίζεται η μεταφρασμένη πρόταση.



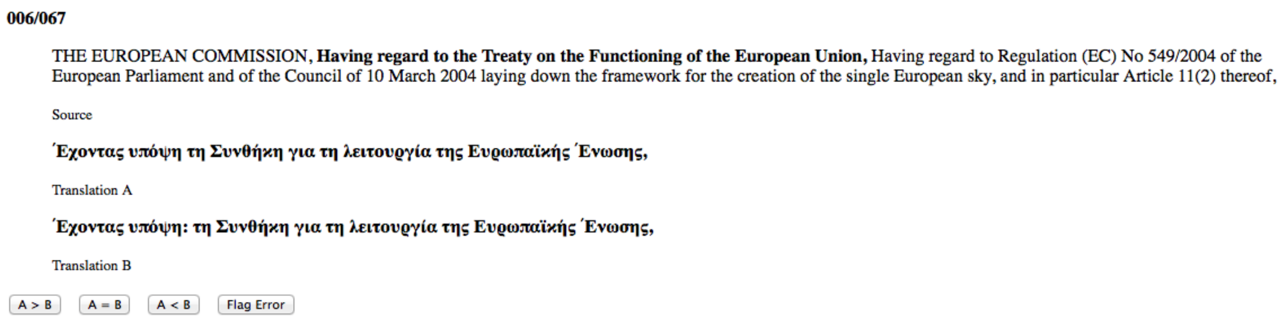
Εικόνα 4: Περιβάλλον επισημείωσης Ελέγχου Ποιότητας στο Appraise

Στην Εικόνα 4 παρουσιάζεται το περιβάλλον του Appraise στη συγκεκριμένη εργασία. Πάνω αριστερά αναφέρεται ο αριθμός της τρέχουσας πρότασης υπό αξιολόγηση (3 από 103 στο συγκεκριμένο κείμενο), ακολουθούν η πρωτότυπη και η μεταφρασμένη πρόταση και, τέλος, οι ετικέτες επισημείωσης. Μόλις επιλεχθεί μία από τις ετικέτες, εμφανίζεται στη οθόνη η επόμενη πρόταση προς αξιολόγηση.

Οι ετικέτες επισημείωσης είναι οι εξής: “acceptable”, “can easily be fixed”, “none of both”. Ο χρήστης δηλώνει αν θεωρεί ότι η συγκεκριμένη υπό αξιολόγηση πρόταση είναι αποδεκτή, αν μπορεί να διορθωθεί εύκολα και αν δεν ισχύει τίποτα από τα δύο. Στη συγκεκριμένη πλατφόρμα δεν δίνονται επεξηγήσεις/οδηγίες σχετικά με τις κατηγορίες και τις διαβαθμίσεις. Ως αποδεκτή θεωρούμε μια πρόταση η οποία μπορεί να παραμείνει ακριβώς ως έχει σε ένα κείμενο το οποίο θα παραδοθεί στον τελικό παραλήπτη του (ο εικονικός πελάτης/χρήστης). Στη δεύτερη ετικέτα ορίζουμε το "εύκολα" ως το πεδίο αλλαγών που τελειώνει εκεί που αρχίζει η μεταφραστική προσπάθεια. Αυτό πρακτικά σημαίνει ότι σε μια πρόταση μπορεί να υπάρχει από ένα σχεδόν αμελητέο λάθος (ένα κόμμα το οποίο δεν αλλάζει το νόημα της πρότασης ή ένας τόνος) μέχρι πολλά λάθη που διορθώνονται χωρίς μεγάλη μεταφραστική προσπάθεια (λεξιλογική έρευνα, αμφιταλάντευση μεταξύ δύο πιθανών λύσεων, κ.λπ.). Η ετικέτα "None of both" χρησιμοποιείται όταν, για να γίνει η διόρθωση, απαιτείται έρευνα ή ριζικές αλλαγές που προϋποθέτουν επίλυση απαιτητικών μεταφραστικών προβλημάτων. Ένα παράδειγμα αποδεκτής μετάφρασης αποτελεί το εξής: "UNIVERSAL DECLARATION OF HUMAN RIGHTS" → "Οικουμενική Διακήρυξη για τα Ανθρώπινα Δικαιώματα". Παραδείγματα πρότασης που επιδέχεται εύκολη βελτίωση είναι τα εξής: "Article 2" → "άρθρο 2" (μοναδική διόρθωση το αρχικό άλφα σε κεφαλαίο)· "Everyone has the right to recognition everywhere as a person before the law." → "Ο καθένας έχει το δικαίωμα να αναγνωρίζονται παντού ως πρόσωπο ενώπιον του νόμου." (διόρθωση "αναγνωρίζονται" → "αναγνωρίζεται")· "Furthermore, no distinction shall be made on the basis of the political, jurisdictional or international status of the country or territory to which a person belongs, whether it be independent, trust, non-self-governing or under any other limitation of sovereignty." → "Επιπλέον, δεν γίνεται διάκριση με βάση του πολιτικού, νομικού ή διεθνούς καθεστώτος της χώρας ή του εδάφους στο οποίο ένα άτομο ανήκει, αν είναι ανεξάρτητη, μη αυτοδιοικούμενο ή υπό οποιονδήποτε άλλον περιορισμό κυριαρχίας." (μεγάλη πρόταση με περισσότερα λάθη, τα οποία όμως διορθώνονται εύκολα: "Επιπλέον, δεν **πρέπει να** γίνεται διάκριση **βάσει** του πολιτικού, νομικού ή διεθνούς καθεστώτος της χώρας ή του εδάφους στο οποίο **ανήκει** ένα άτομο, **είτε** είναι ανεξάρτητη, μη αυτοδιοικούμεν**η** ή υπό οποιονδήποτε άλλον περιορισμό κυριαρχίας."). Παράδειγμα ετικέτας "None of both" αποτελεί η πρόταση: "When the condition is mild, all the ways needed to relax as much as possible the spastic muscles and stimulate the hypotonic muscles (including various kinds of massage, such as general massage, segmentary massage, nerve-point massage, etc., postisometric relaxation, microresonance therapy, etc.) combined with correcting drug therapy, and homeopathic and homotoxicological therapy;" → "Όταν η πάθηση είναι ήπια, όλοι οι δρόμοι που χρειάζονται για να χαλαρώσετε όσο το δυνατόν περισσότερο σπαστικών μυών και τόνωση των μυών υποτονικό (συμπεριλαμβανομένων των διαφόρων ειδών μασάζ, όπως η γενική μασάζ, τμηματικό μασάζ, νεύρο-σημείο μασάζ, κλπ, μετα-ισομετρική χαλάρωση, θεραπεία microresonance, κλπ) σε συνδυασμό με διορθωτική φαρμακευτική αγωγή και με ομοιοπαθητική και homotoxicological θεραπεία?" (υπάρχουν αρκετά γραμματικά και λεξικά λάθη, καθώς και αμετάφραστες λέξεις, οι οποίες είναι όροι, συνεπώς απαιτείται έρευνα για την απόδοσή τους· μια αποδεκτή απόδοση αυτής της πρότασης είναι η εξής: "Όταν η κατάσταση είναι ήπιας μορφής, χρειάζεται να χαλαρώσουν όσο το δυνατόν περισσότερο οι σπαστικοί μύες και να τονωθούν οι υποτονικοί μύες (ακόμη και με διάφορα είδη μασάζ, όπως γενικό μασάζ, τμηματικό μασάζ, μασάζ νευρικών απολήξεων, κ.λπ., μετα-ισομετρική χαλάρωση, θεραπεία με σκέδαση, κ.λπ.) σε συνδυασμό με διορθωτική φαρμακευτική αγωγή και με ομοιοπαθητική και ομοτοξικολογική θεραπεία."). Φυσικά το πόσο εύκολα μπορεί να διορθωθεί ένα κείμενο εξαρτάται και από τους παράγοντες που επηρεάζουν το πόσο εύκολα μπορεί να μεταφραστεί ένα κείμενο, όπως είναι ο βαθμός πολυπλοκότητας, αμφισημίας και ευκρίνειας καθώς και η ορολογία του πρωτότυπου κειμένου.

3-way Ranking (κατάταξη δύο συστημάτων)

Σε αυτή την εργασία επισημειώνονται οι προτάσεις ανα ζέυγη συστημάτων επιλέγοντας μία από τις τρεις ετικέτες: A>B, A=B και A<B.

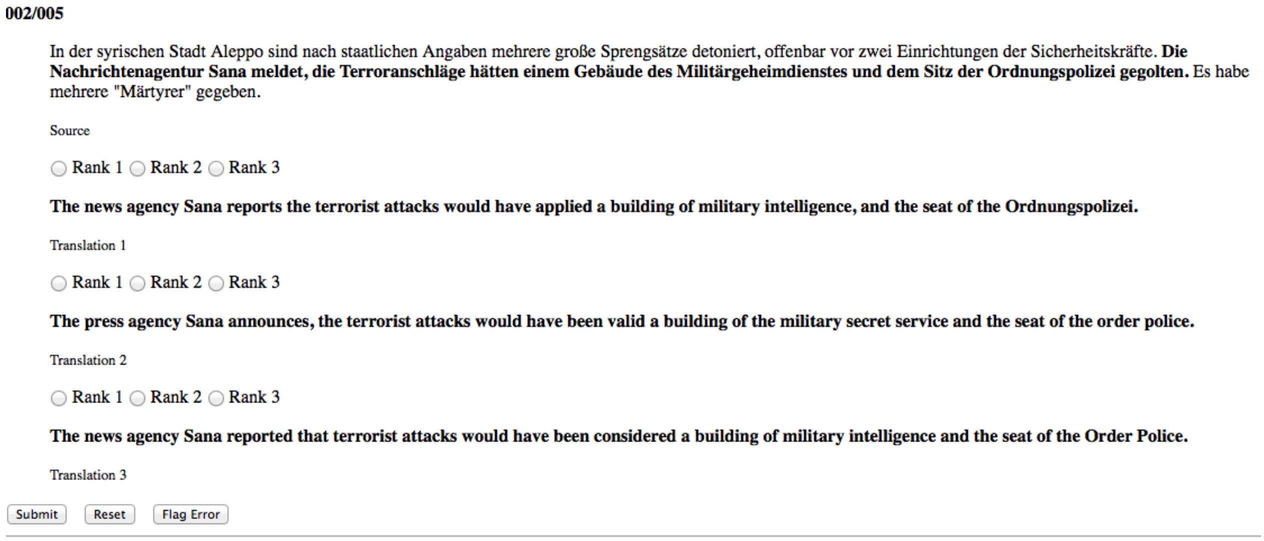


Εικόνα 5: Περιβάλλον επισημείωσης κατάταξης δύο συστημάτων στο Appraise

Στην Εικόνα 5 φαίνεται το περιβάλλον αυτής της εργασίας αξιολόγησης, όπου πάλι εμφανίζεται ο αριθμός της τρέχουσας πρότασης στο σύνολο των προτάσεων του κειμένου, η πρωτότυπη πρόταση σε περιβάλλον και οι δύο υπό αξιολόγηση προτάσεις.

Ranking (Κατάταξη)

Σε αυτή την εργασία παρουσιάζονται οι προτάσεις με τον ίδιο τρόπο όπως στην προηγούμενη. Μπορούν να εμφανιστούν τουλάχιστον δύο μεταφράσεις. Ο χρήστης καλείται να επιλέξει μεταξύ 1, 2 και 3 για να χαρακτηρίσει την κάθε μετάφραση (στο παράδειγμα στην Εικόνα 6, όπου έχουμε 3 μεταφράσεις) και οι ισοβαθμίες επιτρέπονται από το σύστημα.



002/005

In der syrischen Stadt Aleppo sind nach staatlichen Angaben mehrere große Sprengsätze detoniert, offenbar vor zwei Einrichtungen der Sicherheitskräfte. **Die Nachrichtenagentur Sana meldet, die Terroranschläge hätten einem Gebäude des Militärgeheimdienstes und dem Sitz der Ordnungspolizei gegolten.** Es habe mehrere "Märtyrer" gegeben.

Source

Rank 1 Rank 2 Rank 3

The news agency Sana reports the terrorist attacks would have applied a building of military intelligence, and the seat of the Ordnungspolizei.

Translation 1

Rank 1 Rank 2 Rank 3

The press agency Sana announces, the terrorist attacks would have been valid a building of the military secret service and the seat of the order police.

Translation 2

Rank 1 Rank 2 Rank 3

The news agency Sana reported that terrorist attacks would have been considered a building of military intelligence and the seat of the Order Police.

Translation 3

Submit Reset Flag Error

Εικόνα 6: Περιβάλλον επισημείωσης κατάταξης μεταξύ περισσότερων συστημάτων στο Appraise

3.1.2.3 Άλλες τεχνικές

Υπάρχει βιβλιογραφική αναφορά και σε άλλη μια τεχνική παρεμφερή με την προαναφερθείσα εργασία κατάταξης, στην οποία, αντί να πραγματοποιείται η επισημείωση σε επίπεδο πρότασης, πραγματοποιείται σε επίπεδο συντακτικών συστατικών [Callison-Burch, Ch. κ.ά., 2007]. Αυτές οι τεχνικές κατάταξης έχει παρατηρηθεί ότι έχουν υψηλότερη συμφωνία μεταξύ επισημειωτών από ό,τι οι εργασίες επισημείωσης πιστότητας και γραμματικότητας [Callison-Burch, Ch. κ.ά., 2007].

3.2.2 Μέθοδοι αξιολόγησης χωρίς άμεσα εκπεφρασμένη αξιολογική κρίση

Σε αυτή την κατηγορία τεχνικών μη αυτοματοποιημένης αξιολόγησης υπάγονται μέθοδοι στις οποίες επεμβαίνει η ανθρώπινη κρίση έμμεσα. Αυτό μπορεί να γίνει είτε μέσω της χρήσης των λεγόμενων ημι-αυτόματων μετρικών, είτε μέσω της εκτέλεσης εργασιών που προϋποθέτουν την κατανόηση του μηχανικά μεταφρασμένου κειμένου, είτε μέσω της ταξινόμησης, ανάλυσης και διόρθωσης λαθών των μεταφρασμάτων. Σε όλες αυτές τις περιπτώσεις ο άνθρωπος χρησιμοποιεί την κρίση του αλλά δεν εκφέρει άμεσα αξιολογική κρίση για το μετάφρασμα ή το σύστημα, όπως είδαμε ότι συμβαίνει στις προαναφερθείσες μεθόδους. Θεωρούμε ότι σε αυτές τις τεχνικές η αξιολόγηση είναι έμμεση.

3.2.2.1 Ημι-αυτόματες μετρικές

Οι ημι-αυτόματες μετρικές (human-in-the-loop evaluation) αποτελούν παραλλαγές αυτόματων μετρικών με επέμβαση επισημειωτών, όπως στις περιπτώσεις των HTER, HBLEU και HMETEOR [Snover, M. κ.ά., 2006]. Για να γίνει κατανοητό πώς λειτουργούν αυτές οι μετρικές, ακολουθεί μια σύντομη παρουσίαση του HTER.

HTER

Η μετρική HTER (human-targeted TER) [Snover, M. κ.ά., 2006] αποτελεί παραλλαγή του δείκτη TER, για τον οποίο έγινε λόγος στο Κεφάλαιο 3.1.1.1. Με τον δείκτη TER (Translation Edit Rate) υπολογίζεται ο αριθμός των ενεργειών διόρθωσης που απαιτούνται σε ένα μηχανικό μετάφρασμα προκειμένου αυτό να ταυτιστεί με μια μετάφραση-αναφορά. Η παραλλαγή έγκειται στο ότι στον HTER ο επισημειωτής δημιουργεί μια καινούρια μετάφραση-αναφορά με τις ελάχιστες δυνατές διορθώσεις είτε στο μετάφρασμα είτε στη μετάφραση-αναφορά, με στόχο αυτή η καινούρια μετάφραση να είναι γραμματική και να μεταδίδει το νόημα της δοθείσας μετάφρασης-αναφορά. Συγκεκριμένα οι ενέργειες διόρθωσης στον TER είναι οι εξής: εισαγωγή, απαλοιφή, αντικατάσταση μιας λέξης και μετακινήσεις ακολουθιών λέξεων, όλες με το ίδιο κόστος, και ορίζεται ως ο λόγος του αριθμού ενεργειών διόρθωσης προς τον μέσο αριθμό λέξεων στις μεταφράσεις-αναφορά. Στον HTER χρησιμοποιούνται επισημειωτές οι οποίοι είναι ικανοί ομιλητές της γλώσσας-στόχος και μία ή περισσότερες μεταφράσεις-αναφορά. Οι επισημειωτές προβαίνουν στις ελάχιστες ενέργειες διόρθωσης, όπως προαναφέρθηκε, δημιουργώντας μια καινούρια μετάφραση. Τέλος, γίνεται υπολογισμός του ελάχιστου TER χρησιμοποιώντας μόνο αυτή την καινούρια μετάφραση, η οποία όμως δεν λαμβάνεται υπόψη στον υπολογισμό του μέσου αριθμού λέξεων των μεταφράσεων-αναφορά. Η μέθοδος αυτή, αν και πολυέξοδη, παρουσίασε πολύ υψηλή συσχέτιση με ανθρώπινες κρίσεις και γενικότερα θεωρείται λιγότερο υποκειμενική από τις καθαρά ανθρώπινες κρίσεις, παρόλο που παρατηρούνται διαφορές που οφείλονται στην εκπαίδευση, τις δεξιότητες και την προσπάθεια των επισημειωτών. Οι κριτικές βεβαίως δεν λείπουν, καθώς έχει υποστηριχτεί ότι δεν αντανακλά την προσπάθεια που καταβάλλει ο επισημειωτής, όπως στην περίπτωση που μπαίνει στη διαδικασία να διορθώσει κάτι, πραγματοποιεί κάποιες αλλαγές, τις οποίες όμως τελικά αναιρεί, με αποτέλεσμα να έχει καταβάλει προσπάθεια χωρίς αυτό να αντικατοπτρίζεται στις ενέργειες διόρθωσης που καταγράφηκαν [Lacruz, I. κ.ά., 2014].

3.2.2.2 Αξιολόγηση μέσω εκτέλεσης εργασίας (task-based)

Για να αξιολογηθεί ένα σύστημα μετάφρασης, μπορεί να κριθεί η αποτελεσματικότητα του μεταφράσματος σε σχέση με μια συγκεκριμένη εργασία. Συγκεκριμένα παραδείγματα τέτοιου είδους αξιολόγησης που έχουν πραγματοποιηθεί είναι τα εξής: α) ζητήθηκε από ανθρώπους να εντοπίσουν τις βασικές πληροφορίες του κειμένου και β) ζητήθηκε από ανθρώπους να απαντήσουν ερωτήσεις περιεχομένου πάνω στο μετάφρασμα [Sanders, G. κ.ά., 2011]. Με αυτή τη μέθοδο η αξιολόγηση του συστήματος γίνεται με ιδιαίτερα έμμεσο τρόπο. Εξετάζεται το κατά πόσο αποδίδονται συγκεκριμένες έννοιες του πρωτότυπου κειμένου και φυσικά δεν κρίνεται σε μεγάλο βαθμό η ποιότητα του λόγου, εκτός από τον βαθμό στον οποίο αυτή επηρεάζει την κατανόηση. Μπορεί ένα κείμενο να λάβει χαμηλή αξιολόγηση άλλου τύπου και υψηλή σε μια τέτοια μέθοδο αλλά και το αντίστροφο [Bonnie, D., κ.ά., 2011]. Το ζητούμενο βέβαια είναι τι είδους αξιολόγηση χρειαζόμαστε κάθε φορά[31].

3.2.2.3 Ταξινόμηση και ανάλυση λαθών – Μετεπεξεργασία (post-editing)

Τέλος, θα δούμε συγκεκριμένες μεθόδους μη αυτοματοποιημένης αξιολόγησης στις οποίες χρησιμοποιείται ανάλυση, ταξινόμηση και διόρθωση λαθών. Θα παρουσιάσουμε κάποιες από τις πιο διαδεδομένες τυπολογίες λαθών και τα εργαλεία επισημείωσης που προσφέρονται για την ταξινόμηση και τη διόρθωση λαθών.

3.2.2.3.1 MQM

Μία από τις πλέον εξελιγμένες μετρικές τέτοιου τύπου είναι η MQM (Multidimensional Quality Metrics – Πολυδιάστατη μετρική ποιότητας). Αυτό το σύστημα μέτρησης έχει αναπτυχθεί από το QTLaunchPad[32], στα πλαίσια του οποίου ορίζεται η ποιότητα της μηχανικής ή μη μετάφρασης και η διαδικασία μέτρησής της με ομοιογενή πρότυπα. Για την αξιολόγηση ποιότητας μετάφρασης βάσει MQM ακολουθείται η εξής διαδικασία: εντοπισμός χαρακτηριστικών που διαδραματίζουν κάποιο ρόλο στη μετάφραση (παράμετροι/διαστάσεις), επιλογή κατηγοριών λαθών που ανταποκρίνονται στα χαρακτηριστικά των διαστάσεων και επισημείωση στο εργαλείο translate5- εξαγωγή βαθμολογίας.

[31] Αυτού του είδους η αξιολόγηση αναφέρεται κυρίως στη λεγόμενη “gisting translation”, στη μετάφραση δηλαδή όπου αναφέρονται μόνο τα βασικά νοήματα του πρωτότυπου, ενώ στην παρούσα εργασία επικεντρωνόμαστε στον στόχο του μηχανικού μεταφράσματος το οποίο αποτελεί πλήρη απόδοση του πρωτότυπου και μπορεί να χρησιμοποιηθεί ως έχει.

[32] Συνεργατική ερευνητική πρωτοβουλία χρηματοδοτούμενη από την Ευρωπαϊκή Επιτροπή για το διάστημα 2012-2014 με στόχο την προετοιμασία για ένα νέο είδος έρευνας και συνεργασίας μεταξύ κατασκευαστών συστημάτων μηχανικής μετάφρασης και επιχειρήσεων του τομέα της γλώσσας, ώστε να ξεπεραστούν τα εμπόδια ποιότητας στη μηχανική και μη μετάφραση με τη χρήση γλωσσικής τεχνολογίας. Όλες οι πληροφορίες σχετικά με την εδώ αναφερόμενη επισημείωση βρίσκεται στην ιστοσελίδα του (http://www.qt21.eu/launchpad/, τελευταία προσπέλαση 19/06/2015).

Παράμετροι MQM

Στην τελευταία ενημέρωση της ιστοσελίδας του QTLaunchPad[33], ορίζονται δώδεκα παράμετροι ή αλλιώς “διαστάσεις” που πρέπει να λάβει κανείς υπόψη πριν ξεκινήσει την αξιολόγηση:

1) Γλώσσα/συνδυασμός γλώσσας με γεωγραφική τοποθεσία (locale): η γλώσσα ή η γλώσσα και η γεωγραφική τοποθεσία στην οποία απευθύνεται το κείμενο (π.χ. κείμενο για χρήση από γαλλόφωνους αναγνώστες στον Καναδά).
2) Πεδίο: ο τομέας του επιστητού στον οποίο ανήκει το κείμενο (π.χ. δίκαιο, φαρμακολογία).
3) Ορολογία: ορολογικοί πόροι ή συμβάσεις που ισχύουν για το πρωτότυπο ή το μεταφρασμένο κείμενο. Η ορολογία είναι ένα από τα χαρακτηριστικά της λεγόμενης “ειδικής γλώσσας”, η οποία δεν είναι παρά ένα υποσύνολο ενός συνόλου, της φυσικής γλώσσας δηλαδή [Cabré, T., 1999]. Υπάρχουν διάφορα στοιχεία που καθορίζουν μια ειδική γλώσσα, όπως το θέμα, η πρόθεση, οι συνθήκες, το λεξιλόγιο, η σύνταξη και η επικοινωνία μεταξύ ειδικών του ίδιου ή συναφών πεδίων [Cabré, T., 1999], σημαντικό κριτήριο είναι όμως η ανάγκη ύπαρξης ειδικών γνώσεων για την κατανόησή της [Lerat, P., 1995]. Για τους όρους, όπως και για την ειδική γλώσσα, έχουν διατυπωθεί πλείστοι ορισμοί (σύμφωνα με τον Διεθνή Οργανισμό Τυποποίησης (ISO) είναι η “απόδοση μιας συγκεκριμένης έννοιας μιας εξειδικευμένης γλώσσας σε μια γλωσσική έκφραση”, όπου *έννοια* είναι “...το νοητικό κατασκεύασμα που στοχεύει στην ταξινόμηση μεμονωμένων αντικειμένων του εσωτερικού και του εξωτερικού κόσμου...” [ISO704, 1986]) και δεν είναι πάντα εύκολη η διάκριση μεταξύ λέξης και όρου, μπορούμε όμως να επικεντρωθούμε στην εξής διαφορά μεταξύ τους: “η διαφορά [τους] έγκειται στο ότι οι λέξεις δεν χρειάζεται να ορισθούν για να γίνουν αντιληπτές από τους φυσικούς ομιλητές μιας γλώσσας, ενώ οι όροι αν και είναι λεξικές μονάδες δηλώνουν συγκεκριμένες έννοιες στο πλαίσιο ενός γνωστικού τομέα” [Φωτοπούλου, Α., 2003].
4) Κειμενικό είδος (π.χ. μυθιστόρημα, τεχνικό εγχειρίδιο χρήσης, διαφήμιση, κ.λπ.). Κειμενικά είδη, κατά Bakhtin [1986], είναι οι σχετικά σταθεροί τύποι εκφωνημάτων που αναπτύσσονται σε κάθε τομέα του επιστητού, στα οποία εκφωνήματα εκφράζονται οι συγκεκριμένες συνθήκες και στόχοι του κάθε τομέα μέσω του θέματος, του γλωσσικού ύφους (λεξιλογικές, φρασεολογικές και γραμματικές επιλογές) και κυρίως της συνθετικής δομής τους. Ως παραδείγματα αναφέρει μεταξύ άλλων σύντομες απαντήσεις καθημερινών διαλόγων, καθημερινές αφηγήσεις, σύντομες τυποποιημένες στρατιωτικές διαταγές, καθώς και όλη την γκάμα επιχειρησιακών εγγράφων και κοινωνικών ή πολιτικών σχολιασμών.
5) Κοινό στο οποίο απευθύνεται η μετάφραση (π.χ. χρήστες ενός συγκεκριμένου μοντέλου πλυντηρίου).

[33] 21/11/2014, http://www.qt21.eu/launchpad/content/dimensions

6) Σκοπός[34]: ο στόχος του κειμένου (π.χ. πληροφόρηση αναγνώστη σχετικά με τα δικαιώματά του βάσει προτεινόμενου νόμου).
7) Γλωσσικό μητρώο: το γλωσσικό μητρώο που πρέπει να χρησιμοποιηθεί στη γλώσσα-στόχος. Πρόκειται για το επίπεδο επισημότητας της γλώσσας. Ο Newmark [1988] όρισε 8 επίπεδα σε ένα συνεχές από το πιο επίσημο στο πιο ανεπίσημο.
8) Ύφος: "Ο ιδιαίτερος τρόπος με τον οποίο χρησιμοποιεί κάποιος τη γλώσσα στον προφορικό ή στο γραπτό λόγο ανάλογα με την περίσταση επικοινωνίας", σύμφωνα με το Λεξικό της Κοινής Νεοελληνικής του Τριανταφυλλίδη (π.χ. εάν πρέπει να χρησιμοποιηθεί συγκεκριμένος οδηγός ή να χρησιμοποιηθούν/αποφευχθούν συγκεκριμένα υφολογικά χαρακτηριστικά).
9) Ανταπόκριση περιεχομένου: κατά πόσο θα έπρεπε να ανταποκρίνεται το μετάφρασμα στο πρωτότυπο (π.χ. περιληπτική μετάφραση).
10) Τρόπος παρουσίασης μεταφράσματος (π.χ. σε μορφή υπότιτλων, σε μορφή ηχητικής ειδοποίησης, κ.λπ.).
11) Τύπος αρχείου (π.χ. html).
12) Τεχνολογία παραγωγής: η τεχνολογία και το λογισμικό που χρησιμοποιούνται στη μεταφραστική διαδικασία (π.χ. μηχανική μετάφραση, μεταφραστική μνήμη, κ.λπ.).

Τυπολογία λαθών MQM

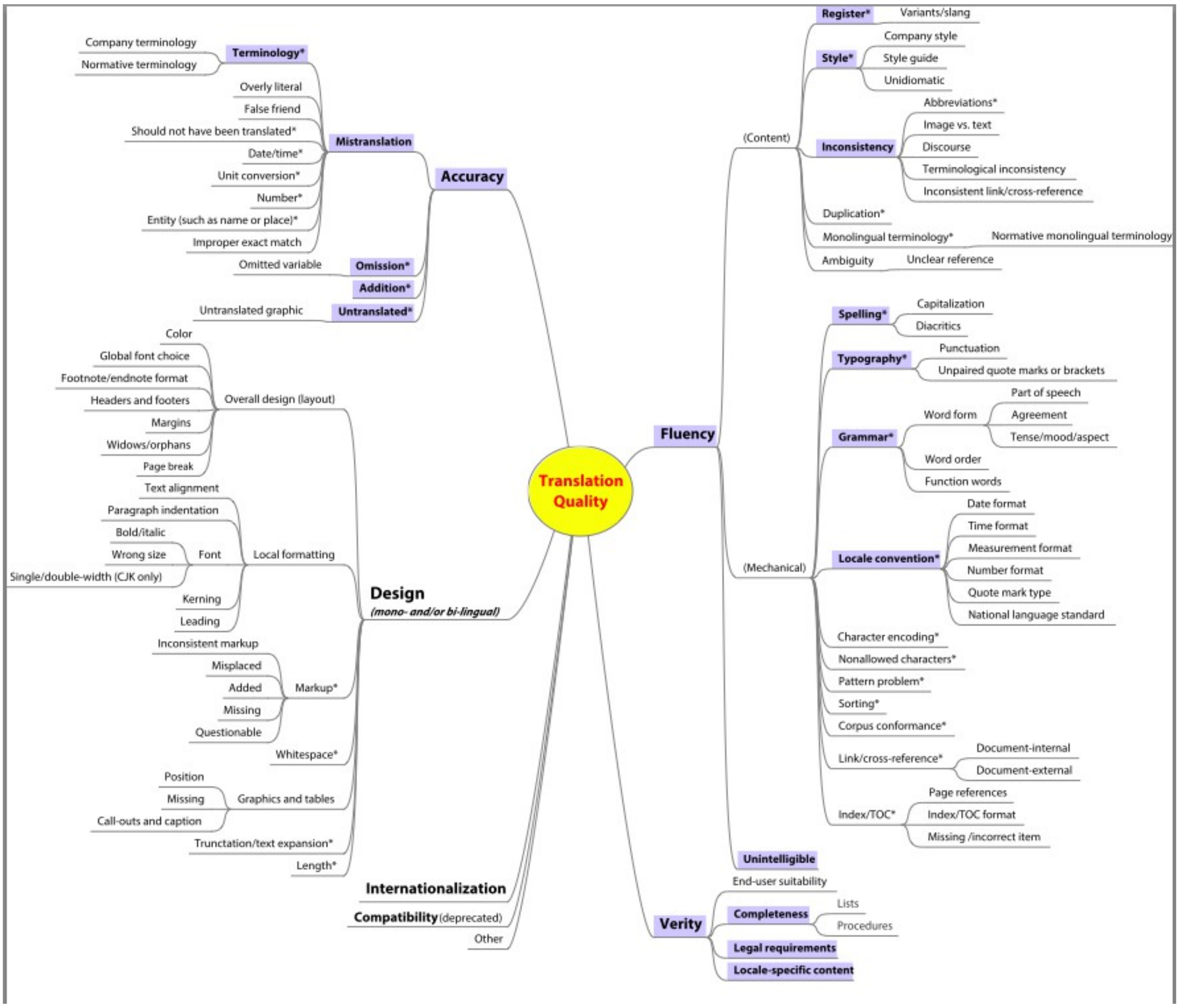
Αφού απαντηθούν οι ερωτήσεις σχετικά με τις παραπάνω διαστάσεις και σε συνάρτηση με αυτές, ο χρήστης επιλέγει ποιες κατηγορίες λαθών από τις υπάρχουσες θα χρησιμοποιήσει. Στην τυπολογία λαθών MQM προβλέπεται μια σειρά από κατηγορίες και υποκατηγορίες πιθανών λαθών (issues), οι οποίες εφαρμόζονται στη μετάφραση αλλά μπορούν να εφαρμοστούν και στο πρωτότυπο, δεδομένου ότι η ποιότητα της μετάφρασης εξαρτάται σε μεγάλο βαθμό από την ποιότητα του πρωτότυπου. Ο επισημειωτής, όπως θα δούμε στη συνέχεια, επιλέγει το απόσπασμα όπου θεωρεί ότι υπάρχει πρόβλημα (δεν περιορίζεται σε μία μόνο λέξη δηλαδή) και μια όσο το δυνατόν πιο συγκεκριμένη υποκατηγορία για να το κατατάξει. Σε περίπτωση που ένα φαινόμενο δεν μπορεί να ενταχθεί σε μια από τις υποκατηγορίες, εντάσσεται στην υπερκατηγορία.

Για τις ανάγκες μιας αξιολόγησης όπως αυτή της παρούσας εργασίας θα επιλέγαμε τις κατηγορίες που εμφανίζονται στο ακόλουθο ιεραρχικό σχήμα[35]. Μέσα στο σχήμα αναφέρονται οι κατηγορίες στα αγγλικά και στη συνέχεια παρατίθενται ορισμοί και παραδείγματα για την κάθε κατηγορία.

[34] Βλ. θεωρία του Σκοπού [Vermeer, H. J., 2000]
[35] Για την πλήρη λίστα των κατηγοριών, βλ. http://www.qt21.eu/launchpad/content/list-mqm-issue-types (τελευταία ενημέρωση 21/11/2014).

- Accuracy
    - Mistranslation
        - Terminology
        - Overly literal
        - False friend
        - Should not have been translated
    - Omission
    - Untranslated
    - Addition
- Fluency
    - Mechanical
        - Typography
            - Punctuation
        - Grammar
            - Word form
                - Part of speech
                - Agreement
                - Tense/mood/aspect
            - Word order
            - Function words
    - Unintelligible

Η πλήρης λίστα απεικονίζεται και ως εξής:



Εικόνα 7: Πλήρης τυπολογία λαθών MQM

Ορισμοί

Στη συνέχεια παρουσιάζονται οι ορισμοί με παραδείγματα από το δικό μας σώμα κειμένων. Παραδείγματα από κατηγορίες και υποκατηγορίες τους, όπως είναι αναμενόμενο, συμπίπτουν.

Accuracy	Πιστότητα
Ορισμός	Ο βαθμός στον οποίο το κείμενο-στόχος εκφράζει με ακρίβεια το νόημα του κειμένου-πηγή.
Γονέας	-
Παιδιά	Λανθασμένη Μετάφραση, Παράλειψη, Αμετάφραστο, Προσθήκη

Παράδειγμα	“2. Can IBS **lead** to cancer?” → “**μόλυβδος** 2.Can IBS στον καρκίνο;”
Σε αυτό το παράδειγμα η λέξη “lead” είναι ρήμα και σημαίνει “οδηγώ”, μεταφράζεται όμως σαν να επρόκειτο για άλλο λήμμα, το ουσιαστικό “lead”, που σημαίνει “μόλυβδος”, οπότε και δεν μεταφέρεται η σημασία του πρωτότυπου στη μετάφραση. Εντάσσεται στην υποκατηγορία Λανθασμένη Μετάφραση.	

Mistranslation	Λανθασμένη Μετάφραση
Ορισμός	Το περιεχόμενο του κειμένου-στόχος δεν ανταποκρίνεται ακριβώς στο περιεχόμενο του κειμένου-πηγή.
Γονέας	Accuracy
Παιδιά	Ορολογία, Καταχρηστική Κατά Λέξη Μετάφραση, Ψευδόφιλη Λέξη, Κείμενο που δεν θα έπρεπε να έχει μεταφραστεί
Παράδειγμα	“To provide a diagnosis for IBS, the doctor will first order specific **tests** to rule out conditions with similar symptoms to those of the irritable bowel syndrome and to check whether you meet the Rome III criteria.” → “Να παρέχουν μια διάγνωση για το σύνδρομο ευερέθιστου Εντέρου, ο γιατρός θα διατάξει πρώτα συγκεκριμένες **δοκιμές**, να αποκλείσει όρους με παρόμοια συμπτώματα με αυτά του συνδρόμου του ευερέθιστου εντέρου και να ελέγξει κατά πόσον πληροίτε τα κριτήρια της Ρώμης III.”
Σε αυτή την περίπτωση αποδίδεται η λέξη του πρωτότυπου με μία από τις πιθανές λέξεις απόδοσης στην ελληνική, η οποία όμως δεν είναι ορθή βάσει συγκειμένου.	

Terminology	Ορολογία
Ορισμός	Μεταφράζεται κάποιος όρος με όρο που δεν είναι ο αναμενόμενος για το πεδίο.
Γονέας	Λανθασμένη Μετάφραση
Παιδιά	-
Παράδειγμα	“4. Do I have a food **intolerance**?” → “4.Do έχω μια **αδιαλλαξία** τροφίμων;”
Πρόκειται για πολυλεκτικό όρο (“food intolerance”) ο οποίος δεν αποδίδεται ως τέτοιος αλλά το κάθε στοιχείο του ξεχωριστά.	

Overly literal	Καταχρηστική Κατά Λέξη Μετάφραση
Ορισμός	Η κατά λέξη μετάφραση αποτελεί μία από τις άμεσες τεχνικές της μετάφρασης κατά Vinay, J. P. και Darbelnet, J. [1958], κατά την οποία το μήνυμα μεταφέρεται από τη μια γλώσσα στην άλλη μεταφράζοντας ακριβώς τις λέξεις του πρωτότυπου στη γλώσσα-στόχος, χωρίς να προσκρούει ούτε στο γλωσσικό αισθητήριο του φυσικού ομιλητή της γλώσσας-στόχος ούτε στις συντακτικές δομές της, οι οποίες, στην προκειμένη περίπτωση, συμπίπτουν με αυτές της γλώσσας αφετηρίας, ούτε στους κανόνες και στους εκφραστικούς τρόπους της γλώσσας-στόχος. Όσο πιο συγγενείς γενετικά είναι μεταξύ τους δυο γλώσσες και όσο μεγαλύτερη είναι η αλληλεπίδρασή τους, τόσο πιο εύκολη και συχνή είναι και η χρήση της εν λόγω τεχνικής. Η καταχρηστική κατά λέξη μετάφραση υπερβαίνει αυτά τα όρια, συνεπώς προσκρούει στο γλωσσικό αισθητήριο, στις

	συντακτικές δομές ή στους κανόνες και στους εκφραστικούς τρόπους της γλώσσας-στόχος. Παύει να αποτελεί μεταφραστική τεχνική και γίνεται στοιχείο προς διόρθωση.
Γονέας	Λανθασμένη Μετάφραση
Παιδιά	-
Παράδειγμα	"Colonoscopy is performed to rule out other conditions but only **on the ground of** indications." → "Η κολονοσκόπηση εκτελείται για να αποκλείσει άλλους όρους αλλά μόνο **στο έδαφος των** ενδείξεων."
Ιδιωτισμός ο οποίος μεταφράζεται κατά λέξη χωρίς να υπάρχει αντιστοιχία στη γλώσσα-στόχος.	

False friend	Ψευδόφιλη Λέξη
Ορισμός	Ψευδόφιλες λέξεις ή μονάδες είναι οι λέξεις δύο διαφορετικών γλωσσών των οποίων οι φωνολογικές μορφές συμπίπτουν ή είναι παρεμφερείς αλλά το νόημα των οποίων διαφέρει [Αναστασιάδη-Συμεωνίδη, Α. και Βλέτση, Ε., 2009]
Γονέας	Λανθασμένη Μετάφραση
Παιδιά	-
Παράδειγμα	-
Για αποσαφήνιση της έννοιας παρέχουμε κάποια παραδείγματα ψευδόφιλων λέξεων από το ζεύγος γλωσσών αγγλικά-ισπανικά: embarrassed (ντροπιασμένος, αμήχανος) / embarazada (έγκυος), parent (γονέας) / pariente (συγγενής). Στη δική μας έρευνα δεν προέκυψαν τέτοιου είδους λάθη, όπως είναι και αναμενόμενο για τη μηχανική μετάφραση, δεν αποκλείουμε όμως το ενδεχόμενο σε ένα μεγαλύτερο σώμα κειμένων να υπήρχαν κάποιες εμφανίσεις λαθών λόγω ψευδόφιλων λέξεων.	

Should not have been translated	Κείμενο που δεν θα έπρεπε να έχει μεταφραστεί
Ορισμός	Έχει μεταφραστεί κείμενο που δεν θα έπρεπε να έχει μεταφραστεί.
Γονέας	Λανθασμένη Μετάφραση
Παιδιά	-
Παράδειγμα	"**Vertex** Announces Presentation of New KALYDECO™ (ivacaftor) Data at European Cystic Fibrosis Society Conference" → "**κορυφή** ανακοινώνει την παρουσίαση των νέων kalydeco ™ (ivacaftor) δεδομένων σε ευρωπαϊκή κοινωνία της διάσκεψης κυστική ίνωση"
Στην προκειμένη περίπτωση μεταφράζεται επωνυμία εταιρείας, η οποία θα έπρεπε να παραμείνει ως έχει.	

Omission	Παράλειψη

Ορισμός	Λείπει περιεχόμενο από τη μετάφραση το οποίο υπάρχει στο κείμενο-πηγή.
Γονέας	Πιστότητα
Παιδιά	-
Παράδειγμα	“"People with cystic fibrosis **experience** significant and progressive loss of lung function during their lives.” → “"οι άνθρωποι με κυστική ίνωση, η σημαντική και τη σταδιακή απώλεια της πνευμονικής λειτουργίας κατά τη διάρκεια της ζωής τους.”
Η έννοια του “experience” δεν εμφανίζεται με κανέναν τρόπο στη μετάφραση· για την ακρίβεια στη μεταφρασμένη πρόταση δεν υπάρχει ρήμα.	

Untranslated	Αμετάφραστο
Ορισμός	Περιεχόμενο που θα έπρεπε να είχε μεταφραστεί αλλά δεν μεταφράστηκε.
Γονέας	Πιστότητα
Παιδιά	-
Παράδειγμα	“Ηλεκτροφυσιολογικό έλεγχο (διοισοφάγιο και διαφλέβιο).” → “**Ilektrofysiologiko** control (**dioisofagio** and **diaflebio**).”
Σε αυτή την πρόταση υπάρχουν τρεις λέξεις που έμειναν αμετάφραστες. Στα εμπορικά συστήματα πραγματοποιείται μεταγραφή στο λατινικό αλφάβητο, ωστόσο η λέξη παραμένει αμετάφραστη.	

Addition	Προσθήκη
Ορισμός	Στο κείμενο-στόχος υπάρχει κείμενο που δεν υπάρχει στο κείμενο-πηγή.
Γονέας	Πιστότητα
Παιδιά	-
Παράδειγμα	“Αναίμακτο εργαστηριακό έλεγχο με ηλεκτροκαρδιογράφημα (ΗΚΓ), υπερηχοκαρδιογράφημα, Holter 24ωρης καταγραφής ρυθμού, δοκιμασία κοπώσεως, δοκιμασία ανακλίσεως (tilt-test).” → “**And as** bloodless **coastal** laboratory testing with electrocardiogram (ECG), a yperichokardiografi, a 24-hour Holter recording rate regulation, stress test, test anaklisews (tilt-test).”
Σε αυτή τη μεταφρασμένη πρόταση παρατηρούνται τρεις λέξεις που έχουν προστεθεί, ενώ δεν υπήρχαν στο πρωτότυπο.	

Fluency	Γραμματικότητα
Ορισμός	Λάθη που μπορούν να εντοπιστούν ανεξάρτητα από το αν το κείμενο στο οποίο βρίσκονται είναι μεταφρασμένο, όπως π.χ. ορθογραφικά λάθη.
Γονέας	-

Παιδιά	Θέμα Μηχανικό ή Μορφής, Ακατανόητο
Παράδειγμα	“... μπορούμε να ελαχιστοποιήσουμε την ταλαιπωρία και να βελτιώσουμε **τη** ποιότητα ζωής.”
Έλλειψη τελικού νι. Υπάγεται στην υποκατηγορία Θέμα Μηχανικό ή Μορφής, σε μεγαλύτερο βάθος στη Γραμματική και τέλος, πιο συγκεκριμένα, στη Μορφολογία.	

Mechanical	Θέμα Μηχανικό ή Μορφής
Ορισμός	Λάθη που δεν επηρεάζουν το λεξικο-σημασιολογικό επίπεδο, επηρεάζουν όμως τη μορφή του κειμένου ή/και της γλώσσας.
Γονέας	Γραμματικότητα
Παιδιά	Τυπογραφικό, Γραμματική
Παράδειγμα	“Τώρα γνωρίζουμε ότι υπάρχουν πάρα πολλά μονοπάτια**,** μεταξύ του εντέρου και του εγκεφάλου (άξονας εγκέφαλος του στομάχου).”
Λανθασμένη χρήση κόμματος. Υπάγεται στην υποκατηγορία Τυπογραφικό και πιο συγκεκριμένα στην υποκατηγορία Στίξη.	

Typography	Τυπογραφικό
Ορισμός	Τυπογραφικά λάθη που δεν αφορούν την ορθογραφία της λέξης αλλά θέματα όπως η στίξη.
Γονέας	Θέμα Μηχανικό ή Μορφής
Παιδιά	Στίξη
Παράδειγμα	“Τώρα γνωρίζουμε ότι υπάρχουν πάρα πολλά μονοπάτια**,** μεταξύ του εντέρου και του εγκεφάλου (άξονας εγκέφαλος του στομάχου).”
Λανθασμένη χρήση κόμματος. Υπάγεται στην υποκατηγορία Στίξη.	

Punctuation	Στίξη
Ορισμός	Λάθος στίξης
Γονέας	Τυπογραφικό
Παιδιά	-
Παράδειγμα	“Τώρα γνωρίζουμε ότι υπάρχουν πάρα πολλά μονοπάτια**,** μεταξύ του εντέρου και του εγκεφάλου (άξονας εγκέφαλος του στομάχου).”
Λανθασμένη χρήση κόμματος.	

Grammar	Γραμματική
Ορισμός	Γραμματικά ή συντακτικά λάθη εκτός από ορθογραφικά.
Γονέας	Θέμα Μηχανικό ή Μορφής
Παιδιά	Μορφολογία, Σειρά Λέξεων, Λειτουργικές Λέξεις
Παράδειγμα	“5. **μου κοιλιά** είναι πρησμένο και νιώθω άβολα.”
Λάθος που υπάγεται στη Σειρά Λέξεων.	

Word form	Μορφολογία
Ορισμός	Μορφολογικά λάθη, στα οποία συμπεριλαμβάνεται και η συμφωνία.
Γονέας	Γραμματική
Παιδιά	Μέρος του λόγου, Συμφωνία, Χρόνος/Έγκλιση/Όψη
Παράδειγμα	“"όχι μόνο kalydeco **οδηγήσει** σε σημαντική αρχική βελτίωση, αλλά τα νέα αυτά δεδομένα μακροχρόνιας έδειξε ότι είναι επίσης, φάνηκε να τροποποιήσει το ποσοστό της πνευμονικής λειτουργίας, η μείωση της απώλειας της πνευμονικής λειτουργίας σε άτομα με g551d μετάλλαξης.”
Λανθασμένη όψη του ρήματος. Υπάγεται στην υποκατηγορία Χρόνος/Έγκλιση/Όψη.	

Part of speech	Μέρος του Λόγου
Ορισμός	Λανθασμένο μέρος του λόγου.
Γονέας	Μορφολογία
Παιδιά	-
Παράδειγμα	“Αυτή τη στιγμή μόνο το 50% των καταναλωτών στο σε απευθείας σύνδεση **κατάστημα** της Ε.Ε..”
Σε θεωρητικό επίπεδο η διάκριση μεταξύ πιστότητας και γραμματικότητας είναι απόλυτη, όπως είναι και απόλυτο το κριτήριο του αν το λάθος γίνεται αντιληπτό εξετάζοντας μόνο το μετάφρασμα ή και το πρωτότυπο. Μάλιστα αυτό ακολουθεί ο επισημειωτής κατά την εκτέλεση της αξιολόγησης για την επιλογή κατηγορίας. Ωστόσο στην πράξη υπάρχουν μηχανικά μεταφρασμένες προτάσεις με τόσο πολλά λάθη που δεν είναι εύκολο να απομονωθούν τα λάθη αυτών των δύο υπερκατηγοριών. Στο συγκεκριμένο παράδειγμα ενδέχεται ο μέσος αναγνώστης μόνο με την ανάγνωση της μετάφρασης να μην εντόπιζε λάθος μέρους του λόγου στη λέξη “κατάστημα”. Η πρωτότυπη πρόταση είναι: “At present only 50% of consumers in the EU shop online.” Η λέξη “shop” χρησιμοποιείται ως ρήμα αλλά μεταφράζεται ως ουσιαστικό. Αυτό δεν αναιρεί τη διάκρισή μας. Απλώς αυτή θα γινόταν πιο κατανοητή, αν εξετάζαμε την ίδια πρόταση μεταφρασμένη από άνθρωπο αντικαθιστώντας το ρήμα που επέλεξε ο μεταφραστής με το ουσιαστικό “κατάστημα”. Μία απόδοση μεταφραστή είναι: “Αυτή τη στιγμή μόνο 50% των καταναλωτών της Ε.Ε. **κάνουν αγορές** στο διαδίκτυο.” → “Αυτή τη στιγμή μόνο 50% των καταναλωτών της Ε.Ε. **κατάστημα** στο διαδίκτυο.”, όπου και είναι πιο ευδιάκριτο το λανθασμένο	

μέρος του λόγου.

Agreement	Συμφωνία
Ορισμός	Δύο ή περισσότερες λέξεις δεν συμφωνούν σε πτώση, αριθμό, πρόσωπο ή άλλα γραμματικά χαρακτηριστικά.
Γονέας	Μορφολογία
Παιδιά	-
Παράδειγμα	“5. μου **κοιλιά** είναι **πρησμένο** και νιώθω άβολα.”
Ασυμφωνία γένους μεταξύ ουσιαστικού και επιθέτου.	

Tense/mood/ aspect	Χρόνος/Έγκλιση/Όψη
Ορισμός	Λάθος σε χρόνο, έγκλιση, όψη ρηματικού τύπου.
Γονέας	Μορφολογία
Παιδιά	-
Παράδειγμα	“Πολλοί IBS ασθενείς **αναπτύξουν** δυσανεξία σε ορισμένες τροφές.”
Λανθασμένη όψη ρηματικού τύπου.	

Word order	Σειρά Λέξεων
Ορισμός	Λανθασμένη σειρά λέξεων
Γονέας	Γραμματική
Παιδιά	-
Παράδειγμα	“5. **μου κοιλιά** είναι πρησμένο και νιώθω άβολα.”
Η κτητική αντωνυμία προηγείται του ουσιαστικού που χαρακτηρίζει, ενώ θα έπρεπε να έπεται.	

Function words	Λειτουργικές Λέξεις
Ορισμός	Λανθασμένη χρήση λειτουργικής λέξης, όπως πρόθεση, βοηθητικό ρήμα και άρθρο.
Γονέας	Γραμματική

Παιδιά	-
Παράδειγμα	“η μέση απόλυτη, από την έναρξη της θεραπείας, της πνευμονικής λειτουργίας, όπως μετρήθηκε από 45, **από** 24 εβδομάδες θεραπείας ήταν 13,5 ποσοστιαίες μονάδες (n = 18).”
Λανθασμένη πρόθεση (“από” αντί για “σε”). Ισχύει το ίδιο που αναφέρθηκε στο παράδειγμα της κατηγορίας “Μέρος του λόγου”. Πρόκειται για μια πρόταση με τόσο πολλά λάθη που είναι δύσκολο να εντοπιστεί το λάθος στην πρόθεση ως τέτοιο. Κατά τη διαδικασία της αξιολόγησης όμως, καθώς ο επισημειωτής προβαίνει και σε διόρθωση της πρότασης[36], αντιλαμβάνεται τον χαρακτήρα του λάθους.	

Unintelligible	Ακατανόητο
Ορισμός	Δεν μπορεί να προσδιοριστεί ο ακριβής χαρακτήρας του λάθους και δεν επιτυγχάνεται ο επικοινωνιακός στόχος. Πρόκειται για κείμενο που δεν μπορεί να γίνει κατανοητό ούτε στο ελάχιστο.
Γονέας	Γραμματικότητα
Παιδιά	-
Παράδειγμα	Σε αγγλική μετάφραση γερμανικού εγχειριδίου αυτοκινήτου εμφανίζεται η ακόλουθη πρόταση: “The brake from whe this કુતારિ સ S149235 part numbr,,.”
Στο σώμα μας δεν εντοπίστηκε τέτοιου είδους λάθος, γι' αυτό και παραθέτουμε το παράδειγμα από τη λίστα του QTLaunchpad.	

Πίνακας 1: Τυπολογία λαθών MQM

Ο διαχωρισμός μεταξύ πιστότητας και γραμματικότητας, όπως ακριβώς περιγράφεται σε αυτή την τυπολογία, ανταποκρίνεται στο στάδιο της ανάπτυξης συστήματος μηχανικής μετάφρασης, οπότε σε επίπεδο αξιοποίησης αποτελεσμάτων γνωρίζουμε ότι τα λάθη της κατηγορίας “Πιστότητα” αφορούν το παράλληλο σώμα εκπαίδευσης του συστήματος και τα λάθη της κατηγορίας “Γραμματικότητα” αφορούν το γλωσσικό μοντέλο του συστήματος.

Επισημείωση σε translate5 και εξαγωγή βαθμολογίας

Το QTLaunchPad παρέχει ιδιαίτερα λεπτομερείς οδηγίες επισημείωσης, με παραδείγματα, συγκεκριμένες ιδιάζουσες περιπτώσεις, καθώς και έναν αλγόριθμο επιλογής κατηγορίας. Το εργαλείο που προσφέρεται για επισημείωση βάσει της παραπάνω τυπολογίας είναι το translate5 της εταιρείας MittagQI[37]. Πρόκειται για εργαλείο ανοιχτού κώδικα, το οποίο δέχεται αρχεία CSV

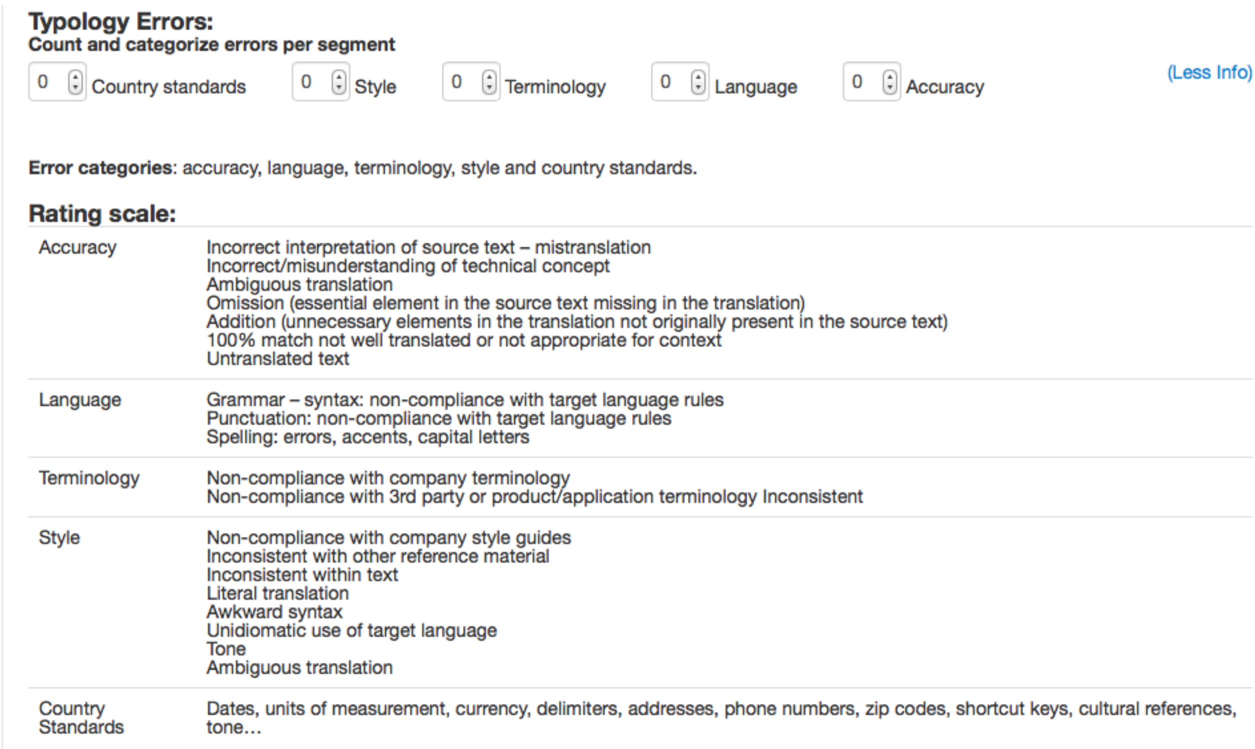
[36] Στο θέμα της ταξινόμησης λαθών με παράλληλη διόρθωση επανερχόμαστε στο Κεφάλαιο 4.2.1.
[37] http://www.mittagqi.com/en/ (τελευταία προσπέλαση 19/06/2015)

(comma-separated value) και εξάγει αποτελέσματα στην ίδια μορφή.

Η εξαγωγή βαθμολογίας δεν είναι απαραίτητη στην αξιολόγηση MQM. Ωστόσο παρέχεται μηχανισμός βαθμολόγησης για ομοιογένεια στα αποτελέσματα. Για την εξαγωγή βαθμολογίας κάθε μετάφρασης μετρώνται οι εμφανίσεις κάθε είδους λάθους. Για τον υπολογισμό της βαθμολογίας συνυπολογίζεται η σοβαρότητα του λάθους, ένα βάρος, το οποίο μπορεί να επιλεγεί, και ο αριθμός λέξεων του κειμένου. Τα επίπεδα σοβαρότητας του λάθους είναι τρία: δευτερεύον (minor: λάθος που δεν παρεμποδίζει την κατανόηση του κειμένου), σοβαρό (major: λάθος που καθιστά δυσνόητο το κείμενο) και κρίσιμο (critical: λάθος που καθιστά τη μετάφραση μη χρησιμοποιήσιμη). Με βάση τη σοβαρότητα του λάθους, το βάρος και τον αριθμό λέξεων, κάθε λάθος συνεπάγεται μια τιμή ποινής (penalty), η οποία αφαιρείται από το ιδανικό 100%. Αυτή η διαδικασία εκφράζεται με την εξής σχέση: $P = (\text{Issues}_{\text{minor}} + 10 \times \text{Issues}_{\text{major}} + 100 \times \text{Issues}_{\text{critical}}) \div \text{Word count}$, όπου P=ποινή, issues=λάθη και word count=αριθμός λέξεων, και minor, major και critical τα τρία προαναφερθέντα επίπεδα σοβαρότητας λάθους. Η ποινή δίνεται σε ποσοστό επί τοις εκατό. Οι ποινές κάθε υποκατηγορίας μιας κατηγορίας προστίθενται. Εάν έχει πραγματοποιηθεί αξιολόγηση και στο πρωτότυπο κείμενο, τότε αφαιρείται η ποινή του κειμένου-πηγή από την ποινή του κειμένου-στόχος. Η τελική βαθμολογία υπολογίζεται αφαιρώντας τις ποινές κάθε υπερκατηγορίας. Αυτό συνεπάγεται ότι εάν έχει αξιολογηθεί και το κείμενο-πηγή και έχει κριθεί ιδιαίτερα προβληματικό, υπάρχει η δυνατότητα συνολικής βαθμολογίας μεγαλύτερης από 100%. Σε αυτή την περίπτωση θεωρείται ότι το κείμενο έχει βελτιωθεί κατά τη διαδικασία της μετάφρασης.

3.2.2.3.2 DQF: Τυπολογία λαθών

Άλλη ευρέως χρησιμοποιούμενη τυπολογία λαθών για αξιολόγηση μετάφρασης είναι αυτή που προτείνεται στα πλαίσια του DQF, αν και έχει ήδη ξεκινήσει η διαδικασία εναρμόνισης DQF και MQM με στόχο να γεφυρωθεί το χάσμα μεταξύ των ορισμών και των προδιαγραφών των δύο μοντέλων [Görög, A., 2014]. Στην Εικόνα 8 φαίνονται οι κατηγορίες λαθών όπως εμφανίζονται κατά την επισημείωση αξιολόγησης. Πρόκειται για 5 κατηγορίες, η καθεμία από τις οποίες περιλαμβάνει αρκετά φαινόμενα.



Typology Errors:
Count and categorize errors per segment

0 Country standards 0 Style 0 Terminology 0 Language 0 Accuracy (Less Info)

Error categories: accuracy, language, terminology, style and country standards.

Rating scale:

Accuracy	Incorrect interpretation of source text – mistranslation Incorrect/misunderstanding of technical concept Ambiguous translation Omission (essential element in the source text missing in the translation) Addition (unnecessary elements in the translation not originally present in the source text) 100% match not well translated or not appropriate for context Untranslated text
Language	Grammar – syntax: non-compliance with target language rules Punctuation: non-compliance with target language rules Spelling: errors, accents, capital letters
Terminology	Non-compliance with company terminology Non-compliance with 3rd party or product/application terminology Inconsistent
Style	Non-compliance with company style guides Inconsistent with other reference material Inconsistent within text Literal translation Awkward syntax Unidiomatic use of target language Tone Ambiguous translation
Country Standards	Dates, units of measurement, currency, delimiters, addresses, phone numbers, zip codes, shortcut keys, cultural references, tone…

Εικόνα 8: Πλήρης τυπολογία λαθών DQF

Από τις παραπάνω κατηγορίες ακολουθούν αυτές που θα μπορούσαν να αξιοποιηθούν σε μια αξιολόγηση όπως η δική μας, με παραδείγματα των φαινομένων, τα οποία εντοπίστηκαν σε δοκιμές που πραγματοποιήθηκαν με το σώμα κειμένων μας.

Accuracy	**Πιστότητα**
Incorrect interpretation of source text – mistranslation	Λανθασμένη μετάφραση
Παράδειγμα	“To provide a diagnosis for IBS, the doctor will first order specific **tests** to rule out conditions with similar symptoms to those of the irritable bowel syndrome and to check whether you meet the Rome III criteria.” → “Να παρέχουν μια διάγνωση για το σύνδρομο ευερέθιστου Εντέρου, ο γιατρός θα διατάξει πρώτα συγκεκριμένες **δοκιμές**, να αποκλείσει όρους με παρόμοια συμπτώματα με αυτά του συνδρόμου του ευερέθιστου εντέρου και να ελέγξει κατά πόσον πληροίτε τα κριτήρια της Ρώμης III.”
Σε αυτή την περίπτωση αποδίδεται η λέξη του πρωτότυπου με μία από τις πιθανές λέξεις απόδοσης στην ελληνική, η οποία όμως δεν είναι ορθή βάσει συγκειμένου.	
Incorrect/misunderstanding of	Λανθασμένη απόδοση τεχνικής έννοιας

technical concept	
Παράδειγμα	“Stress causes the production of specific cytokines in the central nervous system acting on the colon **wall**.” → “Στρες προκαλεί την παραγωγή κυτοκινών συγκεκριμένα στο κεντρικό νευρικό σύστημα ενεργεί στον **τοίχο** του παχέος εντέρου.”
Η λέξη “wall” στα αγγλικά χρησιμοποιείται εδώ ως ιατρικός όρος, μεταφράζεται όμως με την κοινή σημασία της ως “τοίχος” αντί για “τοίχωμα”.	
Omission	Παράλειψη
Παράδειγμα	“Day by day we learn more and more about IBS and we come to terms with **its** complexity.” → “Μέρα με τη μέρα όλο και περισσότερο μαθαίνουμε για το σύνδρομο ευερέθιστου Εντέρου και να δεχτούμε με την πολυπλοκότητα.”
Παραλείπεται η κτητική αντωνυμία.	
Addition	Προσθήκη
Παράδειγμα	“10 Frequently Asked Questions” → “10 ****** συχνές ερωτήσεις”
Προστίθενται αστερίσκοι οι οποίοι δεν υπάρχουν στο πρωτότυπο.	
Untranslated text	Αμετάφραστο κείμενο
Παράδειγμα	“1. Do I definitely have **IBS**?” → “1. σίγουρα έχω **IBS**;”
Αρκτικόλεξο που δεν μεταφράζεται, ενώ θα έπρεπε γιατί δεν είναι κατανοητό στον μέσο αναγνώστη στον οποίο απευθύνεται το κείμενο.	
Language	**Γλώσσα**
Grammar-Syntax	Γραμματική-Σύνταξη
Παράδειγμα	“Ειδικές εξετάσεις θα **επιτρέψει** σε μας για τον εντοπισμό μισαλλοδοξίας στα τρόφιμα, καθώς και την παροχή συμβουλών σχετικά με ειδική διατροφή σχέδια που θα ανακουφίσει τους ασθενείς συμπτώματα.”
Το ρήμα δεν συμφωνεί με το υποκείμενο.	
Punctuation	Στίξη
Παράδειγμα	“Τώρα γνωρίζουμε ότι υπάρχουν πάρα πολλά μονοπάτια**,** μεταξύ του εντέρου και του εγκεφάλου (άξονας εγκέφαλος του στομάχου).”
Λανθασμένη χρήση κόμματος.	
Spelling	Ορθογραφία

Παράδειγμα	“Δηλαδή εάν είστε **πιό** μεγαλύτεροι από 50 (ή 40, σε περίπτωση οικογενειακής ιστορίας του καρκίνος του παχέος εντέρου) ή εάν εκτίθεστε στους παράγοντες κινδύνου.”
Λανθασμένος τονισμός λέξης.	
Style	**Ύφος**
Literal translation	Κατά λέξη μετάφραση
Παράδειγμα	“Colonoscopy is performed to rule out other conditions but only **on the ground of** indications.” → “Η κολονοσκόπηση εκτελείται για να αποκλείσει άλλους όρους αλλά μόνο **στο έδαφος των** ενδείξεων.”
Ιδιωτισμός ο οποίος μεταφράζεται κατά λέξη χωρίς να υπάρχει αντιστοιχία στη γλώσσα-στόχος. Θεωρούμε ότι με τον χαρακτηρισμό “κατά λέξη μετάφραση” εδώ εννοείται “καταχρηστική κατά λέξη μετάφραση”.	
Awkward syntax	Περίεργη σύνταξη
Παράδειγμα	“In this context **work regular follow-up clinic steps funders and defibrillators**.”
Λανθασμένη σειρά ρήματος-υποκειμένου και ασύνδετα ουσιαστικά στη σειρά.	

Πίνακας 2: Τυπολογία λαθών DQF

Τα επίπεδα σοβαρότητας λάθους που ορίζει το DQF είναι τα εξής:

Επίπεδο 1

Κρίσιμα (critical): Λάθη που μπορεί να επιφέρουν επιπλοκές που σχετίζονται με την υγεία ή την ασφάλεια ή με νομικά/οικονομικά θέματα, ή που παραβιάζουν οδηγίες γεωπολιτικής χρήσης, βλάπτουν τη φήμη της εταιρείας, προκαλούν βλάβη της εφαρμογής ή αρνητικές μεταβολές/παρερμηνείες σχετικά με τη λειτουργία ενός προϊόντος ή υπηρεσίας, ή μεταβολές/παρερμηνείες που θα μπορούσαν να θεωρηθούν προσβλητικές.

Επίπεδο 2

Σοβαρά (major): Λάθη που ενδέχεται να προκαλέσουν σύγχυση ή παρερμηνεία στον χρήστη, ή να παρεμποδίσουν την ομαλή λειτουργία του προϊόντος ή της υπηρεσίας λόγω σημαντικής αλλαγής στη σημασία ή του ότι τα λάθη εμφανίζονται σε ορατό ή σημαντικό μέρος του περιεχομένου.

Επίπεδο 3

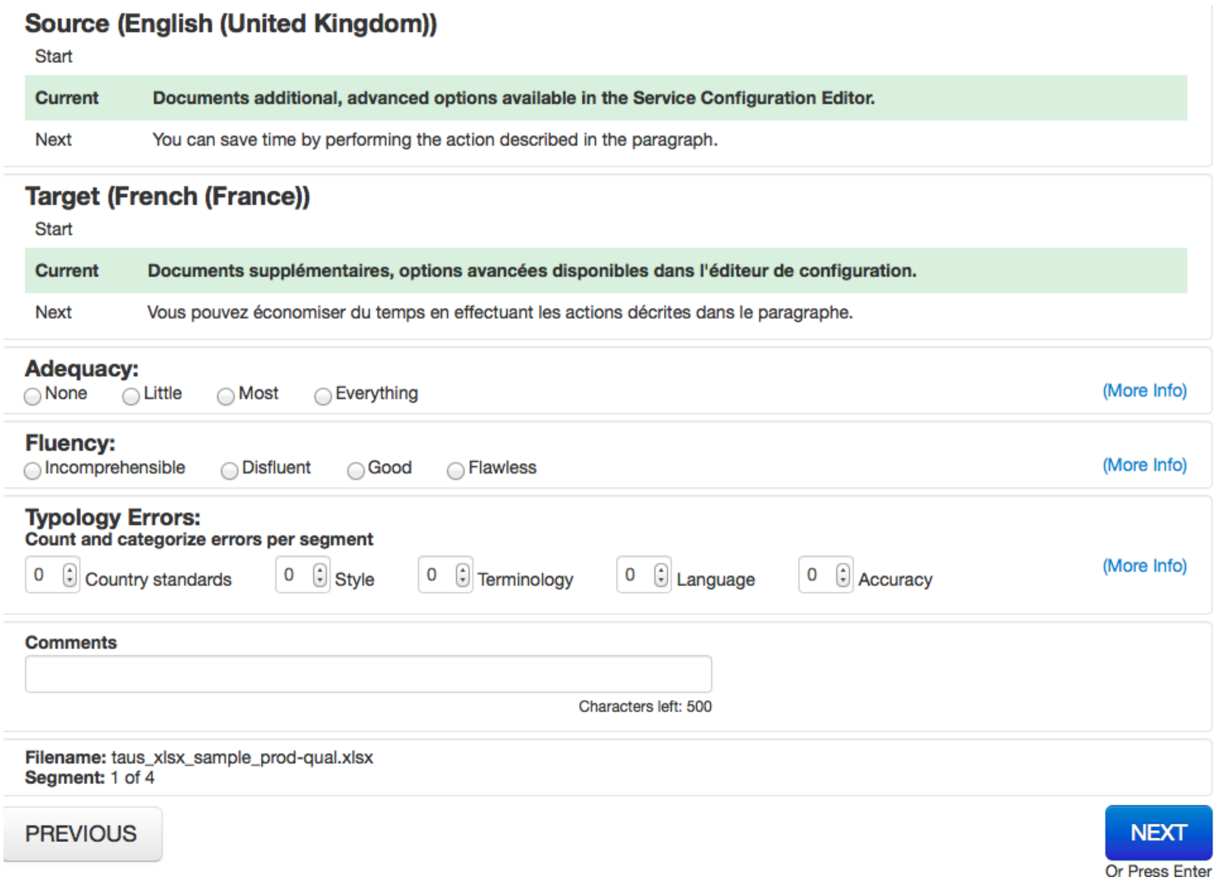
Δευτερεύοντα (minor): Λάθη που δεν έχουν ως αποτέλεσμα απώλεια νοήματος, δεν προκαλούν σύγχυση ή παρερμηνεία στον χρήστη αλλά γίνονται αντιληπτά, επηρεάζουν αρνητικά την ποιότητα του ύφους, τη γραμματικότητα ή την ευκρίνεια, ή λάθη που καθιστούν το περιεχόμενο λιγότερο ελκυστικό.

Επίπεδο 4

Ουδέτερα (neutral): Επιθυμητές αλλαγές σε στοιχεία τα οποία δεν μπορούν να θεωρηθούν λάθη (θέματα υφολογικής επιλογής, λάθη που οφείλονται σε αλλαγές που δεν έχουν ακόμη εφαρμοστεί, κ.λπ.).

Τέλος, προβλέπεται και επίπεδο επαίνου για εξαιρετικές επιδόσεις.

Στην Εικόνα 9 φαίνεται το περιβάλλον επισημείωσης. Εμφανίζεται η πρωτότυπη και η μεταφρασμένη πρόταση σε περιβάλλον, και στο πεδίο Typology Errors επισημειώνονται τα λάθη κάθε κατηγορίας. Αυτό γίνεται πατώντας στο βελάκι που εμφανίζεται δίπλα από τη μηδενική τιμή σε κάθε εμφάνιση λάθους κάθε κατηγορίας. Αυτό σημαίνει ότι ως αποτέλεσμα έχουμε τον αριθμό λαθών κάθε κατηγορίας μέσα σε κάθε πρόταση. Επίσης παρέχεται η δυνατότητα προσθήκης σχολίου έως και 500 χαρακτήρων. Όπως και στις προηγούμενες εργασίες που παρουσιάσαμε στο ίδιο εργαλείο, τα αποτελέσματα εξάγονται στην ίδια μορφή στην οποία εισάγονται τα δεδομένα: υπολογιστικά φύλλα.



Εικόνα 9: Περιβάλλον επισημείωσης ταξινόμησης λαθών DQF

Ως προς την τυπολογία λαθών, παρατηρείται μια συμφόρηση διαφορετικών φαινομένων στις υπάρχουσες κατηγορίες. Αυτό έχει δύο συνέπειες. Από την πλευρά του επισημειωτή, με κατάλληλη πάντα εκπαίδευση, μπορεί να διευκολυνθεί το έργο του, εφόσον δεν θα χρειάζεται να αντιμετωπίζει συνεχώς διλήμματα μεταξύ παρεμφερών κατηγοριών. Από την πλευρά όμως της αξιοποίησης των αποτελεσμάτων, είναι μεν επαρκής αν ο στόχος είναι η βαθμολόγηση συστήματος, περιορίζεται όμως αν στόχος της είναι η μελέτη ή/και αυτόματη εξαγωγή δεδομένων για τα επιμέρους φαινόμενα που περιλαμβάνονται στις κατηγορίες.

3.2.2.3.3 Appraise: Ταξινόμηση λαθών και Μετεπεξεργασία

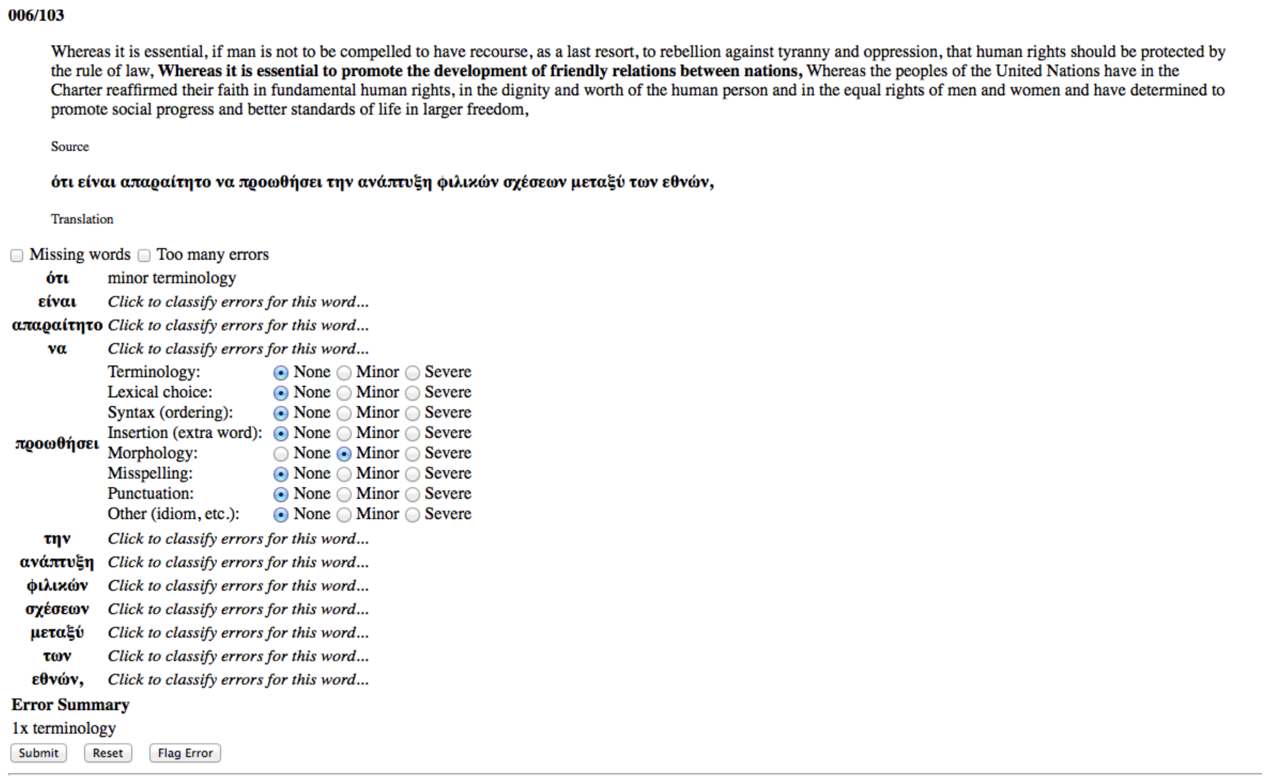
Ταξινόμηση λαθών

Η τρίτη τυπολογία που παρουσιάζουμε είναι η τυπολογία που χρησιμοποιείται στο εργαλείο Appraise για την εργασία Error Classification. Σε αυτή την εργασία επισημείωσης υπάρχουν οι κατηγορίες που εμφανίζονται στην παρακάτω λίστα. Καθώς δεν παρέχονται πληροφορίες σχετικά με τα φαινόμενα ή τις υποκατηγορίες που απαρτίζουν την κάθε κατηγορία, γίνεται εκτενής αναφορά του τρόπου επιλογής των φαινομένων που επιλέξαμε να συμπεριλάβουμε σε κάθε μία από αυτές στο Κεφάλαιο 4.2.2.1.

Κατηγορίες λαθών Appraise:

- Missing Words (Ελλείπουσες Λέξεις)
- Too Many Errors (Υπερβολικός Αριθμός Λαθών)
- Terminology (Ορολογία)
- Lexical Choice (Λεξιλογική Επιλογή)
- Syntax (Ordering) (Σύνταξη [σειρά Λέξεων])
- Insertion (Extra Word) (Προσθήκη Λέξης)
- Morphology (Μορφολογία)
- Misspelling (Λανθασμένη Ορθογραφία)
- Punctuation (Στίξη)
- Other (Idiom, etc.) (Άλλο [Ιδιωματική Έκφραση, κ.λπ.])

Και σε αυτό το εργαλείο δίνεται η επιλογή μεταξύ επιπέδων σοβαρότητας λάθους, σε αυτή την περίπτωση δύο: δευτερεύον (minor) και σοβαρό (severe), για τα οποία επίσης δεν παρέχονται ορισμοί. Στην Εικόνα 9 φαίνεται το περιβάλλον εκτέλεσης της εν λόγω εργασίας. Όπως και στις προηγούμενες εργασίες που παρουσιάσαμε στο Appraise, εμφανίζεται ο αριθμός της τρέχουσας πρότασης, η πρωτότυπη πρόταση σε περιβάλλον και η μεταφρασμένη πρόταση. Ακολουθούν οι επιλογές Missing words και Too many errors, οι οποίες, όπως βλέπουμε, δεν είναι μετρήσιμες. Έπειτα, σε κάθε λέξη ανοίγει μενού με τις κατηγορίες, όπως φαίνεται στην Εικόνα 10, όπου επιλέγεται και το επίπεδο σοβαρότητας του λάθους. Αξίζει να σημειωθεί ότι σε αυτό το περιβάλλον, όπως και στο περιβάλλον επισημείωσης του DQF, η επισημείωση γίνεται σε επίπεδο λέξης. Όταν έχουν ήδη επισημειωθεί κάποια λάθη, εμφανίζεται στο κάτω μέρος περίληψη αυτών. Όπως και στις προηγούμενες εργασίες του Appraise που παρουσιάσαμε, υπάρχει η δυνατότητα submit (υποβολή) και reset (εκ νέου επισημείωση), και τα αποτελέσματα εξάγονται σε αρχείο xml, όπως ακριβώς και τα εισαγόμενα δεδομένα.



Εικόνα 10: Περιβάλλον επισημείωσης Ταξινόμησης Λαθών Appraise

Μετεπεξεργασία (Post-editing)

Ως μετεπεξεργασία ορίζεται η εργασία με την οποία το μηχανικό μετάφρασμα παίρνει τέτοια μορφή στη γλώσσα-στόχος ώστε να μπορεί να χρησιμοποιηθεί [Lacruz, I. κ.ά., 2014]. Στο μεγαλύτερο μέρος τους αυτές οι εργασίες εκτελούνται από επαγγελματίες μεταφραστές [Lacruz, I. κ.ά., 2014], όμως η εκπαίδευση στη μετεπεξεργασία πρόσφατα μόνο άρχισε να εντάσσεται σε μαθήματα μετάφρασης [Lacruz, I. κ.ά., 2014]. Σύμφωνα με τους Lacruz, I. κ.ά. [2014], θεωρείται μια εργασία που διαφέρει κατά πολύ από τη διαδικασία της μετάφρασης και, ως εκ τούτου, οι γνωσιακές (νοητικές) διαδικασίες που απαιτούνται για τις δύο αυτές εργασίες επίσης διαφέρουν. Από εκεί οδηγούνται στο συμπέρασμα ότι η παραδοσιακή εκπαίδευση των μεταφραστών ενδέχεται να μην είναι ιδανική για την εργασία της μετεπεξεργασίας. Θα συμφωνήσουμε ότι πρόκειται για δύο διαφορετικές διαδικασίες, οι οποίες όμως έχουν αρκετά κοινά στάδια. Όταν η μετεπεξεργασία γίνεται με βάση το πρωτότυπο και όχι μια “ιδανική” μετάφραση, εμπεριέχονται σε αυτό τα ακόλουθα στάδια: α) εντοπισμός μεταφραστικών λαθών με αντιπαραβολή πρωτότυπου-μεταφράσματος, β) εντοπισμός γλωσσικών λαθών στη γλώσσα-στόχος, γ) διόρθωση λαθών, δ) έλεγχος αναδιαμορφωμένης πρότασης. Αυτές οι υπο-διαδικασίες ή κάποιες από αυτές δεν είναι απαραίτητα διακριτές, μπορούν δηλαδή να εκτελούνται παράλληλα, και συνήθως επαναλαμβάνονται μέχρι να επιτευχθεί το επιθυμητό αποτέλεσμα. Ο εντοπισμός μεταφραστικών λαθών είναι μια εργασία που μπορεί να πραγματοποιηθεί επιτυχώς από δίγλωσσο άτομο που δεν είναι μεταφραστής, ωστόσο θεωρούμε ότι δεν υπάρχει λόγος αμφισβήτησης του ότι ο μεταφραστής έχει αναπτυγμένη ικανότητα εντοπισμού μεταφραστικών λαθών. Πρόκειται τόσο για μέρος της μεταφραστικής διαδικασίας όσο και για ξεχωριστή εργασία στην οποία επιδίδονται οι μεταφραστές, καθώς συνηθίζεται οι έμπειροι μεταφραστές να επιμελούνται τις μεταφράσεις άλλων μεταφραστών. Αυτό που διαφέρει στη μετεπεξεργασία είναι το είδος των λαθών. Το ίδιο ισχύει και για τον εντοπισμό των γλωσσικών λαθών στη γλώσσα-στόχος, καθώς επίσης πρόκειται για μέρος της εργασίας τους. Όσο για την υπο-διαδικασία διόρθωσης αυτών των λαθών, είναι σχεδόν μια εκ νέου μετάφραση, κατ' εξοχήν έργο του μεταφραστή, γιατί αυτός είναι ο επαγγελματίας που είναι εξοικειωμένος με την εύρεση ορολογίας, της κατάλληλης λέξης ή λεξιλογικής σύναψης, αλλά και με τα εργαλεία που χρησιμεύουν για αυτόν τον σκοπό, και που έχει αναπτύξει τις δεξιότητες οι οποίες απαιτούνται για την απόδοση νοημάτων από μια γλώσσα σε μια άλλη με τη χρήση διαφόρων τεχνικών (παράφραση, χρήση συνωνύμων, κ.λπ.). Ο μεταφραστής επίσης είναι εξοικειωμένος με μεγάλο αριθμό τομέων, όσο κι αν εξειδικεύεται σε κάποιους από αυτούς. Σε αυτό το στάδιο ενδέχεται να απαιτείται διόρθωση με τα λιγότερα δυνατά βήματα και αυτό είναι κάτι στο οποίο δεν έχει απαραίτητα εμπειρία ένας μεταφραστής, αλλά πρόκειται για μια δεξιότητα η οποία βασίζεται στις υπόλοιπες δεξιότητες χειρισμού της γλώσσας που διαθέτει ενας μεταφραστής. Δεν θα έπρεπε να υποτιμάμε τις διαφορές μεταξύ της διόρθωσης μετάφρασης προϊόντος ανθρώπινης εργασίας και μηχανικής επεξεργασίας. Αντιμετωπίζουμε στις δύο περιπτώσεις διαφορετικά φαινόμενα και στην περίπτωση μάλιστα της μηχανικής μετάφρασης τίθεται το θέμα της έκθεσης σε μεγάλο όγκο “τοξικών” κειμένων, για τα οποία εκφράζονται όλο και περισσότερες ενστάσεις στον χώρο των μεταφραστών. Ωστόσο δεν μπορούμε και να αμφισβητήσουμε ότι υπάρχουν επαγγελματίες της γλώσσας που μπορούν να εκπαιδευτούν κατάλληλα για μετεπεξεργασία και αυτοί είναι οι μεταφραστές. Η εκπαίδευσή τους δεν είναι η ιδανική για μετεπεξεργασία αλλά είναι η πλησιέστερη που υπάρχει.

Η μετεπεξεργασία χρησιμοποιείται και ως μέτρο ποιότητας υπολογίζοντας την προσπάθεια και τον χρόνο που καταβάλλεται για αυτήν [Lacruz, I. κ.ά., 2014]. Η έρευνα για τη μετεπεξεργασία είναι ακόμα στα σπάργανα και δεν έχει αποδώσει επαρκή αποτελέσματα [Lacruz, I. κ.ά., 2014]. Η ολοένα αυξανόμενη ανάγκη για αυτήν την τοποθετεί στο κέντρο του ερευνητικού και όχι μόνο ενδιαφέροντος [Lacruz, I. κ.ά., 2014].

Αξίζει να σημειωθεί ότι από το 2010 έχει αρχίσει να εκπονείται έρευνα για αυτόματη μετεπεξεργασία. Πρόκειται για μια πολλά υποσχόμενη νέα τεχνική σε παγκόσμιο επίπεδο, η οποία έχει προς το παρόν αποδώσει συγκεκριμένους καρπούς. Ενδεικτικά αναφέρεται το σύστημα αυτόματης μετεπεξεργασίας Depfix [Rudolf, R., 2014], το οποίο διορθώνει λάθη σε μεταφράσεις από τα αγγλικά προς τα τσέχικα οι οποίες έχουν παραχθεί από στατιστικά συστήματα. Αυτό επιτυγχάνεται μέσω της χρήσης διαφόρων εργαλείων επεξεργασίας φυσικής γλώσσας για την ανάλυση του εισαγόμενου κειμένου και την παραγωγή του εξαγόμενου κειμένου. Το βασικό του χαρακτηριστικό του είναι ότι, εκτός από το στατιστικό μέρος του, περιλαμβάνει κανόνες εντοπισμού και διόρθωσης λαθών (28 κανόνες στη μορφή που είχε το 2014). Τα λάθη εντοπίζονται βάσει επισημειωμένης πληροφορίας που φέρει η κάθε λέξη σε σύγκριση με την αντίστοιχη λέξη στο πρωτότυπο ή/και τις γειτονικές τις λέξεις και οι κανόνες έχουν εξαχθεί από ταξινόμηση λαθών και μετεπεξεργασία που πραγματοποιήθηκε από ανθρώπους [Bojar, O., 2011]. Συνεπώς, τόσο η ταξινόμηση λαθών όσο και η μετεπεξεργασία μπορούν να αξιοποιηθούν σε μια διαδικασία αυτοματοποίησης της μετεπεξεργασίας.

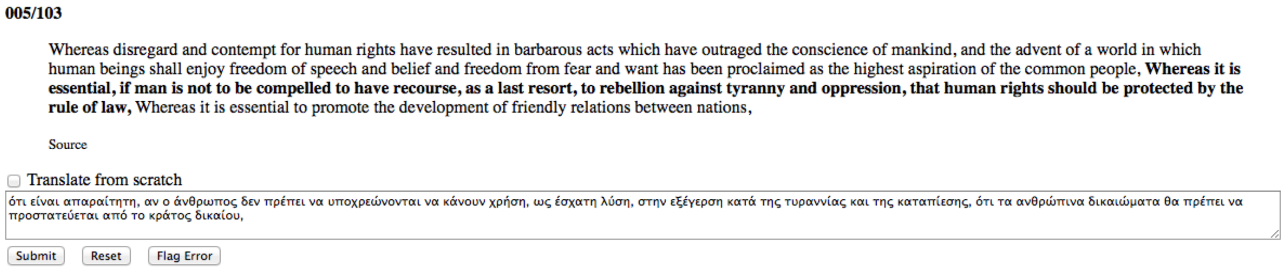
Ως προς τη διαδικασία της μετεπεξεργασίας, παραθέτουμε τις οδηγίες που χρησιμοποιήθηκαν στην αξιολόγηση μηχανικής μετάφρασης GALE[38] το 2007 [Przybocki, M. κ.ά., 2011].

1) Το μηχανικό μετάφρασμα πρέπει να έχει το ίδιο νόημα με τη μετάφραση-αναφορά, χωρίς να προστίθεται περιεχόμενο σε αυτό ή να αφαιρείται περιεχόμενο από αυτό.
2) Το μηχανικό μετάφρασμα πρέπει να είναι τόσο κατανοητό όσο η μετάφραση-αναφορά. Επίσης πρέπει να έχει τον ίδιο βαθμό αμφισημίας με αυτό (ούτε μεγαλύτερο ούτε μικρότερο).
3) Η στίξη πρέπει να είναι κατανοητή και οι μονάδες που λειτουργούν ως προτάσεις πρέπει να έχουν στίξη που να δηλώνει τέλος πρότασης και κατάλληλη χρήση κεφαλαίων. Αλλαγές σε σημεία στίξης δεν πρέπει να γίνονται με βάση τις οδηγίες της παραδοσιακής γραμματικής.
4) Το νόημα πρέπει να αποδοθεί με όσο το δυνατόν λιγότερες διορθώσεις χρησιμοποιώντας κατανοητή γλώσσα. Αν το μηχανικό μετάφρασμα είναι απόλυτα αποδεκτό ως προς τις λέξεις, τις φράσεις και τη στίξη, δεν πρέπει να γίνει καμία αλλαγή.

Σε περίπτωση αμφιβολίας μεταξύ των τεσσάρων ανωτέρω κανόνων, πρέπει να θεωρείται ότι εφαρμόζονται με σειρά προτεραιότητας.

Σε επίπεδο επισημείωσης, στο Appraise η μετεπεξεργασία εκτελείται σε παρόμοιο περιβάλλον με τα προηγούμενα. Το μετάφρασμα εμφανίζεται σε ένα πλαίσιο (Εικόνα 11), όπου ο χρήστης μπορεί να επέμβει, για να προβεί στις επιθυμητές διορθώσεις. Η επιλογή Translate from scratch απλώς απαλείφει το περιεχόμενο του κουτιού, δεν αποτελεί πληροφορία η οποία εμφανίζεται μετά στα αποτελέσματα. Και πάλι, στο τέλος της διαδικασίας, εξάγεται ένα αρχείο xml με το διορθωμένο κείμενο.

[38] Global Autonomous Language Exploitation (http://www.speech.sri.com/projects/GALE/).



Εικόνα 11: Περιβάλλον επισημείωσης Μετεπεξεργασίας στο Appraise

3.2.3 Πλεονεκτήματα μη αυτοματοποιημένης αξιολόγησης

Στα αναφερόμενα πλεονεκτήματα της μη αυτοματοποιημένης σε σύγκριση με την αυτόματη αξιολόγηση, συγκαταλέγονται τα εξής: κατ' αρχάς, με το σκεπτικό ότι οι μεταφράσεις παράγονται για χρήστες που είναι άνθρωποι και όχι μηχανές, η ανθρώπινη κρίση θεωρείται το σωστό μέτρο αξιολόγησης των ιδιοτήτων μιας μετάφρασης [Sanders, G. κ.ά., 2011]· έπειτα, η ανθρώπινη κατανόηση του πραγματικού κόσμου επιτρέπει στον άνθρωπο-κριτή να αξιολογήσει την πρακτική σημασία των λαθών μιας μετάφρασης [Sanders, G. κ.ά., 2011]· τέλος, θεωρείται ότι δεν υπάρχει υποκατάστατο της ανθρώπινης κρίσης για την περίπτωση της μετάφρασης κι έτσι οι ανθρώπινες κρίσεις αποτελούν το σημείο αναφοράς για την ποιότητα της μετάφρασης [Sanders, G. κ.ά., 2011].

3.2.4 Μειονεκτήματα μη αυτοματοποιημένης αξιολόγησης

Μεταξύ των μειονεκτημάτων της μη αυτοματοποιημένης σε σύγκριση με την αυτόματη αξιολόγηση επαναλαμβάνεται κατ' αρχάς η υποκειμενικότητά της [Fordyce, C. S., 2007, Bonnie, D. κ.ά., 2011]. Παραμένει βέβαια το παράδοξο του αρνητικού χαρακτήρα της υποκειμενικότητας ενώ παράλληλα αναγνωρίζεται ότι αυτή η υποκειμενικότητα είναι το σημείο αναφοράς και ό,τι καλύτερο έχουμε μέχρι στιγμής. Επαναλαμβάνονται επίσης τα αρνητικά του υψηλού κόστους, της μη επαναληψιμότητας και του χρονοβόρου χαρακτήρα της [Bonnie, D. κ.ά., 2011]. Άλλο συχνά επαναλαμβανόμενο μειονέκτημα είναι η ασυμφωνία μεταξύ των κριτών [Bonnie, D. κ.ά., 2011], αλλά αυτό είναι ένα αντιμετωπίσιμο φαινόμενο: γίνονται έλεγχοι στατιστικής σημαντικότητας και χρησιμοποιούνται ομάδες ανεξάρτητων κριτών [Bonnie, D. κ.ά., 2011]. Επιπλέον όμως η διαδικασία της αξιολόγησης δεν είναι ίδιον μόνο του χώρου της μηχανικής μετάφρασης και σε οποιαδήποτε αξιολόγηση χρησιμοποιούνται άνθρωποι-κριτές αναπτύσσονται και οι κατάλληλες τεχνικές για την ελαχιστοποίηση των δυσμενών συνεπειών της ασυμφωνίας μεταξύ κριτών. Συγκεκριμένα από τον χώρο της γλώσσας, αξίζει να αναρωτηθεί κανείς πώς αντιμετωπίζεται αυτό το πρόβλημα στον τομέα της γλωσσικής αξιολόγησης, είτε πρόκειται για αξιολόγηση μιας έκθεσης ιδεών στη μητρική γλώσσα του μαθητή είτε πρόκειται για την αξιολόγηση οποιουδήποτε είδους κειμένου, γραπτού ή προφορικού, σπουδαστών ξένης γλώσσας. Η απάντηση είναι: με αυστηρά κριτήρια και εκπαίδευση των αξιολογητών.

4 Διαδικασία και πειράματα αξιολόγησης

Σκοπός των πειραμάτων που παρουσιάζονται σε αυτό το κεφάλαιο είναι η εξαγωγή συμπερασμάτων σχετικά με την κατάταξη των τεσσάρων συστημάτων υπό αξιολόγηση και την απόδοσή τους στα επιμέρους πεδία και κατευθύνσεις, καθώς και σχετικά με τα επιμέρους γλωσσικά φαινόμενα στα οποία παρουσιάζει αδυναμίες το σύστημα Moses, το οποίο είναι υπό ανάπτυξη.

4.1 Σώματα κειμένων

Για την εκπόνηση της παρούσας εργασίας απαιτήθηκε η συλλογή δύο παράλληλων σωμάτων κειμένων. Καθώς επικρατεί μια σύγχυση ως προς τον χαρακτηρισμό “παράλληλο”, ο οποίος μάλιστα συνυπάρχει με άλλους, όπως “μεταφραστικά” και “συγκρίσιμα” σώματα κειμένων, κρίνεται απαραίτητη η αποσαφήνισή του.

Σύμφωνα με την αρχική κατηγοριοποίηση της Mona Baker, υπάρχουν τρία είδη σωμάτων κειμένων για μεταφραστική έρευνα και διδακτική: α) συγκρίσιμα, β) πολύγλωσσα και γ) παράλληλα. Το συγκρίσιμο σώμα κειμένων ορίζεται ως ένα σώμα αποτελούμενο από δύο σώματα στην ίδια γλώσσα, το ένα με πρωτότυπα κείμενα και το άλλο με μεταφρασμένα στην ίδια γλώσσα. Τα πολύγλωσσα σώματα αποτελούνται από δύο σώματα σε δύο γλώσσες με αντίστοιχα πρωτότυπα κείμενα. Τέλος, τα παράλληλα σώματα αποτελούνται από μια συλλογή κειμένων σε μια γλώσσα και μια συλλογή κειμένων σε μετάφρασή τους σε άλλη γλώσσα [Fernandes, L., 2006].

Οι αρχικοί ορισμοί αυτοί, ύστερα από διαφοροποιήσεις και κριτικές, χρησιμοποιούνται πλέον ευρέως με διαφορετικό τρόπο, παρόλο που, εξακολουθεί να υφίσταται μια κάποια σύγχυση σχετικά με τη χρήση τους, όπως προαναφέρθηκε. Η σύγχυση αυτή ενισχύεται αφενός μεν από την ύπαρξη του όρου “μεταφραστικά σώματα κειμένων”, που περιλαμβάνει τα είδη σωμάτων που χρησιμοποιούνται στη μετάφραση και στη μεταφρασεολογία, αφετέρου δε από το γεγονός ότι στον χώρο της μετάφρασης χρησιμοποιείται πολλές φορές ο όρος “παράλληλα σώματα” για να γίνει αναφορά τόσο σε ζεύγη σωμάτων με πρωτότυπα κείμενα και τις μεταφράσεις τους όσο και σε ζεύγη σωμάτων με αντίστοιχα κείμενα σε δύο διαφορετικές γλώσσες. Στην παρούσα εργασία γίνεται η εξής διάκριση: ένα παράλληλο σώμα κειμένων αποτελείται από κείμενα σε μία γλώσσα και μία ή περισσότερες μεταφράσεις των ίδιων κειμένων σε μια άλλη, ενώ ένα συγκρίσιμο σώμα αποτελείται από δύο σώματα σε δύο γλώσσες, τα οποία περιλαμβάνουν κείμενα πρωτότυπα (όχι μεταφρασμένα) με όσο το δυνατόν μεγαλύτερη αντιστοιχία ως προς την έκταση, το θέμα και άλλα χαρακτηριστικά [Lawson, A., 2001].

Σε επόμενο επίπεδο κατηγοριοποίησης, αναφέρονται τρία είδη παράλληλων κειμένων: α) σώμα με κείμενα πρωτότυπα στη γλώσσα Α και με τη μετάφρασή τους στη γλώσσα Β (και Γ…), β) σώμα με ίδιο αριθμό κειμένων πρωτότυπων στη γλώσσα Α και Β και οι αντίστοιχες μεταφράσεις τους (αμφίδρομο [reciprocal] σώμα) και γ) σώμα με μεταφράσεις κειμένων στη γλώσσα Α, Β και Γ, ενώ τα πρωτότυπα κείμενα ήταν γραμμένα στη γλώσσα Ω [Teubert, W., 1996]. Τα σώματα της παρούσας εργασίας αποτελούν ένα αμφίδρομο σώμα, όπως θα φανεί στο Κεφάλαιο 4.1.1. Επίσης γίνονται και άλλων ειδών διακρίσεις ανάλογα με την κατεύθυνση της μετάφρασης [Fernandes, L., 2006], αλλά και μεταξύ γενικών και ειδικών κειμένων. Στην παρούσα εργασία χρησιμοποιούμε μόνο ειδικά κείμενα, για τα οποία γίνεται εκτενέστερος λόγος στο Κεφάλαιο 4.1.2.

Τέλος, παρόλο που συνήθως αναφερόμαστε στα παράλληλα σώματα ως ένα ζεύγος κειμένων σε δύο γλώσσες, καθώς με αυτό τον τρόπο είθισται να τα χρησιμοποιούμε, υπάρχουν σήμερα μεγάλα σώματα κειμένων που περιλαμβάνουν πρωτότυπα και μεταφράσεις σε μεγάλο αριθμό γλωσσών, στα οποία θα αναφερθούμε στο Κεφάλαιο 4.1.2.

4.1.1 Κατάρτιση σωμάτων

Η αξιολόγηση πραγματοποιήθηκε για συστήματα μετάφρασης στο ζεύγος γλωσσών αγγλικά-ελληνικά, σε ευθεία και αντίστροφη[39], εξ ου και η ύπαρξη δύο σωμάτων, ένα για την κάθε κατεύθυνση.

Το κάθε σώμα απαρτίζεται από 15 πρωτότυπα κείμενα και 3 μεταφράσεις για το καθένα από αυτά[40]. Τις μεταφράσεις αυτές τις έχουν επεξεργαστεί 3 διαφορετικοί μεταφραστές. Με τον όρο «επεξεργασία» εδώ νοείται είτε η μετάφραση χωρίς καμία μηχανική υποστήριξη, είτε η μετάφραση με χρήση μηχανικού συστήματος υποβοήθησης, είτε η επιμέλεια μετάφρασης

[39] Ευθεία είναι η μετάφραση που γίνεται από μια ξένη γλώσσα προς τη μητρική γλώσσα του μεταφραστή και αντίστροφη αυτή που γίνεται από τη μητρική προς μια ξένη. Στην περίπτωσή μας θεωρείται ευθεία η μετάφραση από αγγλικά προς ελληνικά και αντίστροφη η μετάφραση από ελληνικά προς αγγλικά.

[40] Πρόκειται για το αρχικό σώμα, το οποίο στη συνέχεια εμπλουτίζεται με μηχανικές μεταφράσεις και επεξεργασμένες μηχανικές μεταφράσεις, όπως θα δούμε στα επόμενα κεφάλαια.

μηχανικού συστήματος. Ο αριθμός των 3 μεταφράσεων-αναφορά ανά πρωτότυπο κρίνεται ικανοποιητικός ώστε να χρησιμοποιηθούν σε ένα σύστημα αυτόματης αξιολόγησης με αρκετά αξιόπιστα αποτελέσματα (βλ. Κεφάλαιο 3.1.1.4).

Ως προς την κατάρτιση των σωμάτων, ήταν σκόπιμο να εντοπίσουμε πρωτότυπα κείμενα με μία υπάρχουσα μετάφραση, προκειμένου να καταστεί εφικτή η συλλογή του αναγκαίου αριθμού παράλληλων κειμένων σε λογικά χρονικά πλαίσια και με τη χρήση περιορισμένων ανθρώπινων πόρων. Υπάρχουσες πηγές τέτοιων κειμένων είναι τα σώματα που διατίθενται από την Ευρωπαϊκή Ένωση και διάφορους οργανισμούς[41], καθώς και ιστότοποι όπου αναρτώνται κείμενα σε αυτές τις δύο γλώσσες (π.χ. ιστοσελίδες θεσμών και οργανισμών με κείμενα πρωτότυπα και τις μεταφράσεις τους σε τουλάχιστον μία γλώσσα, όπως είναι η ιστοσελίδα της Βουλής των Ελλήνων).

4.1.2 Κριτήρια επιλογής πρωτότυπων κειμένων

Προκειμένου να καταστεί εφικτή η αξιοποίηση του περιορισμένου αυτού αριθμού κειμένων και να εξαχθούν όσο το δυνατόν πιο γενικεύσιμα συμπεράσματα, κρίθηκε απαραίτητη η ύπαρξη αντιπροσωπευτικότητας σε διάφορα επίπεδα.

Κατ' αρχάς, θεωρήθηκε απαραίτητη η ύπαρξη κειμένων από διαφορετικούς τομείς του επιστητού, καθώς αυτό συνεπάγεται διαφορετική γλώσσα και διαφορετική ορολογία. Για αυτόν τον σκοπό επιλέγονται τρεις από τους πιο αντιπροσωπευτικούς στην πρακτική της μετάφρασης τομείς, ήτοι της τεχνολογίας, της ιατρικής και του δικαίου και διοίκησης. Ως προς τη γλώσσα, η διαφοροποίησή της αναλύεται στις επόμενες παραγράφους. Ως προς την ορολογία, η ταξινόμηση ενός κειμένου σε έναν τομέα δεν συνεπάγεται απαραίτητα την απουσία ορολογίας από άλλους τομείς. Κρίθηκε επιθυμητό τα κείμενα εργασίας μας να περιέχουν όρους στην πλειονότητά τους, αν όχι αποκλειστικά, από τον τομέα στον οποίο έχει ταξινομηθεί το εν λόγω κείμενο.

Κρίθηκε επίσης απαραίτητη η διαφοροποίηση και στο κειμενικό είδος, διότι και αυτό συνεπάγεται διαφορές στη χρησιμοποιούμενη γλώσσα. Το κειμενικό είδος είναι άρρηκτα συνδεδεμένο με τον

[41] Υπάρχει πληθώρα εμπλουτιζόμενων πολύγλωσσων σωμάτων στο διαδίκτυο προς επεξεργασία. Για πληροφορίες σχετικά με τα υπάρχοντα σώματα, τον όγκο τους, τα δικαιώματα χρήσης, τις γλώσσες που συμπεριλαμβάνονται, τις πηγές, τους τρόπους αξιοποίησής τους και τους λόγους ελεύθερης διάθεσής τους βλ. Steinberger, R. κ.ά., 2014. Βασικό χαρακτηριστικό τους είναι ότι καλύπτουν αρκετούς τομείς του επιστητού και πολλές γλώσσες, μεταξύ των οποίων και ευρωπαϊκές γλώσσες σε κίνδυνο ψηφιακής εξαφάνισης, όπως η ελληνική [Steinberger, R. κ.ά., 2014, Γαβριηλίδου, Μ. κ.ά., 2012]. Ενδεικτικά αναφέρεται ο ιστότοπος http://opus.lingfil.uu.se [Tiedemann, J., 2009], όπου τα σώματα προέρχονται από διάφορες πηγές, μεταξύ των οποίων και η Ευρωπαϊκή Ένωση. Η εν λόγω συλλογή περιέχει μεταξύ άλλων σώματα κειμένων μεταφρασμένης λογοτεχνίας, της Ευρωπαϊκής Κεντρικής Τράπεζας, των πρακτικών του Ευρωκοινοβουλίου και το Ευρωπαϊκό Σύνταγμα.

τομέα στον οποίο ανήκει ένα κείμενο. Π.χ. στον τομέα της ιατρικής είθισται να συναντάει κανείς περιγραφές ιατρικών μηχανημάτων ή φαρμάκων, στον τομέα του δικαίου συμβάσεις ή αποφάσεις και στον τομέα της τεχνολογίας οδηγίες ή περιγραφές συστημάτων. Καθένα από τα είδη αυτά επιτελεί διαφορετικούς σκοπούς, για την επίτευξη των οποίων γίνεται χρήση διαφορετικών γλωσσικών μέσων (π.χ. η μορφή των ρημάτων διαφέρει μεταξύ οδηγιών και περιγραφών, η έκφραση της υποχρέωσης πραγματώνεται με διαφορετικό τρόπο σε μια νομική απόφαση και σε μια περιγραφή φαρμάκου). Ωστόσο, παρόλο που το κειμενικό είδος είναι συνυφασμένο με τον τομέα, στην πράξη συναντώνται κείμενα που ανήκουν σε ένα κειμενικό είδος αλλά αφορούν ένα θέμα από τομέα στον οποίο δεν συνηθίζεται το συγκεκριμένο είδος (π.χ. νομική απόφαση που αφορά ιατρικό θέμα) και αυτό βεβαίως έχει άμεση σχέση με την προαναφερθείσα μεικτή ορολογία.

Η διαφοροποίηση της γλώσσας γίνεται αισθητή μεταξύ διαφορετικών τομέων και ειδών, αλλά και μεταξύ κειμένων του ίδιου τομέα και είδους, λόγω ιδιοσυγκρασιακών διαφορών μεταξύ των συντακτών καθώς και άλλων παραγόντων. Αυτή πραγματώνεται σε επίπεδο κειμένου και σε επίπεδο πρότασης με συγκεκριμένους μηχανισμούς.

Σε επίπεδο κειμένου ένα βασικό χαρακτηριστικό διαφοροποίησης είναι η δομή, η οποία μπορεί να διαφέρει από τον τρόπο παρουσίασης του κειμένου μέχρι το αν το κείμενο αποτελείται από πολλά διακριτά μέρη ή αν σε αυτό για την κατανόηση ενός μέρους απαιτείται η κατανόηση του προηγούμενου. Παράδειγμα κειμένου με αρκετά διακριτά μέρη αποτελεί το τμήμα ενός κειμένου όπου δίνονται οι ορισμοί των εννοιών που θεωρούνται απαραίτητες για την κατανόηση του υπόλοιπου κειμένου, έννοιες οι οποίες απαρτίζονται σε μεγάλο βαθμό από παγιωμένη ορολογία και δεν συνδέονται μεταξύ τους, ή το φύλλο οδηγιών χρήσης ενός φαρμάκου, όπου η δοσολογία γίνεται κατανοητή χωρίς την κατανόηση των ανεπιθύμητων ενεργειών ή των δραστικών συστατικών του φαρμάκου. Αντίθετα, για να γίνει κατανοητή μια παράγραφος από το αιτιολογικό μιας δικαστικής απόφασης, είναι σε μεγάλο βαθμό απαραίτητη η κατανόηση της προηγούμενης ή/και των προηγούμενων παραγράφων. Το ίδιο ισχύει και για την κατανόηση των γεγονότων σε μια έγκληση. Αυτό συμβαίνει γιατί σε αυτές τις περιπτώσεις λόγω οικονομίας παραλείπονται ήδη αναφερθέντα στοιχεία και υπάρχει γενικότερα μεγαλύτερος αριθμός συναναφορών. Από την άλλη πλευρά, σε ένα κείμενο με διακριτά μέρη δεν είναι τόσο εύκολο να πραγματοποιηθεί άρση λεξικής αμφισημίας όσο σε αυτά με λόγο στον οποίο «χτίζονται» τα επιμέρους τμήματα σε στενή αλληλεξάρτηση. Αυτό έχει να κάνει με τους μηχανισμούς συνεκτικότητας και συνοχής ενός κειμένου [De Beaugrande, R.-A., Dressler, W., 1981], των οποίων γίνεται λιγότερο εκτεταμένη χρήση στα κείμενα με διακριτά μέρη, με αποτέλεσμα να δυσχεραίνεται η κατανόηση εξ αιτίας φαινομένων όπως η μη πραγμάτωση της νοηματικής συνάφειας μεταξύ περιόδων και παραγράφων ή η μη επίλυση συναναφοράς (η αδυναμία εντοπισμού δηλαδή της σχέσης μεταξύ δύο τουλάχιστον εκφράσεων που αναφέρονται στην ίδια οντότητα [Crystal, D., 2008]). Τέλος, η δομή μπορεί να παρουσιάζει κι άλλη μία διαφορά που έχει βαρύτητα στη μεταφραστική διαδικασία: ανάλογα με τη μορφή που έχει, συναντώνται όροι εντός συγκειμένου ή εκτός (π.χ. σε μορφή λίστας, περιεχομένων, κ.λπ.). Από την οπτική της μετάφρασης, αυτό είναι κάτι το οποίο αντιλαμβάνεται εμπειρικά και διαισθητικά οποιοσδήποτε μεταφραστής και φυσικός χρήστης της γλώσσας. Σύμφωνα με τον Lyons [Λάιονς, Τ., 2006], η ερμηνεία των περισσότερων γλωσσικών εκφωνημάτων εξαρτάται από τα περικείμενά τους. Άλλωστε η πολυσημία είναι ένα φαινόμενο που συναντάται και στην ορολογία [Newmark, P., 1988, Κελάνδριας, Π., 2013], πράγμα το οποίο καθιστά απαραίτητο το συγκείμενο για την αποσαφήνιση των όρων και την απόδοσή τους. Χωρίς αυτό αυξάνεται ο βαθμός μεταφραστικής δυσκολίας. Στο πλαίσιο του μεταφραστικού μοντέλου των Μπατσαλιά και Σελλά-Μάζη [2004], ο Βαθμός Μεταφραστικής Δυσκολίας μιας μεταφραστικής ενότητας ανεβαίνει όσο αυξάνονται τα επίπεδα στα οποία ανήκει η μεταφραστική μονάδα (σημασιολογικό, λεξικό, μορφολογικό, συντακτικό, πραγματολογικό και υφολογικό) και το άθροισμα των βαθμών όλων των μεταφραστικών ενοτήτων αποτελούν τον βαθμό μεταφραστικής δυσκολίας όλου του κειμένου. Ωστόσο “προκειμένου να μεταφραστούν, άλλες [μεταφραστικές ενότητες] απαιτούν ερμηνεία μηδενικού βαθμού, άλλες πρώτου βαθμού και άλλες, τέλος, δευτέρου βαθμού. Κατ' ακολουθία, θεωρούμε ότι ο Β.Μ.Δ. μιας μεταφραστικής ενότητας εξαρτάται και από αυτήν την ανάλυση και ερμηνεία” [Μπατσαλιά, Φρ., Σελλά-Μάζη, Ε., 2004]. Η έλλειψη συγκειμένου μπορεί να αυξήσει σημαντικά τον απαιτούμενο βαθμό ερμηνείας. Από την οπτική της στατιστικής μηχανικής μετάφρασης, καθώς αυτή γίνεται σε επίπεδο πρότασης, εάν μια πρόταση αποτελείται μόνο από έναν μονολεκτικό όρο, εκτελείται αναζήτηση σε επίπεδο 1-γράμματος, κάτι το οποίο είναι λιγότερο αποδοτικό από την αναζήτηση σε επίπεδο μεγαλύτερων ν-γραμμάτων.

Ωστόσο σε επίπεδο κειμένου γίνονται αντιληπτά και τα χαρακτηριστικά του ύφους και του γλωσσικού μητρώου με μεγαλύτερη ακρίβεια από ό,τι σε επίπεδο πρότασης. Τα ιδιοσυγκρασιακά χαρακτηριστικά του συγγραφέα, ενδεχόμενες λεξιλογικές ή υφολογικές επιλογές του, αλλά και λεξικογραμματικά φαινόμενα που σχετίζονται με το μητρώο ενδέχεται να μην είναι εμφανή σε όλες τις προτάσεις του κειμένου. Για τους σκοπούς της εργασίας μας θεωρήθηκε απαραίτητο να υπάρχει έστω κάποια μικρή διαφοροποίηση στο ύφος των κειμένων.

Σε επίπεδο πρότασης η διαφορά της γλώσσας πραγματώνεται στο μήκος της πρότασης, στην πολυπλοκότητά της (εύρος χρήσης αναφορικών, παρενθετικών και επεξηγηματικών προτάσεων), στον βαθμό συμπύκνωσης του λόγου και ακρίβειας στην έκφραση, στη χρήση διαφορετικών λεξικογραμματικών φαινομένων, όπως η ύπαρξη πολυλεκτικών συνόλων[42] είτε πρόκειται για

[42] “Ιδιοσυγκρασιακές ερμηνείες που ξεπερνούν τα όρια της λέξης” [Sag, I. A. κ.ά., 2002].

όρους είτε όχι, η χρήση άρθρων, η ουσιαστικοποίηση, η ύπαρξη ή μη περιγραφικών επιθέτων και η ύπαρξη όρων που δεν ανήκουν στη γλώσσα του κειμένου, είτε πρόκειται για όρους από τα αγγλικά, είτε για όρους από τα λατινικά, είτε για όρους από παλαιότερες μορφές της ελληνικής.

Η ύπαρξη ποικιλίας σε αυτά τα φαινόμενα στα σώματά μας εξασφαλίζεται σε μεγάλο βαθμό από την επιλογή κειμένων διαφορετικών τομέων και ειδών.

4.1.3 Επιλογή πρωτότυπων κειμένων

Συνεπώς, λαμβάνοντας υπόψη όλα τα παραπάνω επίπεδα διαφοροποίησης των κειμένων μας, αλλά και τους περιορισμούς των υπαρχουσών πηγών κειμένων, αντλήσαμε ένα μεγάλο μέρος αυτών από τα σώματα της Ευρωπαϊκής Ένωσης[43], αξιοποιώντας τα επιμέρους τμήματα που χρειάζονται για να επιτευχθεί η επιθυμητή ποικιλία (τμήματα με καθιερωμένη νομική δομή αποφάσεων, τμήματα με περιγραφές προϊόντων τεχνολογίας, κ.λπ.). Στα πλαίσια αυτού, επιλέχθηκαν στον βαθμό του δυνατού τόσο κείμενα με μακροπερίοδο λόγο όσο και κείμενα με μικρότερες προτάσεις, κείμενα με συνεχόμενη ροή λόγου και με διακριτά μέρη, καθώς και κείμενα με ορολογία εντός και εκτός συγκειμένου. Επίσης για να αποφευχθεί η υπερβολική ομοιομορφία στο ύφος της σύνταξης και για να επιτευχθεί και ο στόχος της ποικιλίας κειμενικών ειδών, αντλήθηκαν κείμενα και από άλλες πηγές, διότι, αν και σε αυτά τα σώματα καλύπτονται διάφοροι τομείς σε λεξιλογικό επίπεδο, υπερισχύουν συγκεκριμένα είδη κειμένων [Steinberger, R. κ.ά., 2014]. Τέλος, φροντίσαμε να μη συμπεριλάβουμε κείμενα από τα σώματα που χρησιμοποιήθηκαν στην εκπαίδευση του υπό ανάπτυξη και υπό αξιολόγηση συστήματος Moses Baseline, για το οποίο γίνεται εκτενής ανάλυση στη συνέχεια.

Ακολουθεί πίνακας με τα κείμενα που αποτέλεσαν τα σώματά μας. Στην ονομασία του κειμένου αναφέρεται ο τομέας (T: Τεχνολογία, M: Ιατρική, L: Δίκαιο και Διοίκηση), η κατεύθυνση (en_gr: μετάφραση από αγγλικά προς ελληνικά, gr_en: μετάφραση από ελληνικά προς αγγλικά) και ο αύξων αριθμός του κειμένου (από 1 έως 5 σε κάθε τομέα). Η ένδειξη en ή gr στο τέλος της ονομασίας δηλώνει ότι πρόκειται αντίστοιχα για το αγγλικό ή το ελληνικό κείμενο, το πρωτότυπο δηλαδή, σε αντίθεση με τις μεταφράσεις, οι ονομασίες των οποίων έχουν κατάληξη HT ή MT.

[43] Αξιοποιήθηκε η βάση EUR-Lex (http://eur-lex.europa.eu), η οποία παρέχει ελεύθερη πρόσβαση και στις 24 επίσημες γλώσσες της Ε.Ε. σε διαφόρων ειδών νομικά έγγραφα, όπως η Επίσημη Εφημερίδα της Ευρωπαϊκής Ένωσης, συνθήκες, οδηγίες, κανονισμοί, αποφάσεις, νομοθετικές προτάσεις, εκθέσεις, διεθνείς συμφωνίες, κ.λπ.

Αγγλικά-Ελληνικά		
Κείμενο	**Λέξεις**	**Χαρακτηριστικά**
Τεχνολογία		
T_en_gr_1_en	1.312	**Είδος**
		Απόσπασμα απόφασης Επιτροπής σχετικά με προδιαγραφές για εξυπηρετητές υπολογιστών και τροφοδοτικά αδιάλειπτης παροχής
		Γλωσσικά και κειμενικά χαρακτηριστικά
		Περιγραφές προϊόντων, χαρακτηριστικών και λειτουργιών τους Αρκετά τεχνικό λεξιλόγιο, απαριθμήσεις, πλειονότητα προτάσεων μικρές
T_en_gr_2_en	1.024	**Είδος**
		Απόσπασμα Κανονισμού σχετικά με τις κατευθυντήριες γραμμές για τα διευρωπαϊκά δίκτυα υποδομών των τηλεπικοινωνιών
		Γλωσσικά και κειμενικά χαρακτηριστικά
		Απόσπασμα όπου αναφέρονται οι υποδομές Μακροπερίοδος λόγος Πολλοί νεολογισμοί Όροι όχι ιδιαίτερα εξειδικευμένοι
T_en_gr_3_en	1.828	**Είδος**
		Ανακοίνωση της Επιτροπής “προς μια ακμάζουσα οικονομία βασιζόμενη στα δεδομένα”
		Γλωσσικά και κειμενικά χαρακτηριστικά
		Όχι ιδιαίτερα εξειδικευμένο Ορισμοί όρων
T_en_gr_4_en	912	**Είδος**
		Δελτίο τύπου για πρόγραμμα πολιτικής για τον προγραμματισμό και την εναρμόνιση της χρήσης του ραδιοφάσματος στην Ε.Ε.
		Γλωσσικά και κειμενικά χαρακτηριστικά
		Κείμενο που απευθύνεται στο ευρύ κοινό Δημοσιογραφικός λόγος με ορολογία από τομείς Ε.Ε. και τεχνολογίας, καθώς και με ονομασίες οργανισμών, κ.λπ.
T_en_gr_5_en	720	**Είδος**
		Δελτίο τύπου Ε.Ε. για τα οφέλη της ψηφιακής οικονομίας

		Γλωσσικά και κειμενικά χαρακτηριστικά
		Κείμενο που απευθύνεται στο ευρύ κοινό Δημοσιογραφικός λόγος με ορολογία από τομείς Ε.Ε. και τεχνολογίας, καθώς και με ονομασίες οργανισμών, κ.λπ. Μικρή λίστα θεμάτων σχετικών με το αντικείμενο με επεξηγήσεις (τίτλοι και στις επεξηγήσεις χρήση γερούνδιου)
	Ιατρική	
M_en_gr_1_en	1.511	**Είδος**
		Παρουσίαση φαρμάκου
		Γλωσσικά και κειμενικά χαρακτηριστικά
		Μακροπερίοδος λόγος, διάφορα είδη προτάσεων και ρηματικών τύπων Εξειδικευμένη ορολογία Σύντομη αναφορά στην ασθένεια και μελέτες που πραγματοποιήθηκαν για το φάρμακο
M_en_gr_2_en	2.621	**Είδος**
		Απόσπαμα οδηγού διατροφής για άτομα που πάσχουν από νόσο του Πάρκινσον
		Γλωσσικά και κειμενικά χαρακτηριστικά
		Κείμενο που απευθύνεται στο ευρύ κοινό, συνεπώς με πολλές επεξηγήσεις και ορολογία όχι ιδιαίτερα εξειδικευμένη, το μεγαλύτερο μέρος της οποίας είναι από τον τομέα της διατροφής Συμβουλές, συνεπώς συνεχείς αναφορές στον αναγνώστη με χρήση προσωπικής αντωνυμίας, και δομές και ρηματικοί τύποι που εξυπηρετούν αυτόν τον σκοπό Μέρη με μακροπερίοδο λόγο και άλλα με απαριθμήσεις, κυρίως συμβουλών
M_en_gr_3_en	2.031	**Είδος**
		Ενημερωτικό για θεραπεία αποκατάστασης ατόμων με κινητικά προβλήματα
		Γλωσσικά και κειμενικά χαρακτηριστικά
		Μακροπερίοδος λόγος με μεγάλο αριθμό ιατρικών όρων αρκετά εξειδικευμένων Περιγραφές δραστηριοτήτων θεραπείας με διάφορους ρηματικούς τύπους, τόσο σε μορφή συνεχούς κειμένου όσο και σε μορφή

		απαρίθμησης
M_en_gr_4_en	472	**Είδος**
		Συχνές ερωτήσεις για το σύνδρομο του ευερέθιστου εντέρου
		Γλωσσικά και κειμενικά χαρακτηριστικά
		10 ερωταπαντήσεις σχετικά με το σύνδρομο Ερωτήσεις Ιατρική ορολογία όχι ιδιαίτερα εξειδικευμένη Απευθύνεται στο ευρύ κοινό
M_en_gr_5_en	2.737	**Είδος**
		Απόσπασμα επιστημονικού άρθρου
		Γλωσσικά και κειμενικά χαρακτηριστικά
		Τυπική δομή άρθρου με τίτλο, περίληψη, λέξεις-κλειδιά, κ.λπ. Μακροπερίοδος λόγος Πολλή και εξειδικευμένη ορολογία Ποικιλία φαινομένων (διάφορα είδη δευτερευουσών προτάσεων, πολλές συνδετικές λέξεις, διάφοροι ρηματικοί τύποι, κ.λπ.) Απευθύνεται σε ειδικούς
Δίκαιο		
L_en_gr_1_en	1.778	**Είδος**
		Οικουμενική Διακήρυξη για τα Ανθρώπινα Δικαιώματα
		Γλωσσικά και κειμενικά χαρακτηριστικά
		Νομικός λόγος
L_en_gr_2_en	1.307	**Είδος**
		Απόφαση του Συμβουλίου
		Γλωσσικά και κειμενικά χαρακτηριστικά
		Νομικός λόγος Τυπική δομή απόφασης
L_en_gr_3_en	1.394	**Είδος**
		Εκτελεστική απόφαση Επιτροπής
		Γλωσσικά και κειμενικά χαρακτηριστικά

		Νομικός λόγος Τυπική δομή απόφασης
L_en_gr_4_en	723	**Είδος**
		Απόσπασμα από καταπίστευμα
		Γλωσσικά και κειμενικά χαρακτηριστικά
		Αρχικό μέρος καταπιστεύματος με ορισμούς
L_en_gr_5_en	1.916	**Είδος**
		Απόσπασμα από κείμενο με ερωταπαντήσεις σχετικά με την κατάργηση της θανατικής ποινής
		Γλωσσικά και κειμενικά χαρακτηριστικά
		Κείμενο κοινωνιολογικού περιεχομένου με νομική ορολογία Απλός λόγος που επιδέχεται ελεύθερη μετάφραση αλλά με συγκεκριμένους όρους που παρουσιάζουν μεταφραστικές δυσκολίες (περιφραστική απόδοση, κ.λπ.) Περιέχει ερωτήσεις, ονομασίες οργανισμών και διεθνών συνθηκών, κ.λπ., καθώς και κάποια μικρά αυτούσια αποσπάσματα από τις συνθήκες
Ελληνικά-Αγγλικά		
Τεχνολογία		
T_gr_en_1_gr	547	**Είδος**
		Παρουσίαση ερευνητικού προγράμματος για επίπεδα ηλεκτρομαγνητικής ακτιβονολίας
		Γλωσσικά και κειμενικά χαρακτηριστικά
		Παρουσίαση στόχου προγράμματος, τμημάτων από τα οποία αποτελείται και συνοπτική περιγραφή λειτουργίας Κείμενο με αρκετά μακροπερίοδο λόγο και όχι πολύ εξειδικευμένη ορολογία
T_gr_en_2_gr	675	**Είδος**
		Απόσπασμα από οδηγό σπουδών Τμήματος Ηλεκτρονικών Μηχανικών και Μηχανικών Υπολογιστών
		Γλωσσικά και κειμενικά χαρακτηριστικά
		Απόσπασμα από τους αντικειμενικούς σκοπούς του τμήματος Μακροπερίοδος λόγος με διαφόρων ειδών προτάσεις Εξειδικευμένη ορολογία

		Όροι στην αγγλική σε παρένθεση Ονομασίες οργανισμών Αρκτικόλεξα
T_gr_en_3_gr	657	**Είδος**
		Απόσπασμα από οδηγό σπουδών Τμήματος Ηλεκτρονικών Μηχανικών και Μηχανικών Υπολογιστών
		Γλωσσικά και κειμενικά χαρακτηριστικά
		Παρουσίαση κάποιων από τα εργαστήρια του τμήματος Πολλοί και εξειδικευμένοι όροι Σε μεγάλο μέρος του κειμένου απαρίθμηση ερευνητικών περιοχών (όροι) Πολλά αγγλικά αρκτικόλεξα
T_gr_en_4_gr	892	**Είδος**
		Απόσπασμα από οδηγό σπουδών Τμήματος Ηλεκτρονικών Μηχανικών και Μηχανικών Υπολογιστών
		Γλωσσικά και κειμενικά χαρακτηριστικά
		Αναλυτικές παρουσιάσεις κάποιων από τα μαθήματα του τμήματος Πλειονότητα των προτάσεων χωρίς ρήμα, απαρίθηση δηλαδή των στοιχείων που αποτελούν την ύλη του μαθήματος, τα οποία είναι στο σύνολό τους τεχνικοί όροι Πλήθος όρων στα αγγλικά σε παρένθεση και αγγλικά αρκτικόλεξα
T_gr_en_5_gr	1.313	**Είδος**
		Απόσπασμα σχετικά με ηλεκτρονικά όργανα και βοηθήματα ναυσιπλοΐας
		Γλωσσικά και κειμενικά χαρακτηριστικά
		Περιγραφές χρήσης ηλεκτρονικών οργάνων-βοηθημάτων ναυσιπλοΐας Πλήθος όρων στα αγγλικά και αρκτικόλεξα (και επεξηγηματικά εντός παρενθέσεων και εκτός) Επεξηγήσεις για το ευρύ κοινό Μακροπερίοδος λόγος με διαφόρων ειδών προτάσεις
	Ιατρική	
M_gr_en_1_gr	904	**Είδος**
		Ενημερωτικό για λαπαροσκόπηση

		Γλωσσικά και κειμενικά χαρακτηριστικά
		Περιγραφή παθήσεων και ιατρικών τεχνικών Ενημερωτικό για το ευρύ κοινό, συνεπώς παρουσία ορολογίας αλλά με επεξηγήσεις για τον μη ειδικό
M_gr_en_2_gr	396	**Είδος**
		Ενημερωτικό για την υποβοηθούμενη αναπαραγωγή
		Γλωσσικά και κειμενικά χαρακτηριστικά
		Περιγραφή τεχνικών υποβοηθούμενης αναπαραγωγής Ενημερωτικό για το ευρύ κοινό, συνεπώς παρουσία ορολογίας αλλά με επεξηγήσεις για τον μη ειδικό Κάποιοι από τους ιατρικούς όρους και στα αγγλικά σε παρένθεση Αρκτικόλεξα στα αγγλικά (ιατρικοί όροι)
M_gr_en_3_gr	330	**Είδος**
		Ενημερωτικό για τον φλεβίτη
		Γλωσσικά και κειμενικά χαρακτηριστικά
		Περιγραφή πάθησης και θεραπειών Ενημερωτικό για το ευρύ κοινό, συνεπώς παρουσία ορολογίας αλλά με επεξηγήσεις για τον μη ειδικό Κάποιοι από τους ιατρικούς όρους και στα αγγλικά σε παρένθεση Λίστα επιπλοκών, η οποία αποτελείται από ιατρικούς όρους εκτός συγκειμένου
M_gr_en_4_gr	158	**Είδος**
		Λίστα υπηρεσιών παιδοκαρδιολογίας
		Γλωσσικά και κειμενικά χαρακτηριστικά
		Ουσιαστικοποιημένη λίστα ιατρικών τεχνικών Μικρές προτάσεις με παρατακτική σύνδεση Αρκετά εξειδικευμένη ορολογία Αρκτικόλεξα Αγγλικοί όροι σε παρένθεση
M_gr_en_5_gr	1.311	**Είδος**
		Ενημερωτικό για υποστήριξη ατόμων με διατροφικές διαταραχές
		Γλωσσικά και κειμενικά χαρακτηριστικά

		Κείμενο όπου αναφέρεται ο ορισμός των διατροφικών διαταραχών και επεξηγήσεις σχετικά με τα διάφορα είδη τους Μακροπερίοδος λόγος Όχι ιδιαίτερα εξειδικευμένη ορολογία Αρκτικόλεξα Αγγλικοί και λατινικοί όροι Λίστα με χαρακτηριστικά γνωρίσματα διαταραχής σε μορφή ουσιαστικών και ρημάτων σε γ΄ ενικό πρόσωπο
	Δίκαιο	
L_gr_en_1_gr	656	**Είδος**
		Απόσπασμα για τη νομοθετική διαδικασία στη Βουλή των Ελλήνων
		Γλωσσικά και κειμενικά χαρακτηριστικά
		Απλός λόγος (όχι νομικός) με περιορισμένο αριθμό τυπικών δομών νομικού λόγου και νομική ορολογία Πλήθος αναφορών σε άρθρα νόμων και αρκτικόλεξα
L_gr_en_2_gr	443	**Είδος**
		Απόσπασμα για τις διαδικασίες κοινοβουλευτικού ελέγχου στη Βουλή των Ελλήνων
		Γλωσσικά και κειμενικά χαρακτηριστικά
		Απλός λόγος (όχι νομικός) με περιορισμένο αριθμό τυπικών δομών και νομική ορολογία Πλήθος αναφορών σε άρθρα νόμων
L_gr_en_3_gr	322	**Είδος**
		Απόσπασμα από ενημερωτικό για δικαιολογητικά και διαδικασία χορήγησης κάρτας πολιτισμού
		Γλωσσικά και κειμενικά χαρακτηριστικά
		Απαρίθμηση δικαιολογητικών και βημάτων διαδικασίας Στην απαρίθμηση δικαιολογητικών ουσιαστικοποιημένες δομές και στα βήματα διαδικασίας ρηματικές Πολλές αναφορές σε στοιχεία ελληνικής πραγματικότητας (οργανισμοί, επίσημα έγγραφα, κ.λπ.)
L_gr_en_4_gr	497	**Είδος**
		Απόσπασμα για τις ειδικές διαδικασίες κοινοβουλευτικού ελέγχου στη Βουλή των Ελλήνων

		Γλωσσικά και κειμενικά χαρακτηριστικά
		Απλός λόγος (όχι νομικός) με περιορισμένο αριθμό τυπικών δομών και νομική ορολογία Πλήθος αναφορών σε άρθρα νόμων
L_gr_en_5_gr	269	**Είδος**
		Απόσπασμα από ενημερωτικό για δικαιολογητικά και διαδικασία για συμπληρωματικές παροχές μητρότητας
		Γλωσσικά και κειμενικά χαρακτηριστικά
		Απαρίθμηση δικαιολογητικών και βημάτων διαδικασίας Στην απαρίθμηση δικαιολογητικών ουσιαστικοποιημένες δομές και στα βήματα διαδικασίας ρηματικές Αναφορές σε στοιχεία ελληνικής πραγματικότητας (οργανισμοί, επίσημα έγγραφα, κ.λπ.)

Πίνακας 3: Κείμενα σώματος: είδη και χαρακτηριστικά

4.1.4 Συλλογή μεταφράσεων

Για τα πειράματα της παρούσας εργασίας χρησιμοποιήθηκαν συνολικά 7 μεταφράσεις ανά κείμενο: 3 μεταφράσεις-αναφορά και 4 μεταφράσεις μηχανικών συστημάτων.

Όσο αφορά τις μεταφράσεις-αναφορά, μία από αυτές συλλέχθηκε από το διαδίκτυο από τις προαναφερθείσες πηγές. Για αυτές τις μεταφράσεις δεν μπορούμε να γνωρίζουμε τι είδους μεταφραστικά εργαλεία χρησιμοποιήθηκαν, ωστόσο βεβαιωθήκαμε ότι σε κάθε περίπτωση είχαν υποστεί επεξεργασία ή/και επιμέλεια από ανθρώπους. Οι υπόλοιπες παράχθηκαν από επαγγελματίες μεταφραστές με την οδηγία να χρησιμοποιήσουν οποιοδήποτε μεταφραστικό εργαλείο επιθυμούν με την παράκληση να γίνει αναφορά στα συγκεκριμένα εργαλεία, αλλά να μη χρησιμοποιήσουν τη μετάφραση που υπάρχει στο διαδίκτυο. Κανένας από τους μεταφραστές δεν χρησιμοποίησε κάποιο σύστημα μηχανικής μετάφρασης. Εργαλεία χρησιμοποιήθηκαν για εύρεση όρων και ήταν τα εξής: η μεταφραστική μνήμη Trados, η ορολογική βάση της Ε.Ε., το εργαλείο Linguee[44], η βάση EUR-Lex, η μηχανή Google Translate, και το παράλληλο σώμα του EMEA[45]. Οι πληροφορίες αυτές είναι απαραίτητες για την ερμηνεία των αποτελεσμάτων της

[44] Μεταφραστικό εργαλείο, το οποίο συνδυάζει ένα κύριο λεξικό και μια μηχανή αναζήτησης, με το οποίο ανασύρονται οι λέξεις με τα αντίστοιχά τους σε συγκείμενο από διάφορες πηγές. [http://www.linguee.gr/ελληνικά-αγγλικά/page/about.php, τελευταία προσπέλαση 20/03/15].

[45] Σώμα κειμένων του Ευρωπαϊκού Οργανισμού Φαρμάκων (EMEA, European Medicines Agency) [http://www.ema.europa.eu/ema/, τελευταία προσπέλαση 20/03/15], το οποίο βρίσκεται ελεύθερα διαθέσιμο στο διαδίκτυο και στο οποίο μπορούν να αναζητηθούν όροι μέσω κάποιου εργαλείου

αυτόματης αξιολόγησης, όπου και χρησιμοποιούνται αυτές οι μεταφράσεις-αναφορά, διότι υποθέτουμε ότι τα στατιστικά συστήματα που αξιολογούμε χρησιμοποιούν σώματα από τα οποία αντλούν υλικό και τα εν λόγω εργαλεία. Καταβλήθηκε κάθε δυνατή προσπάθεια ώστε οι μεταφράσεις που επιλέχθηκαν και όσες παράχθηκαν για τις ανάγκες της παρούσας εργασίας να πληρούν την προϋπόθεση της ποιοτικής μετάφρασης[46]. Οι αναφορές σε αυτές τις μεταφράσεις γίνονται με τη μορφή HT1, HT2 και HT3, όπου HT1 είναι η διαθέσιμη μετάφραση στο διαδίκτυο, όπου αυτή υπήρχε.

Οι αυτόματες μεταφράσεις παράχθηκαν από τρία στατιστικά συστήματα μηχανικής μετάφρασης και ένα σύστημα που βασίζεται σε χρήση κανόνων. Συγκεκριμένα, από τα στατιστικά τα δύο είναι εμπορικά ελεύθερα διαθέσιμα συστήματα, το Google Translate και το Bing Translator, και ένα πρωτότυπο σύστημα υπό κατασκευή, το οποίο αναπτύσσεται από το ΙΕΛ με χρήση του συστήματος εκπαίδευσης στατιστικών συστημάτων μηχανικής μετάφρασης Moses [Koehn, Ph. κ.ά., 2007]. Τέλος, το Systran, επίσης εμπορικό ελεύθερα διαθέσιμο σύστημα, χρησιμοποιήθηκε ως εκπρόσωπος των συστημάτων που βασίζονται σε κανόνες. Οι ονομασίες των μεταφράσεων περιέχουν την πληροφορία του συστήματος με τη μορφή MT1 (Google Translate), MT2 (Bing Translator), MT3 (Systran) και MT4 (Moses Baseline) (π.χ. L_en_gr_1_MT4). Στο εξής για λόγους συντομίας χρησιμοποιείται η ονομασία Google για αναφορά στο Google Translate, η ονομασία Bing για αναφορά στο Bing Translator και η ονομασία Moses για αναφορά στο Moses Baseline.

Ακολουθούν κάποιες πληροφορίες σχετικά με τις ανωτέρω μηχανές.

Google Translate: Στην ιστοσελίδα του Google Translate (https://translate.google.com) δεν τίθεται κάποιος περιορισμός στον αριθμό χαρακτήρων του κειμένου προς μετάφραση και σε καμία περίπτωση δεν αδυνατούσε το σύστημα να δώσει αποτέλεσμα.

Bing Translator: Σε αυτή την εφαρμογή (http://www.bing.com/translator/) τίθεται ο περιορισμός των 5.000 χαρακτήρων. Επίσης σε περιπτώσεις χαρακτήρων που δεν αναγνωρίζονται, το σύστημα αδυνατεί να δώσει αποτελέσματα.

Systran: Το Systran είναι ένα από τα συστήματα που χρησιμοποιούνται στην ιστοσελίδα http://itranslate4.eu και συγκεκριμένα είναι το μοναδικό που λειτουργεί σε αυτή τη σελίδα για το γλωσσικό ζεύγος αγγλικά-ελληνικά. Εδώ παρέχεται η δυνατότητα επιλογής πεδίου (για τα κείμενά μας μπορούσαμε να επιλέξουμε μεταξύ των πεδίων “Γενικό”, “Δίκαιο και Διοίκηση”,

γλωσσικής επεξεργασίας.

[46] Θεωρείται δεδομένο ότι οι μεταφράσεις των φορέων της Ε.Ε. είναι ποιοτικές [Steinberger, R. κ.ά., 2014].

“Τεχνολογία” και “Υγεία και Βιολογία”), χωρίς ωστόσο να παρουσιάζεται καμία διαφορά στο αποτέλεσμα είτε χρησιμοποιείται η επιλογή γενικού κειμένου είτε η επιλογή εξειδικευμένου πεδίου. Τέλος, σε αυτή την εφαρμογή υπάρχει ο περιορισμός των 1.000 χαρακτήρων.

Moses Baseline: Στατιστικό σύστημα υπό ανάπτυξη, το οποίο βασίζεται στο μεταφραστικό σύστημα Moses. Για την εκπαίδευση του συστήματος χρησιμοποιήθηκαν σώματα κειμένων από τους τομείς του δικαίου, της ιατρικής και της ειδησεογραφίας για το μεταφραστικό μοντέλο, σώμα γενικό αντιπροσωπευτικό της αγγλικής γλώσσας για το γλωσσικό μοντέλο στα αγγλικά, και σώματα από τους τομείς του δικαίου, της ιατρικής και της ειδησεογραφίας για το γλωσσικό μοντέλο στα ελληνικά. Τα παράλληλα σώματα που χρησιμοποιήθηκαν στην εκπαίδευσή του αναφέρονται στον Πίνακα 4 και τα μονόγλωσσα στον Πίνακα 5.

Σώμα	**Πεδίο**	**Ζεύγη προτάσεων** κατόπιν αφαίρεσης επαναλαμβανόμενων ζευγών (cleaned)	**Αρ. ελληνικών tokens** (cleaned)	**Αρ. αγγλικών tokens** (cleaned)
JRC-ACQUIS	Δίκαιο	731.436	20.694.965	20.491.476
ECB	Δίκαιο	92.627	2.240.920	2.196.075
SE-TIMES	Ειδησεογραφία	224.460	5.649.873	5.385.768
EMEA	Ιατρική	185.018	3.213.377	3.012.127
ΣΥΝΟΛΟ		**1.233.541**	**31.799.135**	**31.085.446**

Πίνακας 4

Σώμα	**Γλώσσα**	**Ζεύγη προτάσεων** κατόπιν αφαίρεσης επαναλαμβανόμενων ζευγών (cleaned)	**Αρ. tokens** (cleaned)
JRC-ACQUIS	Ελληνικά	731.436	20.694.965
ECB	Ελληνικά	92.627	2.240.920
SE-TIMES	Ελληνικά	224.460	5.649.873
EMEA	Ελληνικά	185.018	3.213.377
ENET	Ελληνικά	1.118.020	30.430.168
ΚΑΘΗΜΕΡΙΝΗ	Ελληνικά	1.857.356	51.495.894
ΝΕΑ	Ελληνικά	2.724.931	63.237.546
ΣΥΝΟΛΟ	**Ελληνικά**	**6.933.848**	**176.962.743**
BNC	**Αγγλικά**	**6.020.864**	**113.383.949**

Πίνακας 5

4.1.5 Επεξεργασία κειμένων

Τα κείμενα αποθηκεύτηκαν σε μορφή απλού κειμένου με κωδικοποίηση utf-8 χωρίς BOM χωρισμένα σε περίοδο ανά σειρά, όπου περίοδος είναι ένα “τμήμα (γραπτού) λόγου ή σύνολο προτάσεων με αυτοτελές διανόημα, του οποίου το τέλος ορίζεται συνήθ. με κάτω τελεία (ή με ερωτηματικό, θαυμαστικό, αποσιωπητικά)”, σύμφωνα με το Λεξικό της Κοινής Νεοελληνικής (Τριανταφυλλίδη).

Ακολούθησε στοίχιση των κειμένων ανά ομάδες, ώστε τα σώματα κειμένων να είναι παράλληλα σε επίπεδο προτάσεων. Αυτή η επεξεργασία είναι απαραίτητη, διότι κατά τη μεταφραστική διαδικασία από ανθρώπους είναι συνηθισμένη πρακτική να αποδίδεται μία περίοδος σε περισσότερες (χωρισμός περιόδου), περισσότερες από μία περιόδους σε μία (ένωση/συγχώνευση περιόδων), ακόμη και μία περίοδος σε καμία (αποδεκτή μη μετάφραση για λόγους π.χ. πραγματολογικούς ή σε περιπτώσεις μεταγλωσσικών πληροφοριών μέσα στο κείμενο). Συνεπώς, εντοπίσαμε όλες αυτές τις περιπτώσεις και ο χωρισμός των περιόδων σε σειρές έγινε έτσι ώστε να υπάρχει αντιστοιχία μεταξύ όλων των κειμένων της ίδιας ομάδας. Αυτό επιτεύχθηκε με τον χωρισμό και τη συγχώνευση περιόδων και στην πράξη σημαίνει ότι όταν μιλάμε για περίοδο, μπορεί να αναφερόμαστε σε δύο ή και τρεις περιόδους μαζί σε κάποια κείμενα, οι οποίες αντιστοιχούν σε μία σε κάποια από τις άλλες μεταφράσεις ή στο πρωτότυπο.

Κατά την επεξεργασία των κειμένων εντοπίστηκαν λάθη σε κάποια από τα πρωτότυπα κείμενα και σε κάποιες από τις μεταφράσεις-αναφορά, τα οποία διορθώθηκαν στην τελική μορφή των στοιχισμένων κειμένων, προκειμένου να μπορέσει να λειτουργήσει το σύστημα αυτόματης αξιολόγησης. Τέτοια λάθη ήταν αναγραμματισμοί, πλεονάζουσες περίοδοι (δηλαδή περίοδοι που δεν αποτελούσαν μετάφραση κάποιου σημείου του πρωτότυπου), ελλείπουσες (μη μεταφρασμένες) περίοδοι, αναδιατεταγμένες περίοδοι μέσα στο κείμενο, επαναλαμβανόμενες περίοδοι, καθώς και χαρακτήρες μη αναγνωρίσιμοι παρά την επιλογή της σωστής κωδικοποίησης.

Τα πρωτότυπα κείμενα μεταφράστηκαν από τα συστήματα στη μορφή που είχαν πριν και μετά τον χωρισμό σε περιόδους και τη στοίχιση. Στα στατιστικά συστήματα αυτό μπορεί να έχει ως αποτέλεσμα διαφορετικά εξαγόμενα λόγω του χωρισμού σε περιόδους που θα κάνει το ίδιο το σύστημα κατά τη διαδικασία της μηχανικής μετάφρασης. Ακολουθούν δύο παραδείγματα.

Διαφορετικά αποτελέσματα της ίδιας πρότασης λόγω διαφορετικού χωρισμού:

- “(α) **Computer Monitor**:” / “(α) **οθόνη του υπολογιστή**:”

- “1. **Μπορώ σίγουρα να έχουν** σύνδρομο ευερέθιστου εντέρου;” / “1. **Έχω σίγουρα έχει** σύνδρομο ευερέθιστου εντέρου;”

Εάν ληφθεί υπόψη ότι σε δύο από τα συστήματα που χρησιμοποιήθηκαν υπήρχε περιορισμός χαρακτήρων, ανάλογα με τον χωρισμό του κειμένου που έπρεπε να γίνει, θα μπορούσαμε να έχουμε διαφορετικά αποτελέσματα. Για λόγους ομοιομορφίας λοιπόν σε όλα τα κείμενα και όλα τα συστήματα, χρησιμοποιήθηκαν για την αξιολόγηση οι μεταφράσεις που δόθηκαν από τα κείμενα στην τελική μορφή (με χωρισμό περιόδων και στοίχιση). Επίσης τα συστήματα, όταν τροφοδοτηθούν με κείμενα με διαχωρισμένες περιόδους, δίνουν μετάφρασμα με τον ίδιο διαχωρισμό.

Ύστερα από όλες τις παραπάνω διαδικασίες επιλογής, μετάφρασης, επεξεργασίας, στοίχισης και διόρθωσης, καταλήγουμε με ένα σύνολο 1.726 προτάσεων (περιόδων), 1.103 προτάσεις στα αγγλικά και 623 προτάσεις στα ελληνικά με τις αντίστοιχες μεταφράσεις τους, αλλά και με ένα εμπλουτισμένο σώμα κειμένων, το οποίο πλέον αποτελείται από 30 πρωτότυπα με 7 μεταφράσεις τους.

Τέλος, τα κείμενα μετατράπηκαν σε μορφή xml, ώστε να μπορούν να χρησιμοποιηθούν στο σύστημα αξιολόγησης Bleu και στο εργαλείο επισημείωσης Appraise.

4.2 Μεθοδολογία αξιολόγησης

Τόσο για την κατεύθυνση από αγγλικά προς ελληνικά όσο και για την κατεύθυνση από ελληνικά προς αγγλικά χρησιμοποιήσαμε μεθόδους και αυτόματης και μη αυτοματοποιημένης αξιολόγησης. Για την αυτόματη αξιολόγηση χρησιμοποιήθηκε η καθιερωμένη πλέον στον χώρο μετρική Bleu[47] για όλα τα κείμενα του σώματος. Με την εν λόγω μετρική εξαγάγαμε αποτελέσματα ανά κείμενο τόσο με τρεις μεταφράσεις-αναφορά όσο και με δύο, καθώς και αποτελέσματα ανά πρόταση. Η χρήση Bleu με δύο μόνο μεταφράσεις-αναφορά επιλέχθηκε, γιατί θεωρήθηκε πιθανό το ενδεχόμενο η πρώτη μετάφραση-αναφορά (HT1) να έχει χρησιμοποιηθεί στο σώμα εκπαίδευσης των εμπορικών στατιστικών συστημάτων, κάτι το οποίο εξασφαλίσαμε ότι δεν συνέβη στην περίπτωση του Moses. Τα αποτελέσματα ανά πρόταση συγκρίθηκαν με τα αντίστοιχα αποτελέσματα της μη αυτοματοποιημένης αξιολόγησης με σκοπό την εξαγωγή συμπερασμάτων σχετικά με τις επιδόσεις της μετρικής Bleu ανά πρόταση.

Από το σύνολο των κειμένων επιλέχθηκαν κάποια για μη αυτοματοποιημένη αξιολόγηση με χρήση

[47] Script “mteval-v13a.pl” (http://www.itl.nist.gov/iad/mig/tests/mt/2009/, τελευταία προσπέλαση 19/06/2015).

διαφόρων τεχνικών, οι οποίες παρουσιάζονται στη συνέχεια. Στα κεφάλαια που ακολουθούν περιγράφεται η διαδικασία επιλογής τυπολογίας λαθών και εργαλείου επισημείωσης για τη μη αυτοματοποιημένη αξιολόγηση, καθώς και οι οδηγίες επισημείωσης που ακολουθήθηκαν.

4.2.1 Ταξινόμηση Λαθών και Μετεπεξεργασία

Η ταξινόμηση λαθών επιλέχθηκε ως τρόπος αξιολόγησης με στόχο την παρατήρηση συγκεκριμένων κατηγοριών λαθών στα μηχανικά μεταφράσματα και τη μετέπειτα δυνατότητα αξιοποίησης των παρατηρήσεων αυτών για βελτίωση του συστήματος. Η διαδικασία της ταξινόμησης λαθών είναι άμεσα εξαρτημένη από τη διόρθωση αυτών των λαθών. Αυτό συμβαίνει γιατί το πρώτο στάδιο της ταξινόμησης είναι ο εντοπισμός του λάθους, και η γνώση του λάθους προϋποθέτει τη γνώση του σωστού, συνεπώς η νοητική διαδικασία εντοπισμού του λάθους εμπεριέχει συνειδητά ή ασυνείδητα την ορθή επιλογή που θα έπρεπε να υπάρχει στη θέση της λανθασμένης, δηλαδή τη διόρθωση ως αποτέλεσμα της μετεπεξεργασίας. Ο Bojar, O. [2011] καταλήγει στο συμπέρασμα ότι οι επισημειωτές χρησιμοποιούν, ακόμη κι αν αυτό δεν γίνεται ρητά, κάποιο αποδεκτό μετάφρασμα και επισημειώνουν τις απαραίτητες αλλαγές προκειμένου να φτάσει το κείμενο σε αυτό το αποδεκτό μετάφρασμα. Επιπλέον, όταν η ταξινόμηση και η διόρθωση γίνονται παράλληλα, η οπτική επαφή με τη διορθωμένη πρόταση συμβάλλει στη βελτιστοποίηση του εντοπισμού των λαθών και στον έλεγχο του αριθμού των επισημειωμένων λαθών, δηλαδή ο επισημειωτής βεβαιώνεται ότι έχει επισημειώσει τα ελάχιστα δυνατά λάθη, εάν αυτό απαιτείται.

Για να γίνουν κατανοητά τα παραπάνω, παραθέτουμε ένα παράδειγμα. Στη μηχανικά μεταφρασμένη πρόταση “Στρες προκαλεί την παραγωγή κυτοκινών συγκεκριμένα στο κεντρικό νευρικό σύστημα ενεργεί στον **τοίχο** του παχέος εντέρου.” εντοπίζεται ένα λάθος στην επιλογή της λέξης “τοίχος”, η οποία χρησιμοποιείται αντί του όρου “τοίχωμα”. Ο επισημειωτής εντοπίζει το λάθος επειδή ακριβώς θεωρεί ότι στη θέση του θα έπρεπε να υπάρχει κάποιο άλλο στοιχείο. Οφείλει λοιπόν να επισημειώσει με την ανάλογη ετικέτα σύμφωνα με την τυπολογία που χρησιμοποιεί, στην προκειμένη περίπτωση κάποια κατηγορία που να σχετίζεται με την απόδοση ορολογίας. Επισημειώνοντάς το όμως δηλώνεται πως η πρόταση θα ήταν σωστή ή θα είχε λιγότερα λάθη, αν αντικαθιστούσαμε τη λανθασμένη με την ορθή λέξη: “Στρες προκαλεί την παραγωγή κυτοκινών συγκεκριμένα στο κεντρικό νευρικό σύστημα ενεργεί στον **τοίχωμα** του παχέος εντέρου.”, ωστόσο δεν ισχύει αυτό, καθώς “δημιουργείται” ένα άλλο λάθος, σε αυτή την περίπτωση λάθος συμφωνίας με το άρθρο (“στον τοίχωμα”).

Με βάση τα παραπάνω, επιλέγουμε να χρησιμοποιήσουμε παράλληλα ανά πρόταση τις εργασίες της ταξινόμησης λαθών και της μετεπεξεργασίας, μεταξύ των οποίων θεωρούμε ότι θα πρέπει να υπάρχει συμφωνία. Συνεπώς, το λάθος συμφωνίας στο παραπάνω παράδειγμα, το οποίο δημιουργείται από τη διόρθωση του λάθους ορολογίας, επισημειώνεται, ώστε ο αριθμός λαθών της ταξινόμησης να συμπίπτει με τον αριθμό διορθώσεων της μετεπεξεργασίας. Αυτή η επιλογή συνάδει και με το συμπέρασμα των Snover, M. κ.ά. [2006], οι οποίοι αποδεικνύουν ότι είναι προτιμότερη η μέθοδος δημιουργίας μιας νέας μετάφρασης-αναφορά (η μετάφραση που προκύπτει από τη μετεπεξεργασία, η διορθωμένη δηλαδή μηχανική μετάφραση) και μέτρησης λαθών από την έκφραση υποκειμενικών κρίσεων.

Για αυτές τις δύο εργασίες επισημείωσης ακολουθείται η τεχνική των ελάχιστων δυνατών διορθώσεων [Przybocki, M. κ.ά., 2011] και η επισημείωση της κάθε πρότασης, εκτός εάν πρόκειται για μικρή πρόταση, γίνεται τμηματικά (ταξινόμηση λαθών σε ένα μικρό τμήμα της πρότασης, μετεπεξεργασία αυτού του τμήματος, κ.ο.κ., και τελικός έλεγχος της μετεπεξεργασμένης πρότασης σε αντιπαράθεση με την πρωτότυπη). Τέλος, με το πέρας της επισημείωσης κάθε κειμένου πραγματοποιείται διόρθωση της επισημείωσης με βάση το μετεπεξεργασμένο κείμενο.

4.2.2 Επιλογή τυπολογίας λαθών

Καθώς δεν παρέχονται ορισμοί των κατηγοριών και των επιπέδων σοβαρότητας λάθους του εργαλείου Appraise, παρουσιάζουμε πρώτα τους δικούς μας ορισμούς αυτών και στη συνέχεια τις συγκρίσεις ανά ζεύγη των τυπολογιών MQM, DQF και Appraise, με στόχο την επιλογή της κατάλληλης τυπολογίας και του κατάλληλου εργαλείου για την αξιολόγηση των κειμένων μας.

4.2.2.1 Ορισμός κατηγοριών Appraise

Βασιζόμενοι σε δοκιμές σε ευθεία και αντίστροφη μετάφραση αλλά και στις πληροφορίες που παρέχονται στις τυπολογίες DQF και MQM, ορίζουμε τις κατηγορίες του Appraise στον Πίνακα 4. Οι όροι που αναφέρονται στους ορισμούς των κατηγοριών νοούνται ως εξής:

Επιθυμητό μετάφρασμα: Κείμενο στη γλώσσα-στόχος το οποίο επιτελεί τη λειτουργία που επιτελεί το πρωτότυπο και φέρει όλες τις πληροφορίες του πρωτότυπου και μόνο αυτές, σε γλωσσικό μητρώο ίδιο ή αντίστοιχο του μητρώου του πρωτότυπου, σε ορθό και κατανοητό λόγο ο οποίος χρησιμοποιείται συνήθως στη γλώσσα-στόχος σε κείμενα από το πεδίο όπου εντάσσεται το μετάφρασμα και ο οποίος δεν προσκρούει στο γλωσσικό αισθητήριο του ικανού αναγνώστη.

Ικανός αναγνώστης: Ο αναγνώστης με επαρκείς γνώσεις της γλώσσας-στόχος και του πεδίου στο οποίο εντάσσεται το κείμενο προς αξιολόγηση.

Λάθος: Οποιοδήποτε στοιχείο του μεταφράσματος παρεμποδίζει τον χαρακτηρισμό του ως επιθυμητό.

Όρος: Οι λέξεις εκείνες οι οποίες ανήκουν σε συγκεκριμένο πεδίο και για την κατανόηση των οποίων χρειάζονται έστω και κάποιες εξειδικευμένες γνώσεις ή έρευνα (π.χ. η λέξη “εξετάσεις” σε ιατρικό κείμενο δεν θεωρείται όρος, η λέξη όμως “δυσανεξία” θεωρείται).

Λειτουργικές λέξεις: τα κλιτικά, οι συμπληρωματικοί δείκτες “να”, “ότι” και “που”, οι προθέσεις, οι σύνδεσμοι, οι δείκτες “θα” και “να”, τα άρθρα και τα βοηθητικά ρήματα.

Στα παρακάτω φαινόμενα γίνεται λόγος πάντα για λέξεις, γιατί η επισημείωση γίνεται σε επίπεδο λέξης.

Missing words	**Ελλείπουσες Λέξεις**
Φαινόμενο:	Μη χρήση λέξης/λέξεων, εξαιρουμένων των λειτουργικών λέξεων, που θα πρέπει να προστεθεί/-ούν για να πληρούνται τα κριτήρια του επιθυμητού μεταφράσματος
Παράδειγμα:	“Everyone has the right to life, liberty and security **of person**.” → “κάθε πρόσωπο έχει δικαίωμα στη ζωή, την ελευθερία και ασφάλεια.”
Δεν πρόκειται για κατηγορία επισημείωσης σε επίπεδο λέξης αλλά χαρακτηρισμός που αποδίδεται σε ολόκληρη την πρόταση. Στο παράδειγμα παρατηρείται απουσία λέξης περιεχομένου, συγκεκριμένα ουσιαστικού. Θεωρείται απαραίτητη η μετάφρασή της, γιατί πρόκειται για νομικό κείμενο.	
Too Many Errors	**Υπερβολικός Αριθμός Λαθών**
Φαινόμενο:	Παρουσία τέτοιου αριθμού και είδους λαθών που καθιστούν την πρόταση μη βελτιώσιμη
Παράδειγμα:	-
Δεν πρόκειται για κατηγορία επισημείωσης σε επίπεδο λέξης αλλά χαρακτηρισμός που αποδίδεται σε ολόκληρη την πρόταση, προκειμένου να μην αξιολογηθεί. Δεν συναντήθηκε τέτοιο φαινόμενο στο σώμα κειμένων μας.	
Terminology	**Ορολογία**
Φαινόμενο:	Χρήση όρου ή λέξης (μη όρου) που θα πρέπει να αντικατασταθεί από (άλλο) όρο για να πληρούνται τα κριτήρια του επιθυμητού μεταφράσματος
Παράδειγμα:	“4. Do I have a food **intolerance**? ” → “4.Do έχω μια **αδιαλλαξία** τροφίμων;”

Πρόκειται για πολυλεκτικό όρο (“food intolerance”) ο οποίος δεν αποδίδεται ως τέτοιος αλλά το κάθε στοιχείο του ξεχωριστά. Η σωστή επιλογή θα ήταν “δυσανεξία”.

Lexical Choice	**Λεξιλογική Επιλογή**
Φαινόμενο:	Επιλογή λέξης, τόσο λέξης περιεχομένου όσο και λειτουργικής λέξης, η οποία δεν πληροί τα κριτήρια του επιθυμητού μεταφράσματος
Παράδειγμα:	“For this reason it is important to seek the psychologist's help to **treat** both the cause and the symptoms of IBS.” → “Για αυτόν τον λόγο είναι σημαντικό να επιδιωχθεί η βοήθεια του ψυχολόγου για να **μεταχειριστεί** και την αιτία και τα συμπτώματα IBS.”
Σε αυτή την κατηγορία υπάγονται μόνο σημασιολογικά λανθασμένες επιλογές. Στο παράδειγμα αποδίδεται η λέξη του πρωτότυπου με μία από τις πιθανές λέξεις απόδοσης στην ελληνική, η οποία όμως δεν είναι ορθή βάσει συγκειμένου.	

Syntax (Ordering)	**Σύνταξη [Σειρά Λέξεων]**
Φαινόμενο:	Λάθος στην ενδεδειγμένη σειρά των λέξεων
Παράδειγμα:	“5. **μου κοιλιά** είναι πρησμένο και νιώθω άβολα.”
Η κτητική αντωνυμία θα έπρεπε να έπεται του ουσιαστικού και όχι να προηγείται, συνεπώς έχουμε λανθασμένη σειρά λέξεων. Παρόλο που σε αυτή την τυπολογία δεν γίνεται η διάκριση μεταξύ πιστότητας και γραμματικότητας, αυτή δεν παύει να υπάρχει. Η λανθασμένη σειρά λέξεων είναι ένα φαινόμενο το οποίο γίνεται αντιληπτό διαβάζοντας μόνο τη μετάφραση (γραμματικότητα), χωρίς να είναι αναγκαία η αντιπαραβολή με το πρωτότυπο. Καθώς η επισημείωση γίνεται σε μία μόνο λέξη, πρέπει να ορίζεται στις οδηγίες η λέξη στην οποία επισημειώνεται το λάθος σύνταξης.	

Insertion (Extra Word)	**Προσθήκη Λέξης**
Φαινόμενο:	Χρήση λέξης ή συμβόλου, συμπεριλαμβανομένων λειτουργικών λέξεων, που θα πρέπει να διαγραφεί για να πληρούνται τα κριτήρια του επιθυμητού μεταφράσματος
Παράδειγμα:	“Μια φουσκωμένη κοιλιά μπορεί να προκαλέσει **τη** σημαντική ταλαιπωρία στον ασθενή.”
Χρήση οριστικού άρθρου σε σημείο που θα έπρεπε να παραλείπεται. Η προσθήκη λέξεων περιεχομένου είναι θέμα πιστότητας, ενώ η προσθήκη λειτουργικών λέξεων είναι θέμα γραμματικότητας. Σε αυτή την τυπολογία δεν γίνεται η διάκριση, συνεπώς στην ίδια κατηγορία παρουσιάζονται τόσο θέματα πιστότητας όσο και γραμματικότητας.	

Morphology	**Μορφολογία**
Φαινόμενο:	Λάθη στην επιλογή χρόνου, έγκλισης, όψης
Παράδειγμα:	“Πολλοί IBS ασθενείς **αναπτύξουν** δυσανεξία σε ορισμένες τροφές.”
Λανθασμένη όψη ρηματικού τύπου. Πρόκειται για φαινόμενο γραμματικότητας, γι' αυτόν τον λόγο και δίνεται στο παράδειγμα μόνο η μετάφραση.	

Φαινόμενο:	Λανθασμένο μέρος του λόγου
Παράδειγμα:	“Αυτή τη στιγμή μόνο το 50% των καταναλωτών στο σε απευθείας σύνδεση **κατάστημα** της Ε.Ε..”

Όπως είδαμε και στον Πίνακα 1, στο συγκεκριμένο παράδειγμα ενδέχεται ο μέσος αναγνώστης μόνο με την ανάγνωση της μετάφρασης να μην εντόπιζε λάθος μέρους του λόγου στη λέξη “κατάστημα”. Η πρωτότυπη πρόταση είναι: “At present only 50% of consumers in the EU shop online.” Η λέξη “shop” χρησιμοποιείται ως ρήμα αλλά μεταφράζεται ως ουσιαστικό. Αυτό δεν αναιρεί τη διάκριση μεταξύ πιστότητας και γραμματικότητας. Απλώς αυτή θα γινόταν πιο κατανοητή, αν εξετάζαμε την ίδια πρόταση μεταφρασμένη από άνθρωπο αντικαθιστώντας το ρήμα που επέλεξε ο μεταφραστής με το ουσιαστικό “κατάστημα”. Μία απόδοση μεταφραστή είναι: “Αυτή τη στιγμή μόνο 50% των καταναλωτών της Ε.Ε. **κάνουν αγορές** στο διαδίκτυο.” → “Αυτή τη στιγμή μόνο 50% των καταναλωτών της Ε.Ε. **κατάστημα** στο διαδίκτυο.”, όπου και φαίνεται πιο ευδιάκριτα το λανθασμένο μέρος του λόγου.

Φαινόμενο:	Λάθος σε αριθμό
Παράδειγμα:	“IBS is linked to many factors such as **gut** flora, mucosal permeability, intolerance to foods as well as genetic factors.” → “Η μεταμόσχευση χλωρίδας **εντέρων** μπορεί ριζικά να αλλάξει τη χλωρίδα. Αυτή η μέθοδος θα χρησιμοποιηθεί σύντομα στο κέντρο μας.”

Η επισημειωμένη λέξη είναι σε πληθυντικό αριθμό αντί για ενικό. Παρά το γεγονός ότι στην MQM αυτό το φαινόμενο υπάγεται στην κατηγορία Λανθασμένη Μετάφραση [Burchardt, A. κ.ά., 2014] και το ότι υπάρχουν περιπτώσεις στις οποίες το λάθος γίνεται αντιληπτό μόνο με αντιπαράθεση πρωτότυπου-μεταφράσματος, το εντάσσουμε στη μορφολογία, ώστε η κατηγορία Λεξιλογική Επιλογή να προσφέρει πληροφορίες αποκλειστικά για το λεξιλόγιο, πληροφορίες που συνδέονται με το γλωσσικό μοντέλο του συστήματος. Συνεπώς, στην κατηγορία Μορφολογία (φαινόμενο “Λάθος σε αριθμό”) του Appraise μπορεί κανείς να συναντήσει τόσο θέματα πιστότητας όσο και γραμματικότητας.

Φαινόμενο:	Λάθος συμφωνίας υποκειμένου-κατηγορήματος, γένους και πτώσης
Παράδειγμα:	5. μου κοιλιά είναι **πρησμένο** και νιώθω άβολα.

Η συμφωνία είναι μορφοσυντακτικό φαινόμενο και δύσκολα κατατάσσεται σε μία από τις δύο κατηγορίες, μορφολογία ή σύνταξη. Στην τυπολογία του Appraise δεν έχουμε μια μορφοσυντακτική υπερκατηγορία. Επιλέγουμε τη μορφολογία, γιατί με αυτόν τον τρόπο η σύνταξη περιορίζεται αποκλειστικά στη Σειρά Λέξεων, όπως ορίζεται στην παρένθεση (Syntax [ordering]), αλλά και ακολουθούμε την ταξινόμηση της πιο λεπτομερούς ιεραρχείας της MQM, όπου η συμφωνία είναι παιδί της μορφολογίας. Στο συγκεκριμένο παράδειγμα έχουμε ασυμφωνία γένους μεταξύ ουσιαστικού και επιθέτου. Στα λάθη συμφωνίας είναι εμφανές ποια λέξη φέρει γραμματικό λάθος, οπότε και δεν χρειάζεται περαιτέρω διευκρίνιση στις οδηγίες.

Φαινόμενο:	**Τελικό νι**

Παράδειγμα:	“Με μια σειρά δοκιμών μπορούμε να καθορίσουμε την αιτία της σε έναν σημαντικό βαθμό και με τη συγκεκριμένη διατροφή επεξεργασίας μπορούμε να ελαχιστοποιήσουμε την ταλαιπωρία και να βελτιώσουμε **τη** ποιότητα ζωής.”
Έλλειψη τελικού νι. Αφορά μόνο την κατεύθυνση αγγλικά-ελληνικά. Πολύ περιορισμένο φαινόμενο (1 εμφάνιση μόνο στα κείμενα δοκιμών), το οποίο είναι και ήσσονος σημασίας για ένα μεταφραστικό σύστημα.	
Φαινόμενο:	**Τονισμός**
Παράδειγμα:	“Δηλαδή εάν είστε **πιό** μεγαλύτεροι από 50 (ή 40, σε περίπτωση οικογενειακής ιστορίας του καρκίνος του παχέος εντέρου) ή εάν εκτίθεστε στους παράγοντες κινδύνου.”
Λανθασμένος τονισμός λέξης. Αφορά μόνο την κατεύθυνση αγγλικά-ελληνικά. Αρκετά περιορισμένο φαινόμενο (οι περισσότερες εμφανίσεις του είναι εικονικές, οφείλονται δηλαδή σε λάθος που δημιουργείται από διόρθωση, όπως συμβαίνει με την ανάγκη προσθήκης αντωνυμίας με αποτέλεσμα τον διπλό τονισμό της λέξης), το οποίο είναι και ήσσονος σημασίας για ένα μεταφραστικό σύστημα.	
Misspelling	**Λανθασμένη Ορθογραφία → Ελλείπουσα Λειτουργική Λέξη**
Φαινόμενο:	Πέρα από τα φαινόμενα του τελικού νι και του τονισμού, τα οποία αφορούν μόνο τα ελληνικά ως γλώσσα-στόχος, είναι ήσσονος σημασίας και εντάχθηκαν στη μορφολογία, δεν παρατηρήθηκαν φαινόμενα λανθασμένης ορθογραφίας στις δοκιμές μας. Επίσης το Moses τυπώνει μόνο πεζούς χαρακτήρες (η σωστή χρήση πεζών-κεφαλαίων θα ήταν το βασικό λάθος που θα μας ενδιέφερε σε αυτή την κατηγορία). Γι' αυτούς τους λόγους επιλέγουμε να αξιοποιήσουμε το υπάρχον πεδίο του εργαλείου για ένα φαινόμενο που δεν προβλέπεται σε επίπεδο κατηγορίας και το οποίο κρίνουμε σκόπιμο να μελετήσουμε: την ελλείπουσα λειτουργική λέξη[48]. Συνεπώς η ετικέτα Misspelling παραπέμπει σε Ελλείπουσα Λειτουργική Λέξη και στο εξής θα αναφέρεται ως τέτοια. Πρόκειται για φαινόμενο γραμματικότητας.
Παράδειγμα:	**“**∅ Κολονοσκόπηση γίνεται για να αποκλείσει άλλες συνθήκες, αλλά μόνο στο έδαφος των ενδείξεων.”
Ελλείπον άρθρο.	
Punctuation	**Στίξη**
Φαινόμενο:	Πλεονάζον, ελλείπον ή λανθασμένο σημείο στίξης που μεταβάλλει τη

[48] Όσο αφορά τις λειτουργικές λέξεις, υπάρχουν τρία είδη λαθών: αντικατάσταση, προσθήκη και παράλειψη. Από αυτά τα τρία το είδος που υπάρχει η δυνατότητα και χρειάζεται περισσότερο να διαχειριστούμε είναι η παράλειψη. Η δυνατότητα υπάρχει, γιατί έχει αποδειχτεί ότι βελτιώσεις στο γλωσσικό μοντέλο επιφέρουν μείωση των ελλειπουσών λέξεων, τόσο των λειτουργικών όσο και των λέξεων περιεχομένου [Nădejde, M., 2011]. Και η χρησιμότητα είναι μεγαλύτερη, γιατί με την παράλειψη λειτουργικής λέξης υπάρχει ενδεχόμενη απώλεια πληροφορίας (π.χ. σε περίπτωση παράλειψης κτητικής αντωνυμίας), ενώ η προσθήκη λειτουργικής λέξης έχει συνήθως μόνο υφολογικό αντίκτυπο στο κείμενο.

	σημασία της πρότασης ή έχει σοβαρό υφολογικό αντίκτυπο
Παράδειγμα:	“"People with cystic fibrosis experience significant and progressive loss of lung function during their lives.” → “"οι άνθρωποι με κυστική ίνωση, η σημαντική και τη σταδιακή απώλεια της πνευμονικής λειτουργίας κατά τη διάρκεια της ζωής τους.”
Σε αυτό το παράδειγμα μπορεί κανείς να παρατηρήσει την πολυπλοκότητα της εργασίας επισημείωσης, διότι ναι μεν η στίξη είναι φαινόμενο γραμματικότητας, στην προκειμένη περίπτωση όμως συνδέεται με το φαινόμενο ελλείπουσας λέξης, το οποίο είναι θέμα πιστότητας. Σε αυτή την πρόταση λείπει το ρήμα (Ελλείπουσα Λέξη → Πιστότητα), οπότε υπάρχει λάθος στίξης (κόμμα μεταξύ υποκειμένου και ρήματος). Η πρόταση διορθωμένη είναι “"οι άνθρωποι με κυστική ίνωση παρουσιάζουν σημαντική και σταδιακή απώλεια της πνευμονικής λειτουργίας κατά τη διάρκεια της ζωής τους.”. Εάν δεν έλειπε το ρήμα αλλά αναφορική αντωνυμία, το κόμμα ενδεχομένως δεν θα αποτελούσε λάθος, αλλά για να το αντιληφθούμε αυτό χρειαζόμαστε πρόσβαση στο πρωτότυπο.	
Other (Idiom, etc.)	**Άλλο [Ιδιωματική Έκφραση, κ.λπ.] → Αμετάφραστη Λέξη**
Φαινόμενο:	Καθώς δεν παρατηρήθηκαν στις δοκιμές μας άλλα φαινόμενα που να μην μπορούν να ενταχτούν σε κάποια από τις υπάρχουσες κατηγορίες και η κατηγορία Αμετάφραστη Λέξη μας είναι απαραίτητη, για να γνωρίζουμε την ποσότητα των λέξεων που δεν μεταφράζει το σύστημα[49], χρησιμοποιούμε αυτή την κατηγορία ως Αμετάφραστη Λέξη, λέξη δηλαδή η οποία δεν μεταφράστηκε ενώ θα έπρεπε, προκειμένου να πληρούνται τα κριτήρια του επιθυμητού μεταφράσματος. Στο εξής θα αναφέρεται ως Αμετάφραστη Λέξη.
Παράδειγμα:	“1. σίγουρα έχω **IBS**;”
Αρκτικόλεξο που δεν μεταφράζεται, ενώ θα έπρεπε, γιατί δεν είναι κατανοητό στον μέσο αναγνώστη στον οποίο απευθύνεται το κείμενο.	

Πίνακας 6: Τυπολογία λαθών Appraise

Ενώ στην τυπολογία του DQF η κάθε κατηγορία περιλάμβανει αρκετά διαφορετικά φαινόμενα, εδώ παρατηρείται αυτό το φαινόμενο στην κατηγορία της μορφολογίας. Ο ακριβής καθορισμός των φαινομένων που υπάγονται σε κάθε κατηγορία είναι κάτι που αφορά τον επισημειωτή και τον βοηθά στο έργο του, ωστόσο από άποψη αξιοποίησης των αποτελεσμάτων για βελτίωση συστήματος, οι 10 υπάρχουσες κατηγορίες αρκούν.

Επίπεδα σοβαρότητας λάθους Appraise

Δευτερεύοντα λάθη (minor): Ό,τι καλύπτεται από το επίπεδο δευτερευόντων λαθών του DQF, ήτοι λάθη που δεν έχουν ως αποτέλεσμα απώλεια νοήματος, δεν προκαλούν σύγχυση ή

49 Οι αμετάφραστες λέξεις είναι λέξεις που δεν συναντήθηκαν κατά την εκπαίδευση του συστήματος [Silva, J., κ.ά., 2012] και ο αριθμός τους αποτελεί χρήσιμη ένδειξη για την ποιότητά του συστήματος.

παρερμηνεία στον χρήστη αλλά γίνονται αντιληπτά, επηρεάζουν αρνητικά την ποιότητα του ύφους, τη γραμματικότητα ή την ευκρίνεια, ή λάθη που καθιστούν το περιεχόμενο λιγότερο ελκυστικό.

Σοβαρά λάθη (severe): Ό,τι καλύπτεται από το επίπεδο σοβαρών λαθών (λάθη που ενδέχεται να προκαλέσουν σύγχυση ή παρερμηνεία στον χρήστη, ή να παρεμποδίσουν την ομαλή λειτουργία του προϊόντος ή της υπηρεσίας λόγω σημαντικής αλλαγής στη σημασία ή του ότι τα λάθη εμφανίζονται σε ορατό ή σημαντικό μέρος του περιεχομένου) και από το επίπεδο κρίσιμων λαθών του DQF (λάθη που μπορεί να επιφέρουν επιπλοκές που σχετίζονται με την υγεία ή την ασφάλεια ή με νομικά/οικονομικά θέματα, ή που παραβιάζουν οδηγίες γεωπολιτικής χρήσης, βλάπτουν τη φήμη της εταιρείας, προκαλούν βλάβη της εφαρμογής ή αρνητικές μεταβολές/παρερμηνείες σχετικά με τη λειτουργία ενός προϊόντος ή υπηρεσίας, ή μεταβολές/παρερμηνείες που θα μπορούσαν να θεωρηθούν προσβλητικές), καθώς και λάθη που, ενώ δεν παρεμποδίζουν σε τόσο σημαντικό βαθμό την κατανόηση, προσκρούουν σε τέτοιο βαθμό στο γλωσσικό αισθητήριο του αναγνώστη που μειώνουν τις πιθανότητες να επαναχρησιμοποιήσει το συγκεκριμένο σύστημα μηχανικής μετάφρασης.

4.2.2.2 Σύγκριση DQF-Appraise

Και οι δύο αυτές τυπολογίες έχουν επίπεδη και όχι ιεραρχική δομή. Υπάρχουν 3 κατηγορίες DQF (Πιστότητα, Γλώσσα και Ύφος) και 9 κατηγορίες Appraise[50], οι οποίες συνδέονται μεταξύ τους με τον ακόλουθο τρόπο.

- Πιστότητα
  - Αμετάφραστη Λέξη
  - Προσθήκη Λέξης
  - Λεξιλογική Επιλογή
  - Ορολογία
  - Ελλείπουσες Λέξεις
  - Ελλείπουσα Λειτουργική Λέξη
- Γλώσσα
  - Μορφολογία
  - Σύνταξη
  - Στίξη
- Ύφος
  - Σύνταξη
  - Λεξιλογική επιλογή

[50] Δεν λαμβάνεται υπόψη η ετικέτα “Υπερβολικός αριθμός λαθών”, γιατί δεν είναι κατηγορία επισημείωσης σε επίπεδο λέξης αλλά χαρακτηρισμός που αποδίδεται σε ολόκληρη την πρόταση, προκειμένου να μην αξιολογηθεί.

Ακολουθεί πιο αναλυτική παρουσίαση του ανωτέρω σχήματος με παραδείγματα φαινομένων. Αναφέρονται μόνο οι κατηγορίες και όχι τα φαινόμενα.

DQF **Appraise**

Πιστότητα **Αμετάφραστη Λέξη**

"1. Do I definitely have **IBS**?" → "1. σίγουρα έχω **IBS**;"

Προσθήκη Λέξης

"10 Frequently Asked Questions" → "10 ** συχνές ερωτήσεις"

Λεξιλογική Επιλογή

"To provide a diagnosis for IBS, the doctor will first order specific **tests** to rule out conditions with similar symptoms to those of the irritable bowel syndrome and to check whether you meet the Rome III criteria." → "Να παρέχουν μια διάγνωση για το σύνδρομο ευερέθιστου Εντέρου, ο γιατρός θα διατάξει πρώτα συγκεκριμένες **δοκιμές**, να αποκλείσει όρους με παρόμοια συμπτώματα με αυτά του συνδρόμου του ευερέθιστου εντέρου και να ελέγξει κατά πόσον πληροίτε τα κριτήρια της Ρώμης III."

Ορολογία

"Stress causes the production of specific cytokines in the central nervous system acting on the colon **wall**." → "Στρες προκαλεί την παραγωγή κυτοκινών συγκεκριμένα στο κεντρικό νευρικό σύστημα ενεργεί στον **τοίχο** του παχέος εντέρου."

Ελλείπουσες Λέξεις

"Everyone has the right to life, liberty and security **of person**." → "κάθε πρόσωπο έχει δικαίωμα στη ζωή, την ελευθερία και ασφάλεια."

Ελλείπουσα Λειτουργική Λέξη

"∅ Κολονοσκόπηση γίνεται για να αποκλείσει άλλες συνθήκες, αλλά μόνο στο έδαφος των ενδείξεων."

Γλώσσα **Μορφολογία**

"Ειδικές εξετάσεις θα **επιτρέψει** σε μας για τον εντοπισμό μισαλλοδοξίας στα τρόφιμα, καθώς και την παροχή συμβουλών σχετικά με ειδική διατροφή σχέδια που θα ανακουφίσει τους ασθενείς συμπτώματα."

Σύνταξη

"5. **μου** κοιλιά είναι πρησμένο και νιώθω άβολα."

Στίξη

""People with cystic fibrosis experience significant and progressive loss of lung function during their lives." → ""οι άνθρωποι με κυστική ίνωση**,** η σημαντική και τη σταδιακή απώλεια της πνευμονικής λειτουργίας κατά τη διάρκεια της ζωής τους.”

Ύφος **Σύνταξη**

“Vertex Pharmaceuticals Incorporated (Nasdaq: VRTX) today announced the presentation of new KALYDECO™ (ivacaftor) data at the 37th European Cystic Fibrosis Society (ECFS) Conference, June 11-14, 2014, in Gothenburg, Sweden.” → “κορυφή pharmaceuticals ενσωματωθεί (nasdaq: vrtx) ανακοίνωσε την παρουσίαση των νέων kalydeco ™ (ivacaftor) τα **δεδομένα** στην 37η ευρωπαϊκής κυστική ίνωση κοινωνία (τεκ), 11-14 ιουνίου 2014, στο γκέτεμποργκ της σουηδίας.”

Λεξιλογική επιλογή

“Colonoscopy is performed to rule out other conditions but only on the **ground** of indications.” → “Κολονοσκόπηση γίνεται για να αποκλείσει άλλες συνθήκες, αλλά μόνο στο **έδαφος** των ενδείξεων.”

Από τις κατηγορίες του DQF η Πιστότητα είναι αυτή που αντιστοιχεί στις περισσότερες κατηγορίες του Appraise. Η Γλώσσα και το Ύφος έχουν αντιστοιχία ένα προς δύο. Και από τις κατηγορίες του Appraise υπάρχουν δύο οι οποίες αντιστοιχούν σε δύο διαφορετικές κατηγορίες του DQF: η Λεξιλογική Επιλογή και η Σύνταξη. Όσο αφορά τη Λεξιλογική Επιλογή, αυτό συμβαίνει γιατί στο Appraise δεν υπάρχει κατηγορία Κατά Λέξη Μετάφραση (φαινόμενο που στο DQF υπάγεται στο Ύφος), οπότε αυτό το φαινόμενο υπάγεται στην κατηγορία Λεξιλογική Επιλογή. Το φαινόμενο της λανθασμένης μετάφρασης (στο οποίο συμπεριλαμβάνεται το φαινόμενο της κατά λέξη μετάφρασης) στο DQF εντάσσεται στην πιστότητα. Πρόκειται για έναν διαχωρισμό (κατά λέξη μετάφραση-λανθασμένη μετάφραση) ο οποίος δεν μας είναι απαραίτητος. Όσο αφορά τη Σύνταξη, η Σύνταξη του Appraise που αντιστοιχεί στην κατηγορία του Ύφους του DQF είναι, όπως αναφέρεται στις οδηγίες του DQF, κάποια “περίεργη σύνταξη”, φαινόμενο το οποίο καλύπτεται και από τη Σύνταξη του Appraise που αντιστοιχεί στη Γλώσσα του DQF. Τα βασικά πλεονεκτήματα που προκύπτουν υπέρ της τυπολογίας του Appraise για την αξιολόγησή μας είναι ότι μπορούμε να έχουμε ξεχωριστά αποτελέσματα για τις αμετάφραστες λέξεις, για τις ελλείπουσες λέξεις και για τις ελλείπουσες λειτουργικές λέξεις, από τις οποίες μπορούμε να πάρουμε χρήσιμη πληροφορία για τη βελτίωση του συστήματος. Όσο αφορά τις αμετάφραστες λέξεις, αυτές σχετίζονται, όπως προαναφέρθηκε, με το σώμα εκπαίδευσης του συστήματος. Όσο αφορά τις ελλείπουσες λειτουργικές λέξεις, στα στατιστικά συστήματα συνήθως η αντιμετώπιση των λειτουργικών λέξεων και των λέξεων περιεχομένου είναι η ίδια, ωστόσο υπάρχει η δυνατότητα διαφορετικής αντιμετώπισής τους [Cui, L. κ.ά., 2011]. Πέρα από αυτή τη δυνατότητα, θεωρούμε χρήσιμη τη διάκριση αυτών των δύο ειδών λέξεων στην ταξινόμηση λαθών, γιατί με αυτήν μπορεί να γίνει και μια περαιτέρω διάκριση μεταξύ επιπέδων σοβαρότητας λάθους. Εκτός από το επίπεδο σοβαρότητας που επισημειώνουμε σε κάθε λέξη, η κάθε κατηγορία δεν έχει την ίδια βαρύτητα με τις άλλες στο αποτέλεσμα. Τα λάθη πιστότητας είναι σοβαρότερα από τα λάθη γραμματικότητας, γιατί δεν μπορούν να γίνουν αντιληπτά από έναν μονόγλωσσο αναγνώστη του μεταφράσματος, συνεπώς μπορεί να αποτύχει ο επικοινωνιακός στόχος του κειμένου, και γιατί για τη διόρθωσή τους απαιτείται μεταφραστική διαδικασία, η οποία με τη σειρά της απαιτεί περισσότερη έρευνα από ό,τι η γλωσσική διόρθωση του κειμένου-στόχος. Τέλος, μας είναι χρήσιμο η στίξη να διακρίνεται από τη μορφολογία, ακριβώς επειδή με αυτόν τον τρόπο διαχωρίζουμε κατηγορίες διαφορετικής σοβαρότητας.

4.2.2.3 Σύγκριση MQM-Appraise

Η σύγκριση αυτή γίνεται μεταξύ μιας ιεραρχικής με μια επίπεδη δομή. Στα αριστερά βρίσκονται οι κατηγορίες του Appraise (και όχι τα επιμέρους φαινόμενα) και στα δεξιά οι υποκατηγορίες του MQM στις οποίες αυτές αντιστοιχούν. Μαζί με τις υποκατηγορίες MQM αναφέρονται και οι κατηγορίες στις οποίες ανήκουν αυτές με τη σειρά τους. Συνολικά είναι 9 κατηγορίες Appraise (και πάλι χωρίς να συμπεριλαμβάνεται η κατηγορία "Υπερβολικός αριθμός λαθών") και 13 υποκατηγορίες MQM.

Ελλείπουσες λέξεις — Παράλειψη < Πιστότητα

Ορολογία — Ορολογία < Λανθασμένη Μετάφραση < Πιστότητα

Λεξ. Επιλογή
- Καταχρηστική Κατά Λέξη Μετάφραση < Λανθασμένη Μετάφραση < Πιστότητα
- Ψευδόφιλη Λέξη < Λανθασμένη μετάφραση < Πιστότητα
- Δεν θα έπρεπε να έχει μεταφραστεί < Λανθασμένη Μετάφραση < Πιστότητα
- Λειτουργικές λέξεις < Γραμματική < Θέμα Μηχανικό ή Μορφής < Γραμματικότητα

Σύνταξη — Σειρά Λέξεων < Γραμματική < Θέμα Μηχανικό ή Μορφής < Γραμματικότητα

Προσθήκη Λέξης
- Προσθήκη < Πιστότητα
- Λειτουργ. Λέξεις < Γραμματική < Θέμα Μηχ. ή Μορφής < Γραμματικότητα

Γραμματικότητα

Μέρος του Λόγου < Μορφολογία < Γραμ. < Θέμα Μηχ. ή Μορφής <

Μορφολογία Συμφωνία < Μορφολογία < Γραμ. < Θέμα Μηχ. ή Μορφής < Γραμματικότητα

Χρόνος/Έγκλιση/Όψη < Μορφ. < Γραμ. < Θέμα Μηχ. ή Μορφής < Γραμματικότητα

Ελλείπουσα Λειτουργική Λέξη — Λειτουργ. Λέξεις < Γραμ. < Θέμα Μηχ. ή Μορφής < Γραμματικότητα

Στίξη — Στίξη < Τυπογραφικό < Θέμα Μηχ. ή Μορφής < Γραμματικότητα

Αμετάφραστη Λέξη — Αμετάφραστο < Πιστότητα

Ακολουθεί πιο αναλυτική παρουσίαση του ανωτέρω σχήματος με παραδείγματα φαινομένων. Αναφέρονται μόνο οι κατηγορίες του Appraise και όχι τα φαινόμενα.

MQM **Appraise**

Πιστότητα

Λανθασμένη Μετάφραση

Ορολογία **Ορολογία**

“4. Do I have a food **intolerance**?” → “4.Do έχω μια **αδιαλλαξία** τροφίμων;”

Καταχρηστική Κατά Λέξη Μετάφραση **Λεξιλογική Επιλογή**

“Colonoscopy is performed to rule out other conditions but only **on the ground of** indications.” → “Η κολονοσκόπηση εκτελείται για να αποκλείσει άλλους όρους αλλά μόνο στο **έδαφος** των ενδείξεων.”

Ψευδόφιλη Λέξη **Λεξιλογική Επιλογή**

Κείμενο που δεν θα έπρεπε να έχει μεταφραστεί **Λεξιλογική Επιλογή**

“**Vertex** Announces Presentation of New KALYDECO™ (ivacaftor) Data at European Cystic Fibrosis Society Conference” → “**κορυφή** ανακοινώνει την παρουσίαση των νέων kalydeco ™ (ivacaftor) δεδομένων σε ευρωπαϊκή κοινωνία της διάσκεψης κυστική ίνωση”

Παράλειψη **Ελλείπουσες Λέξεις**

“"People with cystic fibrosis **experience** significant and progressive loss of lung function during their lives.” → “"οι άνθρωποι με κυστική ίνωση, η σημαντική και τη σταδιακή απώλεια της πνευμονικής λειτουργίας κατά τη διάρκεια της ζωής τους.”

Αμετάφραστο **Αμετάφραστη Λέξη**

“1. Do I definitely have **IBS**?” → “1. σίγουρα έχω **IBS**;”

Προσθήκη **Προσθήκη Λέξης**

“Αναίμακτο εργαστηριακό έλεγχο με ηλεκτροκαρδιογράφημα (ΗΚΓ), υπερηχοκαρδιογράφημα, Holter 24ωρης καταγραφής ρυθμού, δοκιμασία κοπώσεως, δοκιμασία ανακλίσεως (tilt-test).” → “**And as** bloodless **coastal** laboratory testing with electrocardiogram (ECG), a yperichokardiografi, a 24-hour Holter recording rate regulation, stress test, test anaklisews (tilt-test).”

Γραμματικότητα

Θέμα Μηχανικό ή Μορφής

Τυπογραφικό

Στίξη **Στίξη**

“Τώρα γνωρίζουμε ότι υπάρχουν πάρα πολλά μονοπάτια, μεταξύ του εντέρου και του εγκεφάλου...”

Γραμματική

Μορφολογία

Μέρος του Λόγου **Μορφολογία**

“Αυτή τη στιγμή μόνο το 50% των καταναλωτών στο σε απευθείας σύνδεση **κατάστημα** της Ε.Ε..”

Συμφωνία **Μορφολογία**

“5. μου κοιλιά είναι **πρησμένο** και νιώθω άβολα.”

Χρόνος/Όψη/Έγκλιση **Μορφολογία**

“Πολλοί IBS ασθενείς **αναπτύξουν** δυσανεξία σε ορισμένες τροφές.”

Σειρά Λέξεων **Σύνταξη**

“5. **μου** κοιλιά είναι πρησμένο και νιώθω άβολα.”

Λειτουργικές Λέξεις **Ελλείπουσα Λειτουργική Λέξη**

“e Κολονοσκόπηση γίνεται για να αποκλείσει άλλες συνθήκες, αλλά μόνο στο έδαφος των ενδείξεων.”

Προσθήκη Λέξης

“Μια φουσκωμένη κοιλιά μπορεί να προκαλέσει **τη** σημαντική ταλαιπωρία στον ασθενή.”

Λεξιλογική Επιλογή

“**Μια** φουσκωμένη κοιλιά μπορεί να προκαλέσει τη σημαντική ταλαιπωρία στον ασθενή.”

Ακατανόητο

Αν και ο “Υπερβολικός αριθμός λαθών” δεν είναι κατ' εξοχήν κατηγορία, θα μπορούσε να θεωρηθεί αντίστοιχο της κατηγορίας “Ακατανόητο”.

Όπως ήδη αναφέρθηκε, στην MQM ο επισημειωτής επιλέγει το προβληματικό απόσπασμα, το οποίο μπορεί να υπερβαίνει τη μία λέξη[51], ενώ στο Appraise η επισημείωση γίνεται σε επίπεδο λέξης. Γι' αυτόν τον λόγο στα παραδείγματα είναι επισημειωμένη μόνο μία λέξη.

Στην αντιστοιχία των κατηγοριών παρατηρούμε περισσότερες σχέσεις ένα προς ένα από ό,τι στη σύγκριση DQF-MQM. Η Μορφολογία του Appraise χωρίζεται σε 3 υποκατηγορίες στην MQM. Όπως είδαμε στο Κεφάλαιο 3.2.2.3.1, η κατηγορία Μορφολογία του MQM υπάγεται στην ευρύτερη κατηγορία της Γραμματικότητας, η οποία αφορά το γλωσσικό μοντέλο του συστήματος. Από άποψη μετέπειτα βελτίωσης του συστήματος δεν συνεισφέρει ο διαχωρισμός μεταξύ Συμφωνίας, Μέρους του Λόγου και Χρόνου/Έγκλισης/Όψης. Χρησιμεύει στον επισημειωτή σε κάποιες περιπτώσεις, για να αποκλείσει άλλα φαινόμενα και να είναι πιο βέβαιος για την επιλογή του. Ως προς την κατηγορία Λεξιλογική Επιλογή του Appraise, αυτή χωρίζεται σε 4 υποκατηγορίες στο MQM. Επίσης δεν είναι αξιοποιήσιμη η διάκριση μεταξύ Καταχρηστικής Κατά Λέξη Μετάφρασης, Ψευδόφιλης Λέξης και Λέξης που δεν θα έπρεπε να έχει μεταφραστεί. Η μοναδική κατηγορία της MQM που εμφανίζεται σε περισσότερες από 1 κατηγορίες του Appraise είναι οι λειτουργικές λέξεις, γιατί στην MQM η κατηγορία Λειτουργικές Λέξεις περιλαμβάνει και την προσθήκη και την παράλειψη και τη λανθασμένη επιλογή λειτουργικής λέξης και εμπίπτει στην ευρύτερη κατηγορία της Γραμματικότητας, ενώ στην ταξινόμησή μας στο Appraise η λανθασμένη επιλογή εμπίπτει στη Λεξιλογική Επιλογή (Πιστότητα), η προσθήκη λειτουργικής λέξης εμπίπτει στην Προσθήκη Λέξης (Πιστότητα) και η παράλειψη λειτουργικής λέξης αποτελεί ξεχωριστή κατηγορία.

4.2.2.4 Σύγκριση DQF-MQM

Και αυτή η σύγκριση γίνεται μεταξύ μιας επίπεδης και μιας ιεραρχικής δομής. Μάλιστα η επίπεδη είναι η τυπολογία με τις λιγότερες κατηγορίες (3) και η ιεραρχική αυτή με τις περισσότερες (13 υποκατηγορίες). Η σχέση μεταξύ τους είναι η ακόλουθη.

[51] Σύμφωνα με τις οδηγίες του QTLaunchpad, ο επισημειωτής οφείλει να επισημειώνει όσο το δυνατόν μικρότερο απόσπασμα [Burchardt, A. et al, 2014].

- Πιστότητα
  - Ορολογία < Λανθασμένη Μετάφραση < Πιστότητα
  - Ψευδόφιλη λέξη < Λανθασμένη Μετάφραση < Πιστότητα
  - Δεν θα έπρεπε να μεταφραστεί < Λανθασμένη Μετάφραση < Πιστότητα
  - Παράλειψη < Πιστότητα
  - Αμετάφραστο < Λανθασμένη Μετάφραση < Πιστότητα
  - Προσθήκη < Λανθασμένη Μετάφραση < Πιστότητα
  - Λειτουργ. Λέξεις < Μορφολογία < Γραμ. < Θέμα Μηχ. < Γραμματικότητα

- Γλώσσα
  - Στίξη < Τυπογραφικό < Θέμα Μηχ. < Γραμματικότητα
  - Μέρος του Λόγου < Μορφολογία < Γραμ. < Θέμα Μηχ. < Γραμματικότητα
  - Συμφωνία < Μορφολογία < Γραμ. < Θέμα Μηχ. < Γραμματικότητα
  - Χρόνος/Όψη/Έγκλιση < Μορφολογία < Γραμ. < Θέμα Μηχ. < Γραμματικότητα
  - Σειρά Λέξεων < Γραμ. < Θέμα Μηχ. < Γραμματικότητα

- Ύφος
  - Καταχρηστική Κατά Λέξη Μετάφραση

Ακολουθεί πιο αναλυτική παρουσίαση του ανωτέρω σχήματος με παραδείγματα φαινομένων. Αναφέρονται μόνο οι κατηγορίες του DQF και όχι τα φαινόμενα.

MQM	**DQF**
Πιστότητα	
Λανθασμένη Μετάφραση	
Ορολογία	**Πιστότητα**

“4. Do I have a food **intolerance**?” → “4.Do έχω μια **αδιαλλαξία** τροφίμων;”

Καταχρηστική Κατά Λέξη Μετάφραση	**Ύφος**

“Colonoscopy is performed to rule out other conditions but only **on the ground of** indications.” → “Η κολονοσκόπηση εκτελείται για να αποκλείσει άλλους όρους αλλά μόνο στο **έδαφος** των ενδείξεων.”

Ψευδόφιλη λέξη	**Πιστότητα**
Κείμενο που δεν θα έπρεπε να έχει μεταφραστεί	**Πιστότητα**

“**Vertex** Announces Presentation of New KALYDECO™ (ivacaftor) Data at European Cystic Fibrosis Society Conference” → “**κορυφή** ανακοινώνει την παρουσίαση των νέων kalydeco ™ (ivacaftor) δεδομένων σε ευρωπαϊκή κοινωνία της διάσκεψης κυστική ίνωση”

Παράλειψη **Πιστότητα**

“"People with cystic fibrosis **experience** significant and progressive loss of lung function during their lives.” → “"οι άνθρωποι με κυστική ίνωση, η σημαντική και τη σταδιακή απώλεια της πνευμονικής λειτουργίας κατά τη διάρκεια της ζωής τους.”

Αμετάφραστο **Πιστότητα**

“1. Do I definitely have **IBS**?” → “1. σίγουρα έχω **IBS**;”

Προσθήκη **Πιστότητα**

“Αναίμακτο εργαστηριακό έλεγχο με ηλεκτροκαρδιογράφημα (ΗΚΓ), υπερηχοκαρδιογράφημα, Holter 24ωρης καταγραφής ρυθμού, δοκιμασία κοπώσεως, δοκιμασία ανακλίσεως (tilt-test).” → “**And as** bloodless **coastal** laboratory testing with electrocardiogram (ECG), a yperichokardiografi, a 24-hour Holter recording rate regulation, stress test, test anaklisews (tilt-test).”

Γραμματικότητα

Θέμα Μηχανικό ή Μορφής

Τυπογραφικό

Στίξη **Γλώσσα**

“Τώρα γνωρίζουμε ότι υπάρχουν πάρα πολλά μονοπάτια, μεταξύ του εντέρου και του εγκεφάλου...”

Γραμματική

Μορφολογία

Μέρος του Λόγου **Γλώσσα**

“Αυτή τη στιγμή μόνο το 50% των καταναλωτών στο σε απευθείας σύνδεση **κατάστημα** της Ε.Ε..”

Συμφωνία **Γλώσσα**

“5. μου κοιλιά είναι **πρησμένο** και νιώθω άβολα.”

Χρόνος/Όψη/Έγκλιση **Γλώσσα**

“Πολλοί IBS ασθενείς **αναπτύξουν** δυσανεξία σε ορισμένες τροφές.”

Σειρά Λέξεων **Γλώσσα**

“5. **μου** κοιλιά είναι πρησμένο και νιώθω άβολα.”

Λειτουργικές Λέξεις **Πιστότητα**

“e Κολονοσκόπηση γίνεται για να αποκλείσει άλλες συνθήκες, αλλά μόνο στο έδαφος των ενδείξεων.”

Ακατανόητο

Σε αυτή την περίπτωση υπάρχει σχεδόν πλήρης αντιστοίχιση πιστότητας και γραμματικότητας μεταξύ των δύο τυπολογιών (στο DQF η πιστότητα ονομάζεται “Γλώσσα”), με τη διαφορά ότι στην MQM η πιστότητα και η γραμματικότητα είναι υπερκατηγορίες των οποίων οι υποκατηγορίες δεν υπάρχουν στο DQF. Η κατηγορία Ύφος του DQF περιλαμβάνει την Κατά Λέξη Μετάφραση, η οποία στην MQM υπάγεται στην κατηγορία της πιστότητας, αλλά, όπως είδαμε, είναι μια κατηγορία η οποία δεν είναι απαραίτητη στην ταξινόμησή μας, καθώς και την κατηγορία Περίεργη Σύνταξη, η οποία μπορεί να καλυφθεί από την κατηγορία Γλώσσα (φαινόμενο Γραμματική-Σύνταξη). Το φαινόμενο στο οποίο παρατηρείται ασυμφωνία μεταξύ των δύο τυπολογιών είναι οι λειτουργικές λέξεις, οι οποίες, καθώς στο DQF δεν αποτελούν ξεχωριστή κατηγορία, υπάγονται στα φαινόμενα Λανθασμένη μετάφραση, Παράλειψη και Προσθήκη, άρα στην κατηγορία Πιστότητα, ενώ στο MQM πρόκειται ξεκάθαρα για θέμα Γραμματικότητας.

4.2.2.5 Επιλογή τυπολογίας

Συνοψίζοντας, η τυπολογία του DQF διαθέτει πολύ λίγες κατηγορίες, οι οποίες δεν καλύπτουν τις ανάγκες τις εργασίας μας. Συγκεκριμένα δεν παρέχει κατηγορία για τις αμετάφραστες, τις ελλείπουσες λέξεις και τις ελλείπουσες λειτουργικές λέξεις, και δεν κάνει διάκριση μεταξύ στίξης και μορφολογίας. Εμπίπτουν δηλαδή στην ίδια κατηγορία τα λάθη χρόνου, συμφωνίας, αριθμού, κ.λπ. με τα λανθασμένα κόμματα. Αυτό έχει ως αποτέλεσμα την αλλοίωση των αποτελεσμάτων, καθώς καταλήγουμε με έναν αριθμό μορφολογικών λαθών με πολύ διαφορετικό επίπεδο σοβαρότητας. Τέλος, κάνει διάκριση μεταξύ κατά λέξη μετάφρασης και λανθασμένης μετάφρασης, καθώς και μεταξύ σύνταξης και περίεργης σύνταξης, διακρίσεις οι οποίες δεν μας είναι απαραίτητες. Ένα ακόμη βασικό μειονέκτημα είναι το γεγονός ότι η πληροφορία που επισημειώνεται και που αναφέρεται στα εξαγόμενα αποτελέσματα είναι ο αριθμός λαθών κάθε κατηγορίας σε κάθε πρόταση και όχι σε ποια λέξη μέσα στην πρόταση εντοπίστηκε το κάθε λάθος. Οι δύο εναπομείνασες τυπολογίες έχουν σχεδόν τον ίδιο αριθμό κατηγοριών. Το Appraise, παρόλο που είναι αυτό με τις λιγότερες, καλύπτει τις ανάγκες της εργασίας μας. Η λεπτομερής υποκατηγοριοποίηση της MQM, παρότι ιδιαίτερα ενδιαφέρουσα από θεωρητική άποψη, δεν θα αξιοποιούταν στην αξιολόγησή μας, καθώς θα μπορούσαν να μας καλύψουν οι ενδιάμεσες κατηγορίες της, οι οποίες και συμπίπτουν με αυτές του Appraise. Αυτό σημαίνει ότι η Λεξιλογική Επιλογή του Appraise αντιστοιχεί στη Λανθασμένη Μετάφραση της MQM κι η Μορφολογία του Appraise στη Μορφολογία της MQM. Όσο αφορά της λειτουργικές λέξεις, καλύπτουμε την περίπτωση της έλλειψης στο Appraise. Ως προς τα επίπεδα σοβαρότητας λάθους, θεωρούμε τα δύο επίπεδα του Appraise πιο εύχρηστα από τα τρία της MQM. Όσο πιο πολλά επίπεδα υπάρχουν τόσο πιο δυσδιάκριτα είναι τα όρια μεταξύ τους. Η νοητική διεργασία κατά την επισημείωση πολλαπλασιάζεται εάν ο επισημειωτής πρέπει να κάνει συνεχώς λεπτές διακρίσεις μεταξύ των επιπέδων σοβαρότητας παράλληλα με τις λεπτές διακρίσεις μεταξύ κατηγοριών, κάτι το οποίο θέτει σε κίνδυνο την αξιοπιστία της ίδιας της αξιολόγησης[52]. Επίσης όσο περισσότερα είναι τα

[52] Για ανάλυση της νοητικής προσπάθειας που καταβάλλεται κατά την επισημείωση και τους τρόπους

επίπεδα σοβαρότητας λάθους τόσο μεγαλύτερος είναι και ο βαθμός υποκειμενικότητας της αξιολόγησης. Λαμβάνοντας υπόψη όλους τους προαναφερθέντες παράγοντες, επιλέγουμε τη χρήση του εργαλείου Appraise με τις κατηγορίες αυτού όπως ορίστηκαν στον Πίνακα 6.

4.2.3 Οδηγίες επισημείωσης: Ταξινόμηση Λαθών και Μετεπεξεργασία

Με βάση τις υπάρχουσες οδηγίες επισημείωσης [Przybocki, M. κ.ά., 2011], την επιλεχθείσα τυπολογία και τις ανάγκες της αξιολόγησής μας, και ύστερα από δοκιμές και επακόλουθες βελτιώσεις, οι οδηγίες επισημείωσης που ακολουθήθηκαν κατά την ταξινόμηση λαθών και τη μετεπεξεργασία είναι οι εξής:

- Τα λάθη εντοπίζονται και ταξινομούνται με βάση τη διόρθωσή τους.
- Η κατηγορία Misspelling χρησιμοποιείται ως Ελλείπουσα Λειτουργική Λέξη και η ελλείπουσα λειτουργική λέξη επισημειώνεται στη λέξη που ακολουθεί, εκτός εάν πρόκειται για τέλος πρότασης, οπότε και επισημειώνεται στην ακριβώς προηγούμενη.
- Η κατηγορία Other χρησιμοποιείται ως Αμετάφραστη Λέξη.
- Κριτήριο για την επιλογή μεταξύ Ελλείπουσας Λέξης και Ελλείπουσας Λειτουργικής Λέξης είναι αν το λάθος εντάσσεται στη γενικότερη κατηγορία της Πιστότητας (Ελλείπουσα Λέξη) ή της Γραμματικότητας (Ελλείπουσα Λειτουργική Λέξη), σύμφωνα με τον διαχωρισμό που γίνεται στην MQM αλλά και γενικότερα μεταξύ πιστότητας και γραμματικότητας. Αν επισημειώνεται ελλείπουσα λέξη, δεν επισημειώνεται και ενδεχόμενη ελλείπουσα λειτουργική λέξη (π.χ. σε ελλείπον ουσιαστικό δεν επισημειώνεται και η έλλειψη άρθρου αυτού, όπως στην ελλείπουσα φράση “του παρόντος”). Ακόμη και σε περίπτωση που υπάρχουν περισσότερες από μία ελλείπουσες λειτουργικές λέξεις (π.χ. πρόθεση και άρθρο πριν από ουσιαστικό), επιλέγεται μόνο μία φορά η κατηγορία Ελλείπουσα Λειτουργική Λέξη. Στις Ελλείπουσες Λέξεις δεν μπορεί να επιλεγεί επίπεδο σοβαρότητας λάθους.
- Δεν λαμβάνουμε υπόψη κεφαλαία-πεζά και κενά ανάμεσα στα ελληνικά εισαγωγικά και τη λέξη, γιατί βάσει σχεδιασμού το Moses δεν τυπώνει τα μεν και τυπώνει το κενό στα δε.
- Σε περίπτωση που εντοπιστεί ορθογραφικό ή τυπογραφικό λάθος, εντάσσεται στην κατηγορία Λεξιλογική Επιλογή.
- Η επισημείωση γίνεται έχοντας υπόψη τον απλούστερο τρόπο διόρθωσης, ο οποίος είναι αυτός που περιλαμβάνει τα λιγότερα βήματα, δηλαδή τον ελάχιστο δυνατό αριθμό λαθών και διορθώσεων. Π.χ. στο απόσπασμα “people with CF who have the G551D mutation” → “άτομα με cf οι οποίοι έχουν g551d μετάλλαξη” δεν διορθώνουμε την αναφορική αντωνυμία (δύο ενέργειες διόρθωσης: ή δύο αντικαταστάσεις, “τα οποία”, ή μία απαλοιφή και μία

μείωσης αυτής βλ. Przybocki, M. κ.ά. [2011].

αντικατάσταση, “που”) αλλά το ουσιαστικό “άτομα” σε “ασθενείς” (μία ενέργεια διόρθωσης).

- Όταν για ένα λάθος μπορούμε να επιλέξουμε μεταξύ των κατηγοριών 1) Αμετάφραστη Λέξη, 2) Ορολογία και 3) Λεξιλογική Επιλογή, ελέγχουμε με αυτή τη σειρά προτεραιότητας.
- Αν σε μια λέξη έχουμε λάθος Λεξιλογικής Επιλογής και λάθος Μορφολογίας, επισημειώνουμε και τα δύο.
- Αν σε μια λέξη υπάρχει λάθος Λεξιλογικής Επιλογής το οποίο διορθώνεται με δύο λέξεις αντί για μία, τότε επισημειώνουμε και Λεξιλογική επιλογή και Ελλείπουσες Λέξεις/Ελλείπουσα Λειτουργική Λέξη.
- Τα λάθη Σύνταξης επισημειώνονται στην κεφαλή της φράσης όπου υπάρχει το λάθος, εκτός εάν πρόκειται για μικρά ν-γράμματα (έως 3-γράμματα) όπου είναι ευδιάκριτο ποια λέξη βρίσκεται σε λανθασμένη θέση (π.χ. στην περίπτωση “μου κοιλιά” επισημειώνεται στο “μου”). Κεφαλή της φράσης θεωρείται η ορθή κεφαλή της φράσης, ως έχει αφού υποστεί μετεπεξεργασία. Για παράδειγμα, στην πρόταση “Vertex Announces Presentation of New KALYDECO™ (ivacaftor) Data at <u>European Cystic Fibrosis Society Conference</u>” → “κορυφή ανακοινώνει την παρουσίαση των νέων kalydeco ™ (ivacaftor) δεδομένων σε <u>ευρωπαϊκή κοινωνία της διάσκεψης κυστική ίνωση</u>” οι υπογραμμισμένες λέξεις αποτελούν μια ονοματική φράση, η οποία στη μετάφραση περιέχει λάθη στη σειρά των λέξεων. Η διορθωμένη ονοματική φράση είναι “διάσκεψη της ευρωπαϊκής επιστημονικής κοινότητας κυστικής ίνωσης”, της οποίας η κεφαλή είναι η “διάσκεψη” και το λάθος Σύνταξης επισημειώνεται σε αυτήν.
- Το σημείο στίξης εμφανίζεται μαζί με τη λέξη που προηγείται, οπότε αν πρόκειται για λάθος στίξης επιλέγεται η εν λόγω λέξη και η κατηγορία Στίξη. Αν πρόκειται για ελλείπον σημείο στίξης, επισημειώνεται στην προηγούμενη λέξη. Επισημειώνονται μόνο τα λάθη στίξης που μεταβάλλουν το νόημα της πρότασης.
- Όταν επισημειώνεται λάθος Προσθήκης Λέξης, δεν επισημειώνονται άλλα τυχόντα λάθη στην πλεονάζουσα λέξη. Π.χ. αν υπάρχει πλεονάζον άρθρο, στο οποίο υπάρχει και λανθασμένο γένος, δεν επισημειώνεται και μορφολογικό λάθος.
- Υπάρχουν λέξεις οι οποίες σε κάποιες περιπτώσεις λειτουργούν ως λέξεις περιεχομένου και άλλες ως λειτουργικές, όπως είναι το αγγλικό ρήμα “to be”, το οποίο μπορεί να είναι βοηθητικό ρήμα, συνεπώς λειτουργική λέξη, ή λέξη περιεχομένου.
- Κατά την επισημείωση χρησιμοποιούνται οι μεταφράσεις-αναφορά.

Παρατηρήσεις σχετικά με τη διαδικασία επισημείωσης

Ακολουθούν κάποιες παρατηρήσεις σχετικά με τη διαδικασία επισημείωσης σε Ταξινόμηση Λαθών και Μετεπεξεργασία.

Ιδιάζουσες περιπτώσεις

- Στο νομικό κείμενο L_en_gr_4 εμφανίζεται ο όρος "issue", ο οποίος αποδίδεται λανθασμένα από το Moses ως "θέμα", "έκδοση" ή "ζήτημα", ενώ στα πλαίσια του νομικού λόγου και στο συγκεκριμένο κείμενο έχει την έννοια των απογόνων ("m) In this Deed references to "children" and "**issue**" shall be construed so as to exclude an illegitimate child, children or **issue** (but not an adopted or legitimated person) and any persons claiming descent through such illegitimate child, children or **issue**." → "ιγ) στην παρούσα πράξη οι αναφορές στην "παιδιά" και "**θέμα**" νοούνται, ώστε να αποκλείεται μια παράνομη παιδιά, παιδιά ή την **έκδοση** (αλλά όχι και διαθέτουν δημοκρατική νομιμότητα πρόσωπο) και τα πρόσωπα που ζητούν προκαλούνται από τέτοιες παράνομες παιδιά, παιδιά ή **ζήτημα**."). Πρόκειται για μία περίπτωση στην οποία ο αριθμός είναι άμεσα συνδεδεμένος με την έννοια της λέξης, η σημασιολογική αντιστοιχία δηλαδή βρίσκεται μεταξύ των λέξεων "issue" (ενικός αριθμός) και "απόγονοι" (πληθυντικός αριθμός). Στην προκειμένη περίπτωση επισημειώνεται λάθος ορολογίας και όχι μορφολογίας.

Επιλογή επιπέδου σοβαρότητας λάθους

- Το νομικό κείμενο L_en_gr_4 είναι το εισαγωγικό μέρος καταπιστεύματος, στο οποίο δίνονται οι ορισμοί των όρων που εμφανίζονται στο υπόλοιπο κείμενο. Η ακρίβεια είναι ιδιαίτερα σημαντική σε ένα τέτοιου είδους κείμενο. Συνεπώς, το μορφολογικό λάθος επιλογής αριθμού στην ακόλουθη πρόταση είναι σοβαρό: "m) In this Deed references to "children" and "issue" shall be construed so as to exclude an illegitimate **child**, children or issue (but not an adopted or legitimated person) and any persons claiming descent through such illegitimate **child**, children or issue." → "ιγ) στην παρούσα πράξη οι αναφορές στην "παιδιά" και "θέμα" νοούνται, ώστε να αποκλείεται μια παράνομη **παιδιά**, παιδιά ή την έκδοση (αλλά όχι και διαθέτουν δημοκρατική νομιμότητα πρόσωπο) και τα πρόσωπα που ζητούν προκαλούνται από τέτοιες παράνομες **παιδιά**, παιδιά ή ζήτημα."
- Αμετάφραστοι όροι από τα αγγλικά στα ελληνικά, όπως ο όρος "enactment", θεωρούνται σοβαρά λάθη.
- Λανθασμένοι όροι οι οποίοι σχετίζονται με το θέμα του κειμένου και είναι βασικοί στην κατανόησή του, όπως ο όρος "trust" σε κείμενο καταπιστεύματος ("καταπίστευμα", λανθασμένα μεταφρασμένο ως "εμπιστοσύνη") θεωρούνται σοβαρά λάθη.

- Το ίδιο λάθος δεν έχει απαραίτητα την ίδια σοβαρότητα σε όλα τα σημεία του κειμένου. Π.χ. το αμετάφραστο αρκτικόλεξο "CF" ("cystic fibrosis"/"κυστική ίνωση" σε ιατρικό κείμενο) θεωρείται σοβαρό λάθος όταν βρίσκεται μόνο του στην αρχή του κειμένου αλλά δευτερεύον όταν βρίσκεται μέσα στο κείμενο σε παρένθεση μετά από τον ελληνικό όρο.
- Στην κατεύθυνση ελληνικά-αγγλικά, το Moses, σε αντίθεση με τα υπόλοιπα συστήματα που αξιολογήθηκαν, δεν πραγματοποιεί μεταγραφή στο λατινικό αλφάβητο στις αμετάφραστες λέξεις. Αυτό έχει ως αποτέλεσμα όλες οι αμετάφραστες λέξεις του να θεωρούνται σοβαρά λάθη.

Ομοιογένεια

- Δεν επισημειώνουμε λάθη ομοιογένειας, π.χ. στο ίδιο κείμενο (L_en_gr_4) εμφανίζεται τόσο η λεξιλογική επιλογή "η νήσος" όσο και η επιλογή "το νησί", τα οποία δεν διορθώνονται και δεν επισημειώνονται, παρόλο που στη μεταφραστική πρακτική θα έπρεπε να τηρηθεί ομοιογένεια.
- Στο κείμενο L_en_gr_4 παρουσιάζονται όροι με τους ορισμούς τους. Στη μετάφραση του Moses ακολουθούνται δύο τακτικές: α) άνω και κάτω τελεία, β) χρήση του ρήματος "νοείται". Πρόκειται για ανομοιογένεια, η οποία επίσης δεν επισημειώνεται.
- Άλλου είδους ανομοιογένεια που παρουσιάστηκε αλλά δεν επισημειώθηκε είναι στην κατεύθυνση ελληνικά-αγγλικά, σε λίστα με μεικτή χρήση γερουνδίων και ουσιαστικών.

Παράγοντες που επηρεάζουν την ποιότητα επισημείωσης

- Ποιότητα προηγούμενης μετάφρασης. Όπως συμβαίνει σε κάθε είδους αξιολόγηση, ιδιαίτερα όταν αυτή αφορά τη γλώσσα, οι κριτές επηρεάζονται κατά την αξιολόγηση ενός κειμένου από την ποιότητα του/των προηγούμενου/-ων κειμένου/-ων που αξιολόγησαν (αρνητική επιρροή όταν τα προηγούμενα κείμενα ήταν ποιοτικά και θετική όταν τα προηγούμενα ήταν κακής ποιότητας).
- Το μέγεθος του κειμένου προς αξιολόγηση σε συνδυασμό με την ποιότητά του. Εάν το κείμενο είναι ταυτόχρονα και ιδιαίτερο μεγάλο και ιδιαίτερα απαιτητικό ως προς την επισημείωση, καθίσταται πιο δύσκολη ή και αδύνατη η πλήρης επισημείωσή του χωρίς διακοπές, κάτι το οποίο ιδανικά πρέπει να τηρείται.

Λοιπές παρατηρήσεις

- Θεωρούμε ότι η διαγραφή είναι μικρότερη διόρθωση από την αντικατάσταση.

- Η μοναδική περίπτωση στην οποία δεν συμφωνεί ο αριθμός διορθώσεων στη Μετεπεξεργασία με τον αριθμό λαθών στην Ταξινόμηση Λαθών είναι η περίπτωση ύπαρξης τυπογραφικού λάθος στο πρωτότυπο, εξ αιτίας των οποίων παραμένουν αμετάφραστες λέξεις στο κείμενο-στόχος. Σε αυτές τις περιπτώσεις γίνεται η διόρθωση στη μετεπεξεργασία, ώστε να μπορέσει να αξιοποιηθεί το μετεπεξεργασμένο κείμενο, αλλά δεν προσμετράται στα λάθη του συστήματος η αμετάφραστη λέξη.
- Στη μετάφραση προς ελληνικά δημιουργούνται πολλά "πλασματικά" λάθη λόγω του ότι η πρόθεση "σε" συντίθεται με το οριστικό άρθρο. Εάν π.χ. υπάρχει λανθασμένη χρήση άλλης πρόθεσης, όπως "για την", για τη διόρθωσή της απαιτείται και απαλοιφή της πρόθεσης "για" και αντικατάσταση του άρθρου "την".
- Δεν παρατηρήθηκαν περιπτώσεις λανθασμένης ορθογραφίας, ώστε να κριθεί τελικά απαραίτητη η ύπαρξη αυτής της κατηγορίας.

4.3 Αποτελέσματα αξιολόγησης αγγλικά-ελληνικά

4.3.1 Αποτελέσματα αυτόματης αξιολόγησης

Στον Πίνακα 7 παρουσιάζονται τα αποτελέσματα της μετρικής Bleu με τρεις μεταφράσεις-αναφορά για τα κείμενα της κατεύθυνσης αγγλικά-ελληνικά.

EN-GR	GOOGLE	BING	SYSTRAN	MOSES
L_en_gr_1	7660	5735	3987	3577
L_en_gr_2	6157	5219	4729	4993
L_en_gr_3	5248	4351	4252	4342
L_en_gr_4	3371	2743	2216	1772
L_en_gr_5	5366	4612	3369	3900
M_en_gr_1	5563	4357	3053	4055
M_en_gr_2	4296	3807	2858	2937
M_en_gr_3	6574	3346	2714	2543
M_en_gr_4	4386	3972	2937	2995
M_en_gr_5	4967	3676	2792	2959
T_en_gr_1	4425	3617	2768	2664
T_en_gr_2	4784	3781	3773	2883
T_en_gr_3	5356	4064	3349	3190
T_en_gr_4	5652	4589	3901	4120
T_en_gr_5	4952	3892	3302	3745
M.O.	**5250.47**	**4117.40**	**3333.33**	**3378.33**

Πίνακας 7: Αποτελέσματα Bleu με 3 μεταφράσεις-αναφορά για την κατεύθυνση αγγλικά-ελληνικά

Όπως αναφέρεται και στο Κεφάλαιο 3.1.1.4, η βαθμολογία του Bleu κυμαίνεται από 0 έως 1. Οι αριθμοί στον Πίνακα 5, καθώς και στους επόμενους πίνακες με αποτελέσματα Bleu, είναι τα δεκαδικά ψηφία (ο αριθμός 7660 αντιστοιχεί σε 0,7660 κ.ο.κ.) και όπου αναφέρεται 1 είναι η μέγιστη δυνατή βαθμολογία, που δηλώνει πλήρη ταύτιση μηχανικού μεταφράσματος με μετάφραση-αναφορά. Ως προς τα αποτελέσματα του κάθε συστήματος ανά πεδίο, το Systran έχει χαμηλότερη βαθμολογία στα ιατρικά κείμενα από ό,τι στα κείμενα των άλλων πεδίων και το Bing υψηλότερη βαθμολογία στα νομικά από ό,τι στα κείμενα των άλλων πεδίων. Ως προς την κατάταξη των συστημάτων μεταξύ τους, στο σύνολο των κειμένων (μέσος όρος) το Google κατατάσσεται πρώτο, το Bing δεύτερο, το Moses τρίτο και το Systran τέταρτο. Μεταξύ Moses και Systran υπάρχει μικρή διαφορά, καθώς σε κάποια από τα κείμενα το Systran λαμβάνει υψηλότερη βαθμολογία από το Moses. Αξίζει να σημειωθεί πως το γεγονός ότι το σύστημα Moses δεν εκτελεί μέχρι στιγμής μορφοποίηση του μεταφρασμένου κειμένου με βάση το αρχικό έχει αρνητικές επιπτώσεις στη βαθμολογία του στο Bleu. Στα αποτελέσματα του Google η μακράν υψηλότερη βαθμολογία είναι του κειμένου L_en_gr_1_MT1 (7660). Στο ίδιο κείμενο λαμβάνει την υψηλότερη βαθμολογία του και το Bing (5735). Αυτό εξηγείται από το γεγονός ότι το εν λόγω κείμενο είναι η "Οικουμενική διακήρυξη για τα ανθρώπινα δικαιώματα", ένα κείμενο με πολλαπλές εμφανίσεις στο διαδίκτυο, είτε αυτούσιο είτε αποσπασματικά, οι οποίες προφανώς υπάρχουν χωρίς διαφοροποιήσεις μέσα στο μεταφραστικό μοντέλο των Google και Bing, με αποτέλεσμα να αυξάνονται οι πιθανότητες χρήσης αυτούσιων ν-γραμμάτων από αυτό. Στα αποτελέσματα του Google την υψηλότερη βαθμολογία από τα ιατρικά κείμενα έλαβε το M_en_gr_3_MT1 (6574) και από τα τεχνικά το T_en_gr_4_MT1 (5652). Συνεπώς θεωρήσαμε σκόπιμο αυτά τα τρία κείμενα να συμπεριληφθούν στο σώμα της μη αυτοματοποιημένης αξιολόγησης, ώστε να διαπιστώσουμε εάν συμπίπτει η υψηλή βαθμολογία Bleu με την ανθρώπινη κρίση. Η εργασία αξιολόγησης που ενδείκνυται για αυτή την περίπτωση είναι ο Έλεγχος Ποιότητας, ο οποίος αναπτύσσεται στο Κεφάλαιο 3.2.1.2.

Επίσης στα αποτελέσματα μεταξύ των κειμένων του Bing και του Systran υπάρχουν δύο με ελάχιστη διαφορά στη βαθμολογία: το L_en_gr_3 (Bing: 4351, Systran: 4252) και το T_en_gr_2 (Bing: 3781, Systran: 3773). Τα κείμενα αυτά επιλέχθηκαν για εργασία επισημείωσης Κατάταξης, με στόχο να διαπιστώσουμε ποιο από τα δύο είναι πιο ποιοτικό σύμφωνα με την ανθρώπινη κρίση. Η Κατάταξη επίσης παρουσιάζεται στο Κεφάλαιο 3.2.1.2.

Καθώς το μεταφραστικό σύστημα βασισμένο στο Moses είναι υπό ανάπτυξη, η αξιολόγηση που θεωρήθηκε πιο κατάλληλη για τα εξαγόμενα αυτού ήταν η Ταξινόμηση Λαθών, με στόχο την παρατήρηση σε βάθος των φαινομένων και την επεξεργασία πιθανών τρόπων βελτίωσης του συστήματος. Για αυτόν τον σκοπό επιλέχτηκαν τα εξής 4 κείμενα με τα κριτήρια που αναφέρονται στη συνέχεια: L_en_gr_1_MT4, L_en_gr_4_MT4, M_en_gr_1_MT4 και T_en_gr_5_MT4. Το L_en_gr_4_MT4 είναι το κείμενο με τη χαμηλότερη βαθμολογία Bleu (1772) και για αυτόν τον λόγο θεωρήθηκε απαραίτητη η μη αυτοματοποιημένη αξιολόγησή του. Επίσης θεωρήθηκε σκόπιμη η αξιολόγηση τουλάχιστον ενός κειμένου από κάθε τομέα. Τέλος, η Ταξινόμηση Λαθών μας επιτρέπει την παρατήρηση της συμπεριφοράς του συστήματος στις μεγάλες προτάσεις. Ως προς τη μετάφραση μεγάλων προτάσεων και την ποιότητά της, έχει παρατηρηθεί ότι παρουσιάζουν αυξημένο βαθμό δυσκολίας στην επισημείωση, κάτι το οποίο μπορεί να οφείλεται σε δύο παράγοντες: στην ποιότητα της πρότασης και στο γεγονός ότι ο βαθμός δυσκολίας επισημείωσης μιας μικρής πρότασης είναι μικρότερος από αυτόν μιας μεγάλης της ίδιας ποιότητας λόγω μεγαλύτερου όγκου υλικού προς επισημείωση. Συγκεκριμένα οι Koehn, Ph. και Monz, Ch. [2006] αναφέρουν σχετικά με τη μη αυτοματοποιημένη αξιολόγηση στατιστικών συστημάτων: “Σύμφωνα με τους επισημειωτές, είναι σχεδόν αδύνατο να κρίνει κανείς τις μεγάλες προτάσεις. Καθώς η μετάφραση οποιασδήποτε μεγάλης πρότασης είναι κάπως *μπερδεμένη*, ακόμη και η συγκριτική αξιολόγηση μεταξύ συστημάτων ήταν δύσκολη. Κάποιοι επισημειωτές πρότειναν να χωριστούν οι μεγάλες προτάσεις σε γραμματικές προτάσεις και να αξιολογηθούν ξεχωριστά”[53]. Με βάση αυτή την παρατήρηση αλλά και την υπόθεση αυξημένης συντακτικής πολυπλοκότητας στις μεγάλες προτάσεις, θεωρήσαμε αναμενόμενο το σύστημα να έχει χαμηλότερες επιδόσεις σε αυτές και προκειμένου να διαπιστώσουμε αν επιβεβαιώνεται αυτή η υπόθεση, ένα από τα κριτήρια επιλογής κειμένων προς αξιολόγηση είναι και η ύπαρξη μεγάλων προτάσεων σε αυτά. Ως μικρές προτάσεις θεωρούμε τις προτάσεις με 1-14 λέξεις, μεσαίες τις προτάσεις με 15-29 λέξεις και μεγάλες τις προτάσεις με 30 λέξεις τουλάχιστον.

Στον Πίνακα 8 παρουσιάζονται οι βαθμολογίες Bleu με τρεις και με δύο μεταφράσεις-αναφορά, καθώς και η ποσοστιαία διαφορά των δύο βαθμολογιών. Για τις δύο μεταφράσεις-αναφορά, αυτή που επιλέξαμε να αφαιρέσουμε για όλα τα σώματα κειμένων ήταν η HT1, μιας και είναι οι μεταφράσεις τις οποίες συλλέξαμε από το διαδίκτυο και η πιθανότητα αυτά τα κείμενα να έχουν αξιοποιηθεί για την εκπαίδευση των Google και Bing είναι μεγάλη.

[53] Σε δική μας μετάφραση.

EN-GR	Google 3 ref	Google 2 ref	% change	Bing 3 ref	Bing 2 ref	% change	Systran 3 ref	Systran 2 ref	% change	Moses 3 ref	Moses 2 ref	% change
L_1	7660	4088	46,63	5735	4092	28,65	3987	3652	8,40	3577	3198	10,60
L_2	6157	6093	1,04	5219	5106	2,17	4729	4666	1,33	4993	4830	3,26
L_3	5248	5074	3,32	4351	4007	7,91	4252	3994	6,07	4342	3928	9,53
L_4	3371	2413	28,42	2743	2151	21,58	2216	1867	15,75	1772	1519	14,28
L_5	5366	4706	12,30	4612	4126	10,54	3369	2945	12,79	3900	3482	10,72
M_1	5563	4403	20,85	4357	3671	15,74	3053	2714	11,10	4055	3408	15,96
M_2	4296	3573	16,83	3807	3216	15,52	2858	2358	17,49	2937	2512	14,47
M_3	6574	5210	20,75	3346	3113	6,96	2714	2489	8,29	2543	2336	8,14
M_4	4386	3848	12,27	3972	3402	14,35	2937	2653	9,67	2995	2552	14,12
M_5	4967	4528	8,84	3676	3316	9,79	2792	2368	15,19	2959	2631	11,08
T_1	4425	4390	0,79	3617	3586	0,86	2768	2753	0,54	2664	2645	0,71
T_2	4784	4721	1,32	3781	3736	1,19	3773	3751	0,58	2883	2828	1,91
T_3	5356	4815	10,10	4064	3555	12,52	3349	3028	9,58	3190	2774	13,04
T_4	5652	5451	3,56	4589	4316	5,95	3901	3779	3,13	4120	3983	3,33
T_5	4952	4830	2,46	3892	3773	3,06	3302	3273	0,88	3745	3666	2,11
M.O.	**5250,47**	**4542,87**	**12,63**	**4117,40**	**3677,73**	**10,45**	**3333,33**	**3086,00**	**8,04**	**3378,33**	**3087,47**	**8,88**

Πίνακας 8: Αποτελέσματα Bleu με 2 και 3 μεταφράσεις-αναφορά για την κατεύθυνση αγγλικά-ελληνικά

Όπως ήταν αναμενόμενο, όλα τα συστήματα σε όλα τα κείμενα έλαβαν χαμηλότερη βαθμολογία με τις δύο μόνο μεταφράσεις-αναφορά, με το Google και το Bing να παρουσιάζουν τη μεγαλύτερη μείωση, η οποία μάλιστα είναι ιδιαίτερα μεγάλη σε συγκεκριμένα κείμενα, μεταξύ των οποίων και το L_en_gr_1, για το οποίο έγινε λόγος στην ανάλυση του Πίνακα 5. Θεωρούμε ότι η χρήση μεταφράσεων από το διαδίκτυο έχει κάποια επιρροή στα αποτελέσματα, καθώς και η επιλογή των κειμένων, όπως είδαμε στην περίπτωση του L_en_gr_1. Όμως όταν πρόκειται για κείμενα-αναφορά τα οποία δεν συμπεριλαμβάνονται στο σώμα εκπαίδευσης, η διαφορά μεταξύ χρήσης δύο και τριών μεταφράσεων-αναφορά δεν είναι ουσιαστική ή τουλάχιστον όχι γενικεύσιμη στο σώμα της εργασίας μας.

4.3.2 Αποτελέσματα μη αυτοματοποιημένης αξιολόγησης

4.3.2.1 Έλεγχος Ποιότητας (Quality Checking)

Όπως προαναφέρθηκε, τα κείμενα του Google με την καλύτερη βαθμολογία Bleu σε κάθε πεδίο αξιολογήθηκαν με Έλεγχο Ποιότητας. Συγκεκριμένα πρόκειται για τα κείμενα L_en_gr_1_MT1 (7660), M_en_gr_3_MT1 (6574) και T_en_gr_4_MT1 (5652). Όπως είδαμε πιο αναλυτικά στο Κεφάλαιο 3.2.1.2, οι ετικέτες επισημείωσης είναι οι εξής: “acceptable”, “can easily be fixed”, “none of both”, και χρησιμοποιήθηκαν όπως ορίστηκαν στο ανωτέρω κεφάλαιο. Αυτή η εργασία επισημείωσης πραγματοποιήθηκε έχοντας υπόψη ότι ο εικονικός διορθωτής του κειμένου έχει πρόσβαση στο πρωτότυπο και σε επίπεδο πρότασης, όχι κειμένου. Ο διορθωτής του κειμένου είναι εικονικός, γιατί δεν γίνεται παράλληλη διόρθωση του κειμένου, αλλά λαμβάνεται υπόψη κατά την εκτέλεση της αξιολόγησης, καθώς ο επισημειωτής καλείται να απαντήσει αν ο διορθωτής θα μπορούσε εύκολα να διορθώσει την εκάστοτε πρόταση. Τα αποτελέσματα είναι τα εξής:

	Acceptable	Can easily be fixed	None of both	Αριθμός προτάσεων κειμένου	Bleu
L_en_gr_1_MT1	25 (24.27%)	77 (74.76%)	1 (0.97%)	103	7660
M_en_gr_3_MT1	13 (14.44%)	61 (67.78%)	16 (17.78%)	90	6574
T_en_gr_4_MT1	9 (25.71%)	26 (74.29%)	0 (0.00%)	35	5652

Πίνακας 9: Αποτελέσματα Ελέγχου Ποιότητας

Οι βαθμολογίες αντιστρέφονται κατά κάποιο τρόπο. Το κείμενο που είχε τη χαμηλότερη βαθμολογία Bleu (T_en_gr_4_MT1) είναι αυτό που στη μη αυτοματοποιημένη αξιολόγηση έχει το μεγαλύτερο ποσοστό αποδεκτών προτάσεων, ελαφρώς υψηλότερο από το L_en_gr_1_MT1, που ήταν μακράν το καλύτερο βάσει Bleu, και καμία πρόταση “none of both”. Επίσης αυτά τα δύο κείμενα έχουν σχεδόν ίδιο ποσοστό προτάσεων που επιδέχονται εύκολη διόρθωση. Το δε κείμενο M_en_gr_3_MT1 έχει σαφώς μικρότερο ποσοστό αποδεκτών προτάσεων από τα άλλα δύο, έχει μικρότερο ποσοστό προτάσεων που επιδέχονται εύκολη διόρθωση, αλλά και πολύ μεγαλύτερο ποσοστό προτάσεων με την ετικέτα “None of both”, ετικέτα η οποία σχεδόν δεν χρησιμοποιήθηκε στα άλλα δύο κείμενα. Συνολικά παρατηρείται διάσταση μεταξύ της αυτόματης και της μη αυτοματοποιημένης αξιολόγησης ως προς την κατάταξη των κειμένων, ωστόσο στα κείμενα L_en_gr_1_MT1 και T_en_gr_4_MT1 η συσχέτιση θεωρείται υψηλή.

Όσο αφορά τα χαρακτηριστικά των κειμένων T_en_gr_4_MT1 και M_en_gr_3_MT1, το τεχνικό είναι κείμενο που απευθύνεται στο ευρύ κοινό με δημοσιογραφικό λόγο, με ορολογία από τομείς Ε.Ε. και τεχνολογίας, καθώς και με ονομασίες οργανισμών, κ.λπ., ενώ το M_en_gr_3_MT1 έχει μακροπερίοδο λόγο με μεγάλο αριθμό ιατρικών όρων αρκετά εξειδικευμένων και περιγραφές δραστηριοτήτων θεραπείας με διάφορους ρηματικούς τύπους, τόσο σε μορφή συνεχούς κειμένου όσο και σε μορφή απαρίθμησης. Αξίζει επίσης να σημειωθεί ότι το τεχνικό είναι κείμενο της Ε.Ε., στα σώματα της οποίας προφανώς έχει εκπαιδευτεί το Google.

Ως προς τη σχέση μεταξύ ποιότητας μετάφρασης και μήκους πρότασης, ακολουθούν τρεις πίνακες (Πίνακες 10-12), στους οποίους παρουσιάζονται τα μήκη των προτάσεων που κρίθηκαν αποδεκτές, καθώς επίσης και ένας πίνακας με τις προτάσεις οι οποίες έλαβαν τον χαρακτηρισμό None of both και τα μήκη τους. Στους πίνακες αναφέρεται ο αριθμός της πρότασης, το μήκος της πρωτότυπης πρότασης (αριθμός λέξεων) και η βαθμολογία Bleu.

M_en_gr_3_MT1

acceptable sent id	μήκος πρότασης	bleu score
5	13	1
7	6	1
8	11	1
24	11	1
28	22	1
36	9	1
42	22	9889
43	25	6133
50	8	8512
51	23	7861
59	9	1
74	18	1
88	10	1

Πίνακας 10: Μήκη και βαθμολογίες Bleu αποδεκτών προτάσεων του κειμένου M_en_gr_3_MT1

Ο μέσος ορος μήκους αποδεκτής πρότασης σε αυτό το κείμενο είναι 14,38. Οι αποδεκτές προτάσεις είναι μέτριες και μικρές. Επιβεβαιώνεται συσχέτιση βαθμολογίας Bleu με τα αποτελέσματα της μη αυτοματοποιημένης επισημείωσης.

Στο ίδιο κείμενο υπάρχουν περισσότερες προτάσεις με ιδιαίτερα υψηλή βαθμολογία Bleu, οι οποίες όμως έχουν λάβει στη μη αυτοματοποιημένη αξιολόγηση τον χαρακτηρισμό “can easily be fixed”. Ακολουθούν κάποια παραδείγματα, όπου αναφέρονται η βαθμολογία Bleu, η πρωτότυπη και η μεταφρασμένη πρόταση με τις λέξεις που χρήζουν διόρθωση με έντονους χαρακτήρες.

Πρόταση 4 (8180): “The length of treatment is raised gradually from 25 or 30 minutes at the start of the course till 45 to 90 minutes at the end (including breaks for rest).” → “Η διάρκεια της θεραπείας αυξάνεται σταδιακά από 25 ή 30 λεπτά κατά την έναρξη της **πορείας** μέχρι 45 έως 90 λεπτά στο τέλος (συμπεριλαμβανομένων των διαλειμμάτων για ανάπαυση).”

Πρόταση 6 (7938): “The aim of this prior workout is to render less pronounced those posture-tone reflexes which will be handled during the Adeli Suit treatment program.” → “Ο σκοπός αυτής της προθέρμανσης είναι να γίνουν λιγότερο έντονα τα αντανακλαστικά στάσης-τόνου που θα **χειρίζεται** κατά τη διάρκεια του προγράμματος θεραπείας με το κοστούμι Adeli.”

Πρόταση 18 (8409): “3. improving the mobility of the joints;” → “3. βελτίωση της κινητικότητας των αρθρώσεων**?**”

Πρόταση 41 (7253): “When lying on the back, the patient should be taught to raise and turn the head.” → “Όταν **βρίσκεται στο πίσω μέρος**, ο ασθενής πρέπει να μάθει να σηκώνει και να γυρίζει το κεφάλι.”

Στις παραπάνω προτάσεις υπάρχουν κυρίως λάθη λεξιλογικής επιλογής, η διόρθωση των οποίων δεν απαιτεί μεγάλη μεταφραστική προσπάθεια, καθώς δεν πρόκειται για όρους, δεν παρουσιάζουν αμφισημία ή κάποια άλλη μεταφραστική δυσκολία. Στην πρόταση 18 υπάρχει λανθασμένο σημείο στίξης, το οποίο επίσης πολύ εύκολα διορθώνεται.

L_en_gr_1_MT1		
acceptable sent id	μήκος πρότασης	bleu score
1	4	1623
13	18	1
18	11	8612
23	16	1
27	18	7170
30	27	8960
32	11	7911
36	36	7609
40	23	1
43	17	6233
50	18	1
54	17	1
55	22	1
58	10	1
65	12	5946
66	11	1
68	20	9087
76	14	8939
86	7	5133
87	12	1
90	25	1
91	28	8693
97	24	7772
99	20	7998
101	21	1

Πίνακας 11: Μήκη και βαθμολογίες Bleu αποδεκτών προτάσεων του κειμένου L_en_gr_1_MT1

Ο μέσος όρος μήκους αποδεκτής πρότασης σε αυτό το κείμενο είναι 17,68 λέξεις. Οι περισσότερες προτάσεις είναι μέτριου μήκους. Ως προς τα αποτελέσματα Bleu, οι βαθμολογίες είναι ιδιαίτερα υψηλές και συνάδουν με τον χαρακτηρισμό της αποδεκτής πρότασης με μία μόνο εξαίρεση (πρόταση 1).

T_en_gr_4_MT1

acceptable sent id	μήκος πρότασης	bleu score
2	14	7242
7	30	8467
14	3	2364
20	27	5123
24	7	1492
27	2	1
28	25	6971
32	33	8761
34	1	7156

Πίνακας 12: Μήκη και βαθμολογίες Bleu αποδεκτών προτάσεων του κειμένου T_en_gr_4_MT1

Ο μέσος όρος μήκους αποδεκτής πρότασης σε αυτό το κείμενο είναι 15,77 και εμφανίζονται τόσο μικρές όσο και μέτριες αλλά και μεγάλες προτάσεις. Επίσης σε αυτή την περίπτωση η μετρική Bleu παρουσιάζει υψηλή συσχέτιση με τη μη αυτοματοποιημένη επισημείωση με εξαίρεση τις προτάσεις 14 (2364) και 24 (1492). Σε αυτές τις προτάσεις φαίνεται μία από τις αδυναμίες των μετρικών που λειτουργούν με μεταφράσεις-αναφορά. Παραθέτουμε τη μηχανικά μεταφρασμένη πρόταση και τις προτάσεις-αναφορά:

Πρόταση 14
HT1: Αποδοτική διαχείριση ραδιοφάσματος
HT2: Αποτελεσματική διαχείριση ραδιοφάσματος
HT3: Αποδοτική διαχείριση ραδιοφάσματος
MT1: Η αποτελεσματική διαχείριση του ραδιοφάσματος

Πρόταση 24
HT1: Δέσμευση ραδιοφάσματος για έρευνα και εξοικονόμηση ενέργειας
HT2: Εξασφάλιση ποσότητας ραδιοφάσματος για έρευνα και εξοικονόμηση ενέργειας
HT3: Δέσμευση ραδιοφάσματος για έρευνα και εξοικονόμηση ενέργειας
MT1: Δέσμευση φάσματος για την έρευνα και την εξοικονόμηση ενέργειας

Οι διαφορές είναι η επιλογή της λέξης “φάσμα” αντί για “ραδιοφάσμα”, η οποία είναι αποδεκτή ως υπερώνυμο, και η ύπαρξη άρθρων, η οποία ελάχιστα αλλοιώνει τη μορφή της πρότασης και την ποιότητά της, ωστόσο σε επίπεδο σύγκρισης Bleu με τις μεταφράσεις-αναφορά, το σύστημα εντοπίζει κοινά μόνο τα χαμηλότερα ν-γράμματα, με αποτέλεσμα να μειώνεται η τελική βαθμολογία της πρότασης.

None of both sent id	μήκος πρότασης	bleu score
9	53	5419
10	72	5847
13	29	4834
17	20	5194
20	15	3349
21	12	4301
25	74	3870
26	16	4791
31	28	4368
38	22	6676
44	38	6961
49	22	3826
52	17	5985
58	28	9591
66	14	6965
71	23	7227

Πίνακας 13: Μήκη και βαθμολογίες Bleu προτάσεων "None of both" του κειμένου M_en_gr_3_MT1

Στον Πίνακα 13 παρουσιάζονται τα μήκη των προτάσεων του M_en_gr_3_MT1 των οποίων η διόρθωση κρίθηκε ότι είναι πιο απαιτητική. Ο μέσος όρος μήκους αυτών των προτάσεων είναι 30,18, πολύ μεγαλύτερος από τους μέσους όρους μήκους των αποδεκτών προτάσεων. Όσο αφορά τη βαθμολογία Bleu, είναι αρκετά υψηλή. Στον Πίνακα 14 παρουσιάζονται κάποιες από τις προτάσεις με την πιο υψηλή βαθμολογία που χαρακτηρίστηκαν "None of both".

Μετάφραση	**Bleu**
"It is important to train the feet to rest uniformly and take care to keep the upright posture control while maintaining balancing reactions." → "Είναι σημαντικό να εκπαιδεύσουμε τα πόδια για να ξεκουραστούν ομοιόμορφα και να φροντίζει να διατηρεί τον έλεγχο όρθιας στάσης, ενώ διατηρεί τις αντιδράσεις ισορροπίας."	7227
Ενώ η μεταφρασμένη πρόταση δεν παρουσιάζει πολλά λάθη, υπάρχει ένα σημείο σε αυτήν ("να εκπαιδεύσουμε τα πόδια για να ξεκουραστούν ομοιόμορφα"), η διόρθωση του οποίου απαιτεί μεταφραστική πράξη με επιλογές που δεν είναι δεδομένες. Η μεταφραστική προσπάθεια που απαιτείται καθιστά την πρόταση "None of both".	
"When tone (asymmetrical and symmetrical) reflexes appear in the neck, the position of the extremities, as determined by a change in the tone of the flexors and extensors, will also depend on a certain position of the head." → "Όταν τόνο (ασύμμετρα και συμμετρικά) αντανακλαστικά εμφανίζονται στο λαιμό, η θέση των	6961

άκρων, όπως προσδιορίζεται από την αλλαγή στον τόνο των καμπτήρων και των εκτεινόντων, θα εξαρτηθεί επίσης από μία ορισμένη θέση της κεφαλής.”	
Η ύπαρξη πολλών όρων καθιστά την πρόταση “None of both”, διότι απαιτείται έρευνα για την επιβεβαίωση της απόδοσης των όρων στα ελληνικά.	
“During each exercise, special attention should always be paid to arresting faulty synkinesis and synergy and to the position of the head.” → “Κατά τη διάρκεια κάθε άσκησης, θα πρέπει πάντα να δοθεί ιδιαίτερη προσοχή στην αναχαίτιση λανθασμένης συγκινησίας και τη συνέργεια και τη θέση του κεφαλιού.”	6676
Και σε αυτή την πρόταση το πλήθος όρων την καθιστά “None of both”, διότι απαιτείται έρευνα για την επιβεβαίωση της απόδοσης των όρων στα ελληνικά.	

Πίνακας 14: Προτάσεις με υψηλή βαθμολογία Bleu από το κείμενο M_en_gr_3_MT1

Στα παραδείγματα του Πίνακα 14 γίνεται αντιληπτό ότι η μετρική Bleu μας δίνει μια ποσοτική πληροφορία, ουσιαστικά πόσες από τις λέξεις της πρότασης συμπίπτουν με τις ιδανικές μεταφράσεις, αλλά όχι την ποιοτική πληροφορία σχετικά με τον χαρακτήρα των λέξεων που δεν συμπίπτουν, δηλαδή σχετικά με την ποιότητά τους, τον βαθμό δυσκολίας διόρθωσής τους, τον βαθμό πιστότητας και γραμματικότητάς τους. Ας σημειωθεί και σε αυτό το σημείο ότι ο βαθμός δυσκολίας διόρθωσης εξαρτάται και από τον βαθμό δυσκολίας μετάφρασης, που είναι άμεσα συνυφασμένος με κάποια χαρακτηριστικά του πρωτότυπου. Καθώς δεν γίνεται παράλληλη διόρθωση, δεν λαμβάνονται υπόψη οι μεταφράσεις-αναφορά και δεν γίνεται καμία προσπάθεια διόρθωσης, δηλώνεται μόνο το γεγονός ότι απαιτείται μεταφραστική πράξη για τη διόρθωσή τους. Το γεγονός ότι μια πρόταση κρίθηκε “None of both” δεν σημαίνει απαραίτητα ότι θα αλλάξει πολύ κατά τη διορθωσή της.

Από αυτό το μικρό δείγμα προτάσεων που χρησιμοποιήθηκαν στον Έλεγχο Ποιότητας προκύπτουν τα ακόλουθα συμπεράσματα.

Σε επίπεδο κειμένου παρατηρείται διάσταση μεταξύ αυτόματης και μη αυτοματοποιημένης αξιολόγησης ως προς την κατάταξη των κειμένων, η συσχέτιση όμως στα δύο από τα τρία κείμενα θεωρείται υψηλή. Σε επίπεδο πρότασης υπάρχει συσχέτιση στις προτάσεις που κρίνονται ποιοτικές βάσει ανθρώπινης κρίσης. Όσο αφορά τις προτάσεις που κρίθηκε ότι απαιτούν μεγαλύτερη προσπάθεια διόρθωσης, πρόκειται για ένα ποιοτικό κριτήριο, το οποίο δεν μπορεί να συγκριθεί με τα ποσοτικά δεδομένα του Bleu.

Ως προς το μήκος των προτάσεων, οι προτάσεις που κρίθηκαν αποδεκτές είναι μικρότερες από αυτές που απαιτούν μεγαλύτερη προσπάθεια διόρθωσης, χωρίς όμως να παύουν να υπάρχουν και εξαιρέσεις σε αυτό το γενικό συμπέρασμα. Συγκεκριμένα, το μέσο μήκος αποδεκτής πρότασης είναι 15,94 λέξεις, ενώ το μέσο μήκος των προτάσεων που απαιτούν μεγαλύτερη προσπάθεια κατά τη διόρθωση είναι 30,18 λέξεις, κάτι το οποίο αποτελεί ένδειξη του ότι η αυτόματη μετάφραση μικρών προτάσεων είναι πιο ποιοτική από τη μετάφραση μεγάλων προτάσεων.

4.3.2.2 Κατάταξη δύο συστημάτων (3-way ranking)

Η Κατάταξη πραγματοποιήθηκε μεταξύ των κειμένων L_en_gr_3_MT2 και L_en_gr_3_MT3 και των T_en_gr_2_MT2 και T_en_gr_2_MT3, καθώς σε αυτά το Bing και το Systran έλαβαν σχεδόν την ίδια βαθμολογία. Σε αυτή την εργασία επισημειώνονται οι προτάσεις ανα ζεύγη των δύο συστημάτων επιλέγοντας μία από τις τρεις ετικέτες: A>B, A=B και A<B.

Τα αποτελέσματα αυτής της εργασίας είναι τα εξής:

L_en_gr_3_MT2MT3 (67 προτάσεις)	T_en_gr_2_MT2MT3 (51 προτάσεις)
A<B 18 (26.87%)	A<B 26 (50.98%)
A=B 23 (34.33%)	A=B 9 (17.65%)
A>B 26 (38.81%)	A>B 16 (31.37%)

Πίνακας 15: Αποτελέσματα Κατάταξης στο Appraise

Σύμφωνα με τη μη αυτοματοποιημένη αξιολόγηση, στο κείμενο L_en_gr_3 το Bing έχει καλύτερο αποτέλεσμα από το Systran με μεγαλύτερη διαφορά από ό,τι στα αποτελέσματα του Bleu (L_en_gr_3_MT2 (Bing): 4351, L_en_gr_3_MT3 (Systran): 4252). Στο κείμενο T_en_gr_2 έχει καλύτερο αποτέλεσμα το Systran, με μεγάλη μάλιστα διαφορά, σε αντίθεση και πάλι με τα αποτελέσματα του Bleu, όπου η διαφορά μεταξύ των δύο βαθμολογιών ήταν πολύ μικρή (Bing: 3781, Systran: 3773). Όπως και στον Έλεγχο Ποιότητας, σε επίπεδο κειμένου δεν υπάρχει συμφωνία μεταξύ αυτόματης και μη αυτοματοποιημένης αξιολόγησης.

Σε επίπεδο πρότασης, η συμφωνία μεταξύ αυτόματης και μη αυτοματοποιημένης αξιολόγησης είναι στο κείμενο L_en_gr_3 68,65% (σύμπτωση 46 προτάσεων από τις 67) και στο κείμενο

T_en_gr_2 54,90% (σύμπτωση 28 προτάσεων από τις 51).

4.3.2.3 Ταξινόμηση Λαθών και Μετεπεξεργασία (Error Classification και Post-editing)

Οι εργασίες Ταξινόμησης Λαθών και Μετεπεξεργασίας πραγματοποιήθηκαν για τέσσερα κείμενα εξαγόμενα του Moses, τα οποία αντιστοιχούν στον συνολικό αριθμό των 229 προτάσεων. Στον Πίνακα 14 παρουσιάζονται τα αποτελέσματα της Ταξινόμησης Λαθών για την κατεύθυνση αγγλικά-ελληνικά. Συγκεκριμένα παρουσιάζεται ο αριθμός λαθών ανά επίπεδο σοβαρότητας κάθε κατηγορίας σε κάθε κείμενο, ο αριθμός λέξεων του κειμένου, το ποσοστό λαθών κάθε κατηγορίας μέσα στην κάθε μεταφρασμένη πρόταση, καθώς και τα συνολικά ποσοστά ανα κείμενο. Τέλος, αναφέρονται οι συνολικοί αριθμοί και τα συνολικά ποσοστά ανά κατηγορία. Οι στήλες με κόκκινο αντιστοιχούν σε σοβαρά λάθη.

EN-GR	Ελ. Λέξεις	Υπ. Λάθη	Ορολ Δευτ.	Ορολ Σοβ.	Λεξ. Επ. Δευτ.	Λεξ. Επ. Σοβ.	Σύντ. Δευτ.	Σύντ. Σοβ.	Προσ θήκη Δευτ.	Προσ θήκη Σοβ.	Μορ φολο γία Δευτ.	Μορ φολο γία Σοβ.	Ελ. Λειτ. Λέξη Δευτ.	Ελ. Λειτ. Λέξη Σοβ.	Στίξη Δευτ.	Στίξη Σοβ.	Αμετ. Δευτ.	Αμετ. Σοβ.	Σύν. Λέξ.
L_4	17	0	8	34	72	10	4	6	37	13	111	2	17	0	21	0	3	1	716
T_4	16	0	3	0	48	1	9	3	26	1	75	0	15	0	3	0	2	1	754
L_1	21	0	12	2	100	0	5	1	50	3	121	0	58	0	6	0	1	0	1710
M_1	24	0	12	19	75	15	63	16	47	1	141	1	47	0	12	1	13	7	1591
Σύν.	**78**	**0**	**35**	**55**	**295**	**26**	**81**	**26**	**160**	**18**	**448**	**3**	**137**	**0**	**42**	**1**	**19**	**9**	
Ποσοστό %																			Συνολικό ποσοστό
L_4	2,37	0,00	1,12	4,75	10,06	1,40	0,56	0,84	5,17	1,82	15,50	0,28	2,37	0,00	2,93	0,00	0,42	0,14	**49,72**
T_5	2,12	0,00	0,40	0,00	6,37	0,13	1,19	0,40	3,45	2,13	9,95	0,00	1,99	0,00	0,40	0,00	0,27	0,13	**26,92**
L_1	1,23	0,00	0,70	0,12	5,85	0,00	0,29	0,06	2,92	0,18	7,08	0,00	3,39	0,00	0,35	0,00	0,06	0,00	**22,22**
M_1	1,51	0,00	0,75	1,19	4,71	0,94	3,96	1,01	2,95	0,06	8,86	0,06	2,95	0,00	0,75	0,06	0,82	0,44	**31,05**
Σύν.	**7,23**	**0,00**	**2,97**	**6,06**	**26,98**	**2,47**	**6,00**	**2,30**	**14,49**	**2,19**	**41,39**	**0,34**	**10,71**	**0,00**	**4,44**	**0,06**	**1,56**	**0,71**	

Πίνακας 16: Αποτελέσματα Ταξινόμησης Λαθών ανά επίπεδο σοβαρότητας λάθους για την κατεύθυνση αγγλικά-ελληνικά

Οι εμφανίσεις σοβαρών λαθών είναι λίγες και υπάρχει μία κατηγορία (Στίξη) με μηδενικό αποτέλεσμα στο επίπεδο σοβαρού λάθους. Οι κατηγορίες στις οποίες δύο κείμενα (το L_en_gr_4 και το M_en_gr_1) παρουσιάζουν μεγάλο αριθμό σοβαρών λαθών συμπίπτουν: Ορολογία, Λεξιλογική Επιλογή και Σύνταξη. Στο L_en_gr_4 συναντώνται σοβαρά λάθη και στην κατηγορία της Προσθήκης Λέξης. Τα λάθη Μορφολογίας, τα οποία και αποτελούν το μεγαλύτερο μέρος των λαθών (451 λάθη), όπως φαίνεται και στον Πίνακα 15, είναι στη συντριπτική πλειονότητά τους δευτερεύοντα. Μεγάλος αριθμός μορφολογικών λαθών είναι αναμενόμενος όταν η γλώσσα-στόχος είναι μορφολογικά πλούσια, ωστόσο, σύμφωνα με το δείγμα μας, είναι σημαντικότερο να αντιμετωπιστούν τα λεξιλογικά και συντακτικά λάθη από ό,τι τα μορφολογικά.

Στον Πίνακα 17 παρουσιάζεται ο αριθμός και το ποσοστό λαθών συνολικά ανά κατηγορία και όχι ανάλογα με το επίπεδο σοβαρότητας του λάθους.

	Ελ. Λέξεις	Υπερ. Λάθη	Ορολο-γία	Λεξ. Επιλογή	Σύνταξη	Προσθή-κη	Μορφο-λογία	Ελ. Λειτ. Λέξη	Στίξη	Αμετά-φραστο	Σύνολο λέξεων
L_4	17	0	42	82	10	50	113	17	21	4	716
T_5	16	0	3	49	12	27	75	15	3	3	754
L_1	21	0	14	100	6	53	121	58	6	6	1710
M_1	24	0	31	90	79	48	142	47	12	20	1591
Σύνολο	**78**	**0**	**90**	**321**	**107**	**178**	**451**	**137**	**42**	**30**	
Ποσοστό %											Συνολικό ποσοστό
L_4	2,37	0,00	5,87	11,45	1,40	6,98	15,78	2,37	2,93	0,56	**49,72**
T_5	2,12	0,00	0,40	6,50	1,59	3,58	9,95	1,99	0,40	0,40	**26,92**
L_1	1,23	0,00	0,82	5,85	0,35	3,10	7,08	3,39	0,35	0,18	**22,34**
M_1	1,51	0,00	1,95	5,66	4,97	3,02	8,93	2,95	0,75	1,26	**30,99**
Σύνολο	**7,23**	**0,00**	**9,03**	**29,46**	**8,30**	**16,68**	**41,73**	**10,71**	**4,44**	**2,39**	

Πίνακας 17: Αποτελέσματα Ταξινόμησης Λαθών για την κατεύθυνση αγγλικά-ελληνικά

Στον Πίνακα 18 παρουσιάζονται τα συνολικά ποσοστά λαθών ανά κείμενο και οι βαθμολογίες Bleu ανά κείμενο.

	Ποσοστό λαθών	Bleu
L_en_gr_4_MT4	49,72%	1772
T_en_gr_5_MT4	26,92%	3745
L_en_gr_1_MT4	22,34%	3577
M_en_gr_1_MT4	30,99%	4055

Πίνακας 18: Ποσοστά λαθών και βαθμολογία Bleu ανά κείμενο στην κατεύθυνση αγγλικά-ελληνικά

Υπάρχει αρκετά υψηλή συσχέτιση μεταξύ αυτόματης και μη αυτοματοποιημένης αξιολόγησης στα νομικά κείμενα καθώς και στο τεχνικό, καθώς το L_en_gr_4_MT4, το οποίο έχει ιδιαίτερα χαμηλή βαθμολογία Bleu, έχει και ιδιαίτερα υψηλό ποσοστό λαθών, και τα T_en_gr_5_MT4 και L_en_gr_1_MT4 έχουν μικρή διαφορά μεταξύ τους τόσο σε βαθμολογία Bleu όσο και σε ποσοστό λαθών, αν και αντίστροφα. Δεν υπάρχει πλήρης συμφωνία μεταξύ αυτόματης και μη αυτοματοποιημένης αξιολόγησης ως προς το πώς κατατάσσονται τα κείμενα. Σύμφωνα με την αυτόματη μετρική, κατατάσσονται ως εξής: 1) M_en_gr_1_MT4, 2) T_en_gr_5_MT4, 3) L_en_gr_1_MT4, 4) L_en_gr_4_MT4, και σύμφωνα με τη μη αυτοματοποιημένη: 1) L_en_gr_1_MT4, 2) T_en_gr_5_MT4, 3) M_en_gr_1_MT4, 4) L_en_gr_4_MT4.

Από την παρατήρηση των αποτελεσμάτων της αξιολόγησης ανά πρόταση σε συνδυασμό με το μήκος πρότασης και τη βαθμολογία Bleu ανά πρόταση προκύπτουν οι ακόλουθες παρατηρήσεις. Ως προς τη σχέση μήκους-ποιότητας πρότασης, ενώ δεν υπάρχει διαφορά σε επίπεδο παρουσίας λαθών γενικά, παρατηρείται το εξής φαινόμενο: τα σοβαρά λάθη εμφανίζονται κυρίως σε μέτριες και μεγάλες προτάσεις και το πλήθος τους αυξάνεται όσο ανεβαίνει ο αριθμός λέξεων της πρότασης. Το ποσοστό προτάσεων με σοβαρά λάθη στις μικρές προτάσεις είναι 10,11%, στις μεσαίες 27,71% και στις μεγάλες 56,36%. Η υπόθεση ότι αυτό οφείλεται σε αύξηση της συντακτικής πολυπλοκότητας δεν επιβεβαιώνεται, καθώς τα σοβαρά συντακτικά λάθη είναι 25 από τα συνολικά 135 σοβαρά λάθη και αυτά τα 25 δεν είναι συγκεντρωμένα στις μεγάλες προτάσεις. Στον Πίνακα 19 παρουσιάζεται ο αριθμός σοβαρών λαθών ανά κατηγορία.

Κατηγορία	**Αριθμός σοβαρών λαθών**
Σύνταξη	25
Ορολογία	52
Λεξιλογική Επιλογή	27
Προσθήκη Λέξης	18
Μορφολογία	3
Ελλείπουσα Λειτουργική Λέξη	0
Στίξη	1
Αμετάφραστη Λέξη	9

Πίνακας 19: Αριθμός σοβαρών λαθών ανά κατηγορία για την κατεύθυνση αγγλικά-ελληνικά

Ως προς τη συσχέτιση μη αυτοματοποιημένης και αυτόματης αξιολόγησης, στην πλειονότητα των προτάσεων με βαθμολογία Bleu ίση ή μεγαλύτερη του 0,5000 (υψηλή βαθμολογία) υπάρχει μικρός αριθμός σοβαρών λαθών, όπως φαίνεται στον Πίνακα 20. Συγκεκριμένα μόνο 7 από τις 75 προτάσεις με υψηλή βαθμολογία (9,33%) περιέχουν σοβαρά λάθη.

Βαθμολογία Bleu	**Δευτερεύοντα λάθη**	**Σοβαρά λάθη**	**Αριθμός λέξεων μεταφρασμένης πρότασης**
5946	1		4
5958	7		47
5005	3		20

5082	1		23
5852	9	2	68
8409	1		2
5558	6	1	41
5334	2	1	25
6795	4	1	44
5979	4		21
6819	5	4	31
5855	16		29
8378			21
8409	1		2
6819	4	2	46
1			1
5327	6		29
5009	2		12
5192	6		19
7071			2
7071			2
7071			2
7071			2
7071			2
7071			2
7071			2
7071			2
7071			2
7071			2
7071			2
7071			2
7071			2
7071			2
7071			2
7071			2
7071			2
7071			2
7071			2
7071			2
7071			2
7071			2

7071			2
7071			2
7071			2
7071			2
7071			2
7071			2
7071			2
7071			2
5347	1		10
6102			15
5384	1		15
5133			8
7612			11
7027	5		25
5557	6		18
5207			13
5133			7
5774			5
7861	2		23
6279	7		28
5861	7		21
6239	1		12
5360	1		17
5232	4		24
6043	1		20
8932			10
5183	7		26
6484			10
5465	2	1	27
1			6
5675	2		36
5193	2		10
5121	7		27
6389			2

Πίνακας 20: Σχέση σοβαρών λαθών με υψηλή βαθμολογία Bleu για την κατεύθυνση αγγλικά-ελληνικά

Στις χαμηλές βαθμολογίες (βαθμολογία ίση ή χαμηλότερη του 0,2000) δεν παρατηρείται τόσο μεγάλη συσχέτιση (Πίνακας 21). Ναι μεν υπάρχουν 22 προτάσεις με σοβαρά λάθη από τις συνολικά 56 προτάσεις με χαμηλή βαθμολογία (39,28%) και σε κάποιες περιπτώσεις ο αριθμός των σοβαρών λαθών είναι μεγάλος, ωστόσο υπάρχουν πολλές προτάσεις με λίγα ή και κανένα δευτερεύον λάθος. Από το συγκεκριμένο δείγμα συμπεραίνουμε πως η μετρική Bleu αντικατοπτρίζει την ποιοτική μετάφραση σε αρκετά ικανοποιητικό βαθμό, αλλά μπορεί να αποδώσει χαμηλή βαθμολογία και σε μεταφράσεις οι οποίες δεν περιέχουν πολλά λάθη, περιέχουν όμως ν-γράμματα που δεν συμπίπτουν με τα ν-γράμματα των μεταφράσεων-αναφορά.

Βαθμολογία Bleu	**Δευτερεύοντα λάθη**	**Σοβαρά λάθη**	**Αριθμός λέξεων μεταφρασμένης πρότασης**
1285	5		13
1135	15	1	33
1399	9	1	29
1597	3		5
0959	5	1	14
0437			5
1362	9	1	18
0977	4		40
1253	8		37
0747	4		13
1210	1		33
1870	11		28
0366	5		15
1226	12		19
1666	2	1	14
1284	4		7
1463	4		16
1385	8		28
1587	5		36
1823	6		19
1005	6		10
1710	10		21
1580	9		25
0804	10		19
1492	3	1	8
1487	8		14
1531	7		29
0845	2		7

0284	9		20
0552	3		5
0366	7		15
1068			2
1068			2
0534	2		2
0657	4		4
1503	10	2	19
1597			3
0943	5	2	23
0558	11	2	17
0610	6	3	14
0740	8	5	18
0929	3	1	5
1976	10	4	36
1534	4	1	21
1236	6		8
0743	4	1	10
1822	22	3	60
0626	7		13
1387	28	8	43
1505	11	3	24
1625	8	9	41
1249	4		17
0504	3	3	14
0827	7	2	15
0454	6		8
0996	13	1	27

Πίνακας 21: Σχέση σοβαρών λαθών με χαμηλή βαθμολογία Bleu για την κατεύθυνση αγγλικά-ελληνικά

4.4 Αποτελέσματα αξιολόγησης ελληνικά-αγγλικά

4.4.1 Αποτελέσματα αυτόματης αξιολόγησης

Στον Πίνακα 22 παρουσιάζονται τα αποτελέσματα της μετρικής Bleu με τρεις μεταφράσεις-αναφορά για τα κείμενα της κατεύθυνσης ελληνικά-αγγλικά.

GR-EN	GOOGLE	BING	SYSTRAN	MOSES
L_gr_en_1	4167	4088	2896	3814
L_gr_en_2	2374	1426	3104	2262
L_gr_en_3	5400	4893	3218	3287
L_gr_en_4	3301	2764	2206	3475
L_gr_en_5	4982	3950	2545	2911
M_gr_en_1	4064	4093	2016	3119
M_gr_en_2	4671	3867	1956	2698
M_gr_en_3	3285	3304	1719	2457
M_gr_en_4	4089	2131	1705	2459
M_gr_en_5	4471	4340	2086	3194
T_gr_en_1	4950	5186	3105	3874
T_gr_en_2	4938	5064	2730	3563
T_gr_en_3	6442	6515	2736	3678
T_gr_en_4	5485	5264	2943	2715
T_gr_en_5	4753	4012	2373	3101
M.O.	**4491.47**	**4059.80**	**2489.20**	**3107.13**

Πίνακας 22: Αποτελέσματα Bleu με 3 μεταφράσεις-αναφορά για την κατεύθυνση ελληνικά-αγγλικά

Τα συστήματα κατατάσσονται με την ίδια σειρά όπως και στην κατεύθυνση αγγλικά-ελληνικά: 1) Google, 2) Bing, 3) Moses, 4) Systran. Ως προς τα αποτελέσματα ανά πεδίο, το Bing έχει υψηλότερη βαθμολογία στα τεχνικά κείμενα από ό,τι στα κείμενα των άλλων πεδίων, το Systran λαμβάνει χαμηλότερη βαθμολογία στα ιατρικά κείμενα από ό,τι στα κείμενα των άλλων πεδίων, ενώ το Google και το Moses παρουσιάζουν ομοιομορφία ως προς τα αποτελέσματα σε όλα τα πεδία, όπως και στην ευθεία μετάφραση.

Στον Πίνακα 23 παρουσιάζονται τα συνολικά αποτελέσματα των συστημάτων στις δύο κατευθύνσεις.

	Google	Bing	Systran	Moses
EN-GR	5250,47	4117,40	3333,33	3378,33
GR-EN	4491,47	4059,80	2489,20	3107,13

Πίνακας 23: Αποτελέσματα Bleu ανά σύστημα στις δύο κατευθύνσεις

Όλα τα συστήματα λαμβάνουν χαμηλότερη βαθμολογία στις αντίστροφες μεταφράσεις, κάτι το οποίο δεν θεωρείται αναμενόμενο, δεδομένου ότι η ελληνική είναι μια πλούσια μορφολογικά γλώσσα και αυτό μπορεί να επηρεάσει τη γραμματικότητα του μεταφράσματος και συνεπώς τη βαθμολογία της αυτόματης μετρικής αρνητικά. Αυτό εξηγείται, όπως θα φανεί στη συνέχεια, από το γεγονός ότι στην αντίστροφη μετάφραση αυξάνονται άλλα είδη λαθών, με αποτέλεσμα να αντισταθμίζεται η μείωση μορφολογικών λαθών.

Στον Πίνακα 24 παρουσιάζονται οι βαθμολογίες Bleu με τρεις και με δύο μεταφράσεις-αναφορά, καθώς και η ποσοστιαία διαφορά των δύο βαθμολογιών.

GR-EN	Google 3 ref	Google 2 ref	% change	Bing 3 ref	Bing 2 ref	% change	Systran 3 ref	Systran 2 ref	% change	Moses 3 ref	Moses 2 ref	% change
L_1	4167	3908	6,22	4088	3864	5,48	2896	2741	5,35	3814	3667	3,85
L_2	2374	2050	13,65	1426	1314	7,85	3104	2792	10,05	2262	1843	18,52
L_3	5400	3787	29,87	4893	3734	23,69	3218	2326	27,72	3287	2617	20,38
L_4	3301	3072	6,94	2764	2714	1,81	2206	2049	7,12	3475	3244	6,65
L_5	4982	3290	33,96	3950	3119	21,04	2545	2230	12,38	2911	2492	14,39
M_1	4064	3711	8,69	4093	3693	9,77	2016	1891	6,20	3119	2901	6,99
M_2	4671	4272	8,54	3867	1432	11,25	1956	1731	11,50	2698	2501	7,30
M_3	3285	2802	14,70	3304	2935	11,17	1719	1374	20,07	2457	2209	10,09
M_4	4089	3528	13,72	2131	1935	9,20	1705	1268	25,63	2459	2159	12,20
M_5	4471	3939	11,90	4340	3836	11,61	2086	1812	13,14	3194	2821	11,68
T_1	4950	4295	13,23	5186	4658	10,18	3105	2834	8,73	3874	3553	8,29
T_2	4938	4552	7,82	5064	4740	6,40	2730	2501	8,39	3563	3292	7,61
T_3	6442	5415	15,94	6515	5411	16,95	2736	2257	17,51	3678	3173	13,73
T_4	5485	5254	4,21	5264	5027	4,50	2943	2792	5,13	2715	2545	6,26
T_5	4753	4431	6,77	4012	3798	5,33	2373	2258	4,85	3101	2954	4,74
M.O.	**4491,47**	**3887,07**	**13,08**	**4059,80**	**3614,00**	**10,42**	**2489,20**	**2190,40**	**12,25**	**3107,13**	**2798,07**	**10,18**

Πίνακας 24: Αποτελέσματα Bleu με 2 και 3 μεταφράσεις-αναφορά για την κατεύθυνση ελληνικά-αγγλικά

Όπως ήταν αναμενόμενο και όπως παρατηρήθηκε και στην κατεύθυνση αγγλικά-ελληνικά, όλα τα συστήματα σε όλα τα κείμενα έλαβαν χαμηλότερη βαθμολογία με τις δύο μόνο μεταφράσεις-αναφορά. Το Google, όπως και στην ευθεία μετάφραση, παρουσιάζει τη μεγαλύτερη διαφορά, ωστόσο όλα τα συστήματα δείχνουν να επηρεάζονται περισσότερο στα ίδια κείμενα. Μη αναμενόμενος επίσης θεωρείται ο βαθμός στον οποίο επηρεάζεται το Systran, καθώς δεν ανήκει στα στατιστικά συστήματα, στα οποία ενδέχεται να έχει χρησιμοποιηθεί στο σώμα εκπαίδευσής τους η πρώτη μετάφραση-αναφορά.

Από τα νομικά κείμενα τα L_gr_en_1, L_gr_en_2 και L_gr_en_4 προέρχονται από την ίδια πηγή και η πρώτη μετάφραση-αναφορά, η οποία και αφαιρέθηκε σε αυτό το πείραμα, είναι η μετάφραση που υπάρχει στην ίδια ιστοσελίδα από την οποία αντλήθηκαν και τα πρωτότυπα κείμενα. Το ίδιο ισχύει και για τα κείμενα L_gr_en_3 και L_gr_en_5. Η αφαίρεση της πρώτης μετάφρασης-αναφορά επηρεάζει σημαντικά μόνο τα κείμενα L_gr_en_3 και L_gr_en_5 σε όλα τα συστήματα. Μέτρια επιρροή παρατηρείται στο κείμενο L_gr_en_2 επίσης σε όλα τα συστήματα.

Αντίστοιχα παρατηρείται μεγάλη διαφορά σε όλα τα συστήματα στο κείμενο T_gr_en_3 σε σχέση με τα υπόλοιπα του ίδιου πεδίου, παρόλο που τα T_gr_en_2, T_gr_en_3 και T_gr_en_4 είναι όλα αποσπάσματα από το ίδιο κείμενο, κάτι το οποίο αποκλείει το ενδεχόμενο να έχει χρησιμοποιηθεί μόνο το ένα από αυτά στην εκπαίδευση των συστημάτων. Από τα τρία αυτά κείμενα το T_gr_en_3 είναι το μοναδικό στο οποίο υπάρχουν προτάσεις με σχεδόν απόλυτη ομοιότητα μεταξύ HT1 και Google, η οποία ομοιότητα δεν περιορίζεται σε όρους αλλά εντοπίζεται και σε ολόκληρες δομές, όπως στο ακόλουθο παράδειγμα: HT1: "**The Electronics Laboratory is equipped with tools for** design, simulation, layout, prototype development, characterization and control of opto-electronic and microelectronic systems and devices." / Google: "**The Electronics Laboratory is equipped with tools for** design, simulation, layout, prototype development, characterization and control of optoelectronic and microelectronic systems and devices.". Αυτό μας οδηγεί στο συμπέρασμα ότι πιθανώς έχει χρησιμοποιηθεί αυτή η μηχανή στην πρώτη μετάφραση-αναφορά. Οι ομοιότητες μεταξύ των στατιστικών συστημάτων πιθανότατα οφείλονται σε κοινό σώμα εκπαίδευσης.

Όπως και στην ευθεία μετάφραση, επιλέχθηκαν για Ταξινόμηση Λαθών και Μετεπεξεργασία τέσσερις μεταφράσεις του Moses, τουλάχιστον μία από το κάθε πεδίο: L_gr_en_1_MT4, L_gr_en_2_MT4, M_gr_en_2_MT4 και T_gr_en_1_MT4. Τα νομικά είναι τα κείμενα με την υψηλότερη και τη χαμηλότερη βαθμολογία στο πεδίο τους (L_gr_en_1_MT4: 3814, L_gr_en_2_MT4: 2262), το τεχνικό είναι το κείμενο με την υψηλότερη βαθμολογία στο πεδίο του (3874) και το ιατρικό είναι το κείμενο με την υψηλότερη βαθμολογία (2698) από τα κείμενα μεσαίου μεγέθους του πεδίου του.

4.4.2 Αποτελέσματα μη αυτοματοποιημένης αξιολόγησης: Ταξινόμηση Λαθών και Μετεπεξεργασία (Error Classification και Post-editing)

Τα τέσσερα κείμενα τα οποία υποβλήθηκαν σε Ταξινόμηση Λαθών και Μετεπεξεργασία στην αντίστροφη μετάφραση αντιστοιχούν στον συνολικό αριθμό των 102 προτάσεων. Στον Πίνακα 24 παρουσιάζονται τα αποτελέσματα της εργασίας Ταξινόμησης Λαθών για την κατεύθυνση ελληνικά-αγγλικά. Ο πίνακας περιλαμβάνει τις εξής πληροφορίες: αριθμός λαθών ανά επίπεδο σοβαρότητας κάθε κατηγορίας σε κάθε κείμενο, αριθμός λέξεων μεταφρασμένου κειμένου, ποσοστό λαθών κάθε κατηγορίας μέσα στην κάθε πρόταση και συνολικά ποσοστά ανα κείμενο. Επίσης αναφέρονται οι συνολικοί αριθμοί και τα συνολικά ποσοστά ανά κατηγορία. Οι στήλες με κόκκινο αντιστοιχούν σε σοβαρά λάθη.

GR-EN	Ελ. Λέξεις	Υπ. Λάθη	Ορολ Δευτ.	Ορολ Σοβ.	Λεξ. Επ. Δευτ.	Λεξ. Επ. Σοβ.	Σύντ. Δευτ.	Σύντ. Σοβ.	Προσθήκη Δευτ.	Προσθήκη Σοβ.	Μορφολογία Δευτ.	Μορφολογία Σοβ.	Ελ. Λειτ. Λέξη Δευτ.	Ελ. Λειτ. Λέξη Σοβ.	Στίξη Δευτ.	Στίξη Σοβ.	Αμετ. Δευτ.	Αμετ. Σοβ.	Σύν. Λέξ.
L_1	12	0	36	3	33	3	12	0	37	0	14	0	20	0	7	0	0	21	675
L_2	10	0	20	0	48	2	8	1	39	0	5	0	19	0	6	0	0	12	450
M_2	0	0	5	1	46	2	14	0	34	0	15	0	19	0	5	0	0	28	414
T_1	12	0	8	0	47	1	27	0	38	1	14	0	11	0	10	0	0	18	590
Σύν.	**34**	**0**	**69**	**4**	**174**	**8**	**61**	**1**	**148**	**1**	**48**	**0**	**69**	**0**	**28**	**0**	**0**	**79**	
Ποσοστό %																			Συνολικό ποσοστό
L_1	1,78	0,00	5,33	0,44	4,89	0,44	1,78	0,00	5,48	0,00	2.07	0,00	2,96	0,00	1,04	0,00	0,00	3,11	**29,33**
L_2	2,22	0,00	4,44	0,00	10,67	0,44	1,78	0,22	8,67	0,00	1,11	0,00	4,22	0,00	1,33	0,00	0,00	2,67	**37,78**
M_2	0,00	0,00	1,21	0,24	11,11	0,48	3,38	0,00	8,21	0,00	3,62	0,00	4,59	0,00	1,21	0,00	0,00	6,76	**40,82**
T_1	2,03	0,00	1,38	0,00	7,97	0,17	4,58	0,00	6,44	0,17	2,37	0,00	1,86	0,00	1,69	0,00	0,00	3,05	**31,69**
Σύν.	**6,03**	**0,00**	**12,34**	**0,69**	**34,63**	**1,54**	**11,51**	**0,22**	**28,80**	**0,17**	**9,18**	**0,00**	**13,64**	**0,00**	**5,27**	**0,00**	**0,00**	**15,59**	

Πίνακας 25: Αποτελέσματα Ταξινόμησης Λαθών ανά επίπεδο σοβαρότητας λάθους για την κατεύθυνση ελληνικά-αγγλικά

Όπως και στην ευθεία μετάφραση, η παρουσία σοβαρών λαθών δεν είναι σημαντική. Στην αντίστροφη οι κατηγορίες με μηδενικά αποτελέσματα είναι τρεις (Στίξη, η οποία συμπίπτει και με την ευθεία, Ελλείπουσα Λειτουργική Λέξη και Μορφολογία).

Στον Πίνακα 26 παρουσιάζεται ο αριθμός και το ποσοστό λαθών συνολικά ανά κατηγορία και όχι ανάλογα με το επίπεδο σοβαρότητας του λάθους.

	Ελλ. Λέξεις	Υπερβ. Λάθη	Ορολογία	Λεξ. Επιλογή	Σύνταξη	Προσθήκη	Μορφολογία	Ελ. Λειτ. Λέξη	Στίξη	Αμετάφραστο	Σύνολο Λέξεων
L_gr_en_1_MT4	12	0	39	36	12	37	14	20	7	21	675
L_gr_en_2_MT4	10	0	20	50	9	39	5	19	6	12	450
M_gr_en_2_MT4	0	0	6	48	14	34	15	19	5	28	414
T_gr_en_1_MT4	12	0	8	48	27	39	14	11	10	18	590
	34	0	73	182	62	149	48	69	28	79	Συνολικό Ποσοστό
L_gr_en_1_MT4	1.78%	0.00%	5.78%	5.33%	1.78%	5.48%	2.07%	2.96%	1.04%	3.11%	29.33%
L_gr_en_2_MT4	2.22%	0.00%	4.44%	11.11%	2.00%	8.67%	1.11%	4.22%	1.33%	2.67%	37.78%
M_gr_en_2_MT4	0.00%	0.00%	1.45%	11.59%	3.38%	8.21%	3.62%	4.59%	1.21%	6.76%	40.82%
T_gr_en_1_MT4	2.03%	0.00%	1.36%	8.14%	4.58%	6.61%	2.37%	1.86%	1.69%	3.05%	31.69%
Σύνολο	6.03%	0.00%	13.03%	36.17%	11.74%	28.97%	9.18%	13.64%	5.27%	15.59%	

Πίνακας 26: Αποτελέσματα Ταξινόμησης Λαθών για την κατεύθυνση ελληνικά-αγγλικά

Στην αντίστροφη μετάφραση η Μορφολογία δεν είναι πλέον, όπως στην ευθεία, η κατηγορία με τον μεγαλύτερο αριθμό λαθών, κάτι το οποίο είναι αναμενόμενο, εφόσον η γλώσσα-στόχος (αγγλικά) δεν είναι μορφολογικά τόσο πλούσια. Η κατηγορία με τα περισσότερα λάθη είναι η Λεξιλογική Επιλογή.

Στον Πίνακα 27 παρουσιάζονται τα συνολικά ποσοστά λαθών ανά κείμενο και οι βαθμολογίες Bleu ανά κείμενο.

	Ποσοστό λαθών	Bleu
L_gr_en_1_MT4	29,33%	3814
L_gr_en_2_MT4	37,78%	2262
M_gr_en_2_MT4	40,82%	2698
T_gr_en_1_MT4	31,69%	3874

Πίνακας 27: Ποσοστά λαθών και βαθμολογία Bleu ανά κείμενο για την κατεύθυνση ελληνικά-αγγλικά

Υπάρχει συσχέτιση σε ικανοποιητικό βαθμό σε όλα τα κείμενα, καθώς τα δύο κείμενα με την υψηλότερη βαθμολογία Bleu (L_gr_en_1_MT4 και T_gr_en_1_MT4) έχουν και τα χαμηλότερα ποσοστά λαθών, καθώς και τα δύο κείμενα με τη χαμηλότερη βαθμολογία (L_gr_en_2_MT4 και M_gr_en_2_MT4) έχουν τα μεγαλύτερα ποσοστά λαθών.

Ακολουθεί συγκριτικός πίνακας με τα συνολικά ποσοστά λαθών ανά κατηγορία στις δύο κατευθύνσεις.

	Ελλ. λέξεις	Υπερβ. λάθη	Ορολ.	Λεξ. επ.	Σύνταξη	Προσθ.	Μορφ.	Ελ. λειτ. λέξη	Στίξη	Αμετάφρ.	Μέσο ποσοστό
EN-GR	7,23%	0,00%	9,03%	29,46%	8,30%	16,68%	41,73%	10,71%	4,44%	2,39%	32,49%
GR-EN	6,03%	0,00%	13,03%	36,17%	11,74%	28,97%	9,18%	13,64%	5,27%	15,59%	34,91%

Πίνακας 28: Συνολικά ποσοστά λαθών ανά κατηγορία στις δύο κατευθύνσεις

Οι διαφορές που παρουσιάζονται στις κατηγορίες λαθών είναι αναμενόμενες δεδομένης της κατεύθυνσης της μετάφρασης. Τα μορφολογικά λάθη στη μετάφραση προς τα αγγλικά είναι πολύ λιγότερα από ό,τι στη μετάφραση προς τα ελληνικά, τα οποία είναι πολύ πιο πλούσια μορφολογικά γλώσσα. Τα λάθη Προσθήκης Λέξης (πλεονάζουσα λέξη) επίσης παρουσιάζουν μεγάλη αύξηση προς τα αγγλικά. Αυτό πιθανώς σχετίζεται με το γεγονός ότι τα κείμενα που έχουν μεταφραστεί από τα ελληνικά στα αγγλικά είναι συνήθως μικρότερα σε μήκος από ό,τι τα αντίστοιχα πρωτότυπά τους όταν η μετάφραση είναι ποιοτική· στην προκειμένη περίπτωση προφανώς δεν έχουν πραγματοποιηθεί όλες οι απαραίτητες μετατροπές. Τα λάθη Σύνταξης επίσης παρουσιάζουν μικρή αύξηση στην αντίστροφη μετάφραση, φαινόμενο το οποίο εξηγείται από την αυστηρή σειρά λέξεων στην αγγλική. Αύξηση λαθών στην αντίστροφη μετάφραση παρατηρείται και στις κατηγορίες της Ορολογίας και της Λεξιλογικής επιλογής, κάτι το οποίο εξαρτάται από το μεταφραστικό μοντέλο. Πιθανώς το γεγονός ότι, λόγω έλλειψης παράλληλων κειμένων με πρωτότυπη γλώσσα τα ελληνικά, χρησιμοποιούνται στην εκπαίδευση των συστημάτων τα ίδια παράλληλα σώματα που χρησιμοποιούνται και στην ευθεία (σώματα με πρωτότυπα κείμενα στα αγγλικά και μεταφρασμένα στα ελληνικά) έχει ως αποτέλεσμα το μεταφραστικό μοντέλο να μην είναι τόσο αποδοτικό όσο στην ευθεία. Στην κατηγορία της Ελλείπουσας Λειτουργικής Λέξης υπάρχει μικρή αύξηση στην αντίστροφη κατεύθυνση, φαινόμενο επίσης αναμενόμενο λόγω της υποχρεωτικής χρήσης προσωπικών αντωνυμιών αλλά και του μεγάλου αριθμού βοηθητικών ρημάτων που χρησιμοποιούνται στην αγγλική. Τέλος, στην αντίστροφη μετάφραση το Moses αφήνει πολύ περισσότερες λέξεις αμετάφραστες από ό,τι στην ευθεία. Αυτό συμβαίνει είτε γιατί οι συγκεκριμένες λέξεις δεν υπάρχουν στο σώμα εκπαίδευσης είτε γιατί ενώ υπάρχουν δεν επιλέγονται λόγω μικρότερης πιθανότητας που τους ανατίθεται. Το συνολικό ποσοστό σημειώνει μικρή αύξηση στην αντίστροφη κατεύθυνση, το οποίο συνάδει με τη μικρή πτώση στη βαθμολογία της αυτόματης μετρικής.

Όσο αφορά τη συσχέτιση μη αυτοματοποιημένης και αυτόματης αξιολόγησης σε επίπεδο πρότασης, σε μεγάλο μέρος των προτάσεων με βαθμολογία Bleu ίση ή μεγαλύτερη του 0,5000 υπάρχει μικρός αριθμός σοβαρών λαθών, όπως φαίνεται στον Πίνακα 29. Συγκεκριμένα 6 από τις 19 προτάσεις με υψηλή βαθμολογία (31,57%) περιέχουν σοβαρά λάθη. Αξίζει να σημειωθεί ότι 82 από τα συνολικά 96 σοβαρά λάθη είναι αμετάφραστες λέξεις.

Βαθμολογία Bleu	**Δευτερεύοντα λάθη**	**Σοβαρά λάθη**	**Αριθμός λέξεων μεταφρασμένης πρότασης**
1			2
6996	1	1	8
6155	4	1	18
1			2
5507	10		5
5564	6		32
5133	10	2	36
1			1
6231	2		21
5504	3		19
5268	1		16

1			2
5491	4		7
7071		1	2
5383	5		21
5788	4		39
6043	1		5
6418	2		10
5416	3	1	10

Πίνακας 29: Σχέση σοβαρών λαθών με υψηλή βαθμολογία Bleu για την κατεύθυνση ελληνικά-αγγλικά

Στις χαμηλές βαθμολογίες (βαθμολογία ίση ή χαμηλότερη του 0,2000) παρατηρείται μεγαλύτερη συσχέτιση από ό,τι στην ευθεία μετάφραση (Πίνακας 30). 18 από τις 27 προτάσεις (16,66%) περιέχουν σοβαρά λάθη.

Βαθμολογία Bleu	Δευτερεύοντα λάθη	Σοβαρά λάθη	Αριθμός λέξεων μεταφρασμένης πρότασης
1697	8	1	23
0534	6		5
1238	1		9
1383	2	1	4
1797	2		4
1255	3		8
1797	1	1	3
1454	2	1	7
1426	3		6
0987	14		20
1263	19	3	47
1708	2	1	18
0673	18	1	26
1882	8	2	32
1551	4		12
1736	11	7	29
1892	9	1	14
1652	4		7
1163	9		15
1150	7	1	16
1668	10	1	17

1276	13	1	34
0965	2	1	3
1440	11	2	18
1124	11	2	29
1340	2	1	17
1778	3	1	13

Πίνακας 30: Σχέση σοβαρών λαθών με χαμηλή βαθμολογία Bleu για την κατεύθυνση ελληνικά-αγγλικά

4.5 Πείραμα με αντιστροφή μεταφράσεων που αξιολογήθηκαν αγγλικά-ελληνικά

Ο μεγάλος αριθμός μορφολογικών λαθών στην ευθεία μετάφραση αποδόθηκε στον μορφολογικό πλούτο της ελληνικής. Προκειμένου να ελεγθεί αυτή η υπόθεση, θεωρήθηκε σκόπιμο να επισημειωθούν κείμενα με όσο το δυνατόν μεγαλύτερο βαθμό αντιστοιχίας με τα κείμενα που επισημειώθηκαν στην κατεύθυνση αγγλικά-ελληνικά. Για αυτό το πείραμα χρησιμοποιήθηκαν μεταφράσεις από τα ελληνικά προς τα αγγλικά στις οποίες το κείμενο-πηγή ήταν μία από τις μεταφράσεις-αναφορά που αξιολογήθηκαν στην ευθεία κατεύθυνση. Αυτά τα νέα κείμενα-πηγή μεταφράστηκαν από το Moses και υποβλήθηκαν μόνο σε μη αυτοματοποιημένη Ταξινόμηση Λαθών και Μετεπεξεργασία, καθώς δεν υπήρχε σώμα μεταφράσεων-αναφορά για αυτόματη αξιολόγηση.

4.5.1 Ταξινόμηση Λαθών και Μετεπεξεργασία (Error Classification και Post-editing)

Τα αποτελέσματα της Ταξινόμησης Λαθών παρουσιάζονται στον Πίνακα 30. Αναφέρεται ο αριθμός λαθών ανά επίπεδο σοβαρότητας κάθε κατηγορίας σε κάθε κείμενο, ο αριθμός λέξεων του μεταφρασμένου κειμένου, το ποσοστό λαθών κάθε κατηγορίας μέσα στην κάθε πρόταση, τα συνολικά ποσοστά ανά κείμενο, καθώς και οι συνολικοί αριθμοί και τα συνολικά ποσοστά ανά κατηγορία. Οι στήλες με κόκκινο αντιστοιχούν σε σοβαρά λάθη.

HT1 MT4	Ελ. Λέξεις	Υπ. Λάθη	Ορολ Δευτ.	Ορολ Σοβ.	Λεξ. Επ. Δευτ.	Λεξ. Επ. Σοβ.	Σύντ. Δευτ.	Σύντ. Σοβ.	Προσθήκη Δευτ.	Προσθήκη Σοβ.	Μορφολογία Δευτ.	Μορφολογία Σοβ.	Ελ. Λειτ. Λέξη Δευτ.	Ελ. Λειτ. Λέξη Σοβ.	Στίξη Δευτ.	Στίξη Σοβ.	Αμετ. Δευτ.	Αμετ. Σοβ.	Σύν. Λέξ.
T_5	10	0	13	2	57	1	15	0	26	0	12	0	13	0	17	0	0	24	800
L_1	18	0	13	0	174	3	20	1	115	6	29	0	35	0	3	0	0	18	1900
M_1	11	0	24	9	65	6	84	5	91	3	36	0	24	0	25	0	0	36	1792
L_4	18	0	19	1	51	0	7	0	37	0	10	0	12	0	43	1	0	20	1900
Σύν.	**57**	**0**	**69**	**12**	**347**	**10**	**126**	**6**	**269**	**9**	**87**	**0**	**84**	**0**	**88**	**1**	**0**	**98**	
Ποσοστό %																			Συνολικό ποσοστό
T_5	1,25	0,00	1,63	0,25	7,13	0.13	1,88	0,00	3,25	0,00	1,50	0,00	1,63	0,00	2,13	0,00	0,00	3,00	**23,75**
L_1	0,95	0,00	0,68	0,00	9,16	0,16	1,05	0,05	6,05	0,32	1,53	0,00	1,84	0,00	0,16	0,00	0,00	0,95	**22,89**
M_1	0,61	0,00	1,34	0,50	3,63	0,33	4,69	0,28	5,08	0,17	2,01	0,00	1,34	0,00	1,40	0,00	0,00	2,01	**23,38**
L_4	0,95	0,00	1,00	0,05	2,68	0,00	0,37	0,00	1,95	0,00	0,53	0,00	0,63	0,00	2,26	0,05	0,00	1,05	**11,53**
Σύν.	**3,76**	**0,00**	**4,65**	**0,80**	**22,59**	**0,62**	**7,98**	**0,33**	**16,33**	**0,48**	**5,56**	**0,00**	**5,44**	**0,00**	**5,94**	**0,05**	**0,00**	**7,01**	**81,55**

Πίνακας 31: Αποτελέσματα Ταξινόμησης Λαθών ανά επίπεδο σοβαρότητας λάθους στα κείμενα του πειράματος αντιστροφής

Και σε αυτή την ομάδα κειμένων τα σοβαρά λάθη είναι πολύ λιγότερα από τα δευτερεύοντα. Παρατηρούνται επίσης πολλές αμετάφραστες λέξεις, οι οποίες στην αντίστροφη αποτελούν πάντα σοβαρό λάθος λόγω μη μεταγραφής στο λατινικό αλφάβητο. Οι κατηγορίες της Μορφολογίας και της Ελλείπουσας Λειτουργικής Λέξης δεν περιέχουν κανένα σοβαρό λάθος.

Στον Πίνακα 32 παρουσιάζεται ο αριθμός και το ποσοστό λαθών συνολικά ανά κατηγορία και όχι ανάλογα με το επίπεδο σοβαρότητας του λάθους.

	Ελλ. Λέξεις	Υπερβ. Λάθη	Ορολογία	Λεξ. Επιλογή	Σύνταξη	Προσθήκη	Μορφολογία	Ελ. Λειτ. Λέξη	Στίξη	Αμετάφραστο	Σύνολο Λέξεων
T_en_gr_5_HT1_MT4	10	0	15	58	15	26	12	13	17	24	800
L_en_gr_1_HT1_MT4	18	0	13	177	21	121	29	35	3	18	1900
M_en_gr_1_HT3_MT4	11	0	33	71	89	94	36	24	25	36	1792
L_en_gr_4_HT1_MT4	18	0	20	51	7	37	10	12	44	20	1900
σύνολο	57	0	81	357	132	278	87	84	89	98	Συνολικό Ποσοστό
T_en_gr_5_HT1_MT4	1.25%	0.00%	1.88%	7.25%	1.88%	3.25%	1.50%	1.63%	2.13%	3.00%	23.75%
L_en_gr_1_HT1_MT4	0.95%	0.00%	0.68%	9.32%	1.11%	6.37%	1.53%	1.84%	0.16%	0.95%	22.89%
M_en_gr_1_HT3_MT4	0.61%	0.00%	1.84%	3.96%	4.97%	5.25%	2.01%	1.34%	1.40%	2.01%	23.38%
L_en_gr_4_HT1_MT4	0.95%	0.00%	1.05%	2.68%	0.37%	1.95%	0.53%	0.63%	2.32%	1.05%	11.53%
σύνολο	3.76%	0.00%	5.45%	23.21%	8.32%	16.81%	5.56%	5.44%	5.99%	7.01%	

Πίνακας 32: Αποτελέσματα Ταξινόμησης Λαθών στα κείμενα του πειράματος αντιστροφής

Η κατηγορία με τα περισσότερα λάθη είναι η Λεξιλογική Επιλογή και ακολουθεί σε ποσοστό η Προσθήκη, κατηγορία αναμενόμενο να έχει υψηλά ποσοστά στην αντίστροφη μετάφραση, όπως ήδη είδαμε. Η υπόθεσή μας για την κατηγορία της Μορφολογίας επιβεβαιώνεται, καθώς βρίσκεται στις κατηγορίες με τα λιγότερα λάθη.

Στον Πίνακα 33 παρουσιάζονται τα ποσοστά λαθών ανά κατηγορία στις δύο κατευθύνσεις, όπου και παρατηρούνται οι τρεις κατηγορίες λαθών που είχαν μεγάλη διαφορά στις δύο κατευθύνσεις: Μορφολογία, Ελλείπουσα Λειτουργική Λέξη και Αμετάφραστες Λέξεις. Παρατηρείται επίσης σημαντική συνολική βελτίωση στην απόδοση του συστήματος, κάτι το οποίο οφείλεται στο ότι τα κείμενα τα οποία παρουσίαζαν μεταφραστικές δυσκολίες στην αρχική ευθεία μετάφραση (πρωτότυπα κείμενα στα αγγλικά) κατά τη διαδικασία της μετάφρασης από άνθρωπο τροποποιήθηκαν με τέτοιο τρόπο που ως αποτέλεσμα παρουσιάζουν λιγότερες μεταφραστικές δυσκολίες στη δεύτερη διαδικασία μετάφρασης προς τα αγγλικά. Είναι συνηθισμένο φαινόμενο στη μετάφραση να απλοποιούνται πολύπλοκες συντακτικές δομές, να χωρίζονται υπερβολικά μεγάλες προτάσεις και καταβάλλεται εν γένει προσπάθεια το μετάφρασμα να είναι κατανοητό, μέλημα το οποίο δεν απασχολεί απαραίτητα τον συντάκτη οποιουδήποτε πρωτότυπου κειμένου. Το ίδιο φαινόμενο έχει παρατηρηθεί από τον Schwartz, L. [2014] για τις κατευθύνσεις αγγλικά-ρώσικα-αγγλικά. Η σύγκριση παρατίθεται στον Πίνακα 34.

	Ελλ. λέξεις	Υπερβ. λάθη	Ορολ.	Λεξ. επ.	Σύνταξη	Προσθ.	Μορφ.	Ελ. λειτ. λέξη	Στίξη	Αμετάφρ.	Μέσο ποσοστό
EN-GR	7,23%	0,00%	9,03%	29,46%	8,30%	16,68%	41,73%	10,71%	4,44%	2,39%	32,49%
GR-EN	3,76%	0,00%	5,45%	23,21%	8,32%	16,81%	5,56%	5,44%	5,99%	7,01%	20,39%

Πίνακας 33: Ποσοστά λαθών ανά κατηγορία στις δύο κατευθύνσεις για τα κείμενα του πειράματος αντιστροφής

Στον Πίνακα 34 παρουσιάζεται η σύγκριση μεταξύ των δύο κατευθύνσεων σε επίπεδο συνολικού ποσοστού λαθών ανά κείμενο.

	En-gr	Gr-en
L_en_gr_4_MT4	49.72%	11.47%
T_en_gr_5_MT4	26.92%	23.75%
L_en_gr_1_MT4	22.22%	22.74%
M_en_gr_1_MT4	31.05%	20.15%

Πίνακας 34: Συνολικά ποσοστά λαθών στις δύο κατευθύνσεις για τα κείμενα του πειράματος αντιστροφής

Το πρωτότυπο κείμενο L_en_gr_4 στα αγγλικά είναι το κείμενο που παρουσιάζει τις περισσότερες μεταφραστικές δυσκολίες. Κατά τη διαδικασία μετάφρασής του προς τα ελληνικά καθίσταται πιο κατανοητό, με αποτέλεσμα να παρουσιάζει λιγότερες μεταφραστικές δυσκολίες κατά την αντίστροφη μετάφραση, γι' αυτό και παρουσιάζει τη μεγαλύτερη διαφορά από όλα τα κείμενα. Από αυτό το μικρό δείγμα εξάγεται το συμπέρασμα ότι πρέπει να υπάρχει κάποιος βαθμός σύμπτωσης μεταξύ μεταφραστικής δυσκολίας για ανθρώπους και για μηχανές, συμπέρασμα το οποίο θα έπρεπε να επιβεβαιωθεί από μεγαλύτερο όγκο δεδομένων.

5 Συμπεράσματα

Σε αυτό το κεφάλαιο συνοψίζονται τα συμπεράσματα των πειραμάτων της παρούσας εργασίας ως προς πέντε επίπεδα: α) κατάταξη των τεσσάρων συστημάτων, β) επιδόσεις κάθε συστήματος ανά πεδίο, β) φαινόμενα που παρατηρούνται στα εξαγόμενα του Moses ανά κατεύθυνση και προτάσεις βελτίωσής του, γ) σχέση μεταξύ μήκους και ποιότητας μεταφρασμένης πρότασης, και δ) συσχέτιση της αυτόματης μετρικής Bleu με τη μη αυτοματοποιημένη αξιολόγηση ανά κείμενο και ανά πρόταση.

Κατάταξη των τεσσάρων συστημάτων

Όπως φαίνεται στον Πίνακα 23, και στις δύο κατευθύνσεις τα τέσσερα συστήματα κατατάσσονται, βάσει Bleu, ως εξής: 1) Google, 2) Bing, 3) Moses, 4) Systran. Επίσης όλα τα συστήματα λαμβάνουν χαμηλότερη βαθμολογία στην αντίστροφη μετάφραση. Η μη αυτοματοποιημένη αξιολόγηση σε μικρό δείγμα εξαγόμενων των Google, Bing και Systran επιβεβαιώνει τα αποτελέσματα του Bleu σε μεγάλο βαθμό, χωρίς να παρατηρείται όμως απόλυτη συμφωνία. Σύμφωνα με τον Έλεγχο Ποιότητας, 2 από τα 3 κείμενα του Google με ιδιαίτερα υψηλή βαθμολογία Bleu θεωρούνται αρκετά ποιοτικά, και σύμφωνα με την Κατάταξη μεταξύ Bing και Systran, οι συσχετισμοί της αυτόματης μετρικής επιβεβαιώνονται με μεγαλύτερη όμως διαφορά μεταξύ των δύο αυτών συστημάτων.

Επιδόσεις κάθε συστήματος ανά πεδίο

Το Systran και στις δύο κατευθύνσεις λαμβάνει χαμηλότερη βαθμολογία Bleu στα ιατρικά κείμενα από ό,τι στα κείμενα των άλλων πεδίων. Το Bing στην ευθεία λαμβάνει υψηλότερη βαθμολογία στα νομικά και στην αντίστροφη στα τεχνικά. Το Google και το Moses παρουσιάζουν ομοιομορφία ως προς τα αποτελέσματα σε όλα τα πεδία και στις δύο κατευθύνσεις.

Moses: συμπεράσματα ανά κατεύθυνση με προτάσεις βελτίωσης

Σύμφωνα με την Ταξινόμηση Λαθών, τα εξαγόμενα του Moses στην κατεύθυνση αγγλικά-ελληνικά περιέχουν μεγάλο αριθμό μορφολογικών λαθών, τα οποία όμως υπάγονται στην κατηγορία των δευτερευόντων και όχι των σοβαρών λαθών. Αντίθετα, σοβαρά λάθη παρατηρούνται στις κατηγορίες Ορολογία, Λεξιλογική Επιλογή και Σύνταξη. Τα μορφολογικά λάθη επηρεάζουν μεν αρνητικά τη γραμματικότητα του κειμένου, ωστόσο για την επίτευξη του στόχου της μετάφρασης είναι μεγαλύτερης σημασίας η αντιμετώπιση των λεξιλογικών λαθών, ιδιαίτερα αν αυτά είναι σοβαρά. Στην αντιμετώπιση των μορφολογικών λαθών γίνεται αναφορά στο Κεφάλαιο 6. Όσο αφορά τα λεξιλογικά λάθη (Ορολογία και Λεξιλογική Επιλογή), για την αντιμετώπισή τους απαιτείται μεγαλύτερο ή/και καλύτερο παράλληλο σώμα εκπαίδευσης. Τέλος, τα συντακτικά λάθη στην τυπολογία μας αφορούν αυστηρά σειρά λέξεων και όχι άλλα γραμματικά φαινόμενα, συνεπώς αντιμετωπίζονται με ρύθμιση της αναδιάταξης (reordering).

Στην κατεύθυνση ελληνικά-αγγλικά διαφοροποιούνται σε κάποιο βαθμό οι αδυναμίες του Moses, οι οποίες σχετίζονται με τη φύση της γλώσσας-στόχος. Σε αυτή την κατεύθυνση οι κατηγορίες που μας απασχολούν περισσότερο είναι τα λάθη Προσθήκης Λέξης, Σύνταξης, Ορολογίας και Λεξιλογικής Επιλογής, τα οποία παρουσιάζουν μάλιστα αύξηση στην αντίστροφη μετάφραση. Επίσης μας απασχολεί η κατηγορία της Ελλείπουσας Λειτουργικής Λέξης, στην οποία υπάρχει μικρή αύξηση στην αντίστροφη κατεύθυνση. Η σημαντικότερη ωστόσο παρατήρηση στην αντίστροφη μετάφραση είναι το πλήθος αμετάφραστων λέξεων.

Σχέση μεταξύ μήκους και ποιότητας μεταφρασμένης πρότασης

Όσο αφορά την ευθεία μετάφραση, σύμφωνα με τον Έλεγχο Ποιότητας, ο μέσος όρος λέξεων των αποδεκτών προτάσεων είναι 15,94 λέξεις, ενώ ο μέσος όρος λέξεων των προτάσεων που απαιτούν μεγαλύτερη προσπάθεια κατά τη διόρθωση είναι 30,18. Σύμφωνα με την Ταξινόμηση Λαθών, ενώ δεν υπάρχει διαφορά σε επίπεδο παρουσίας λαθών γενικά, τα σοβαρά λάθη εμφανίζονται σε μέτριες και μεγάλες προτάσεις και το πλήθος τους αυξάνεται όσο ανεβαίνει ο αριθμός λέξεων της πρότασης. Αυτές οι παρατηρήσεις αποτελούν ένδειξη του ότι όσο μεγαλύτερη είναι η πρόταση τόσο μειώνεται η ποιότητα της μετάφρασής της. Αξίζει να σημειωθεί ότι ο αριθμός σοβαρών συντακτικών λαθών δεν είναι τέτοιος που να επιτρέπει την απόδοση αυτού του φαινομένου στην αύξηση της συντακτικής πολυπλοκότητας. Στην αντίστροφη μετάφραση τα δεδομένα δεν επιτρέπουν την εξαγωγή συμπερασμάτων για τη σχέση μήκους και ποιότητας μεταφρασμένης πρότασης.

Συσχέτιση της αυτόματης μετρικής Bleu με τη μη αυτοματοποιημένη αξιολόγηση ανά κείμενο και ανά πρόταση

Ως προς το μικρό επισημειωμένο δείγμα εξαγόμενων των Google, Bing και Systran, τα αποτελέσματα του Bleu σε επίπεδο κειμένου επιβεβαιώνονται σε μεγάλο βαθμό, χωρίς να παρατηρείται απόλυτη συμφωνία. Για το δείγμα εξαγόμενων του Google, σε επίπεδο πρότασης, υπάρχει υψηλή συσχέτιση στις προτάσεις που κρίνονται ποιοτικές, όσο αφορά όμως τις προτάσεις για τις οποίες κρίθηκε ότι απαιτείται μεγαλύτερη προσπάθεια διόρθωσης, πρόκειται για ένα ποιοτικό κριτήριο, το οποίο δεν μπορεί να συγκριθεί με τα ποσοτικά δεδομένα του Bleu. Στις προτάσεις που αξιολογήθηκαν στην Κατάταξη η συμφωνία είναι της τάξης του 61,77%.

Ως προς το δείγμα εξαγόμενων του Moses, σε επίπεδο κειμένων υπάρχει ικανοποιητική συσχέτιση και στις δύο κατευθύνσεις. Η συσχέτιση εξάγεται από το γεγονός ότι τα κείμενα με τις χαμηλότερες βαθμολογίες Bleu περιέχουν μεγαλύτερο ποσοστό λαθών και αντίστροφα.

Για το ίδιο δείγμα σε επίπεδο πρότασης, στην κατεύθυνση αγγλικά-ελληνικά υπάρχει υψηλή συσχέτιση μεταξύ βαθμολογίας Bleu και μη αυτοματοποιημένης αξιολόγησης όσο αφορά τις υψηλές βαθμολογίες Bleu (υψηλή βαθμολογία Bleu-απουσία σοβαρών λαθών). Αυτού του είδους η συσχέτιση δεν αποδίδεται σε κάποιο χαρακτηριστικό της μετρικής, καθώς αυτή δεν λαμβάνει υπόψη κανένα είδος επιπέδου σοβαρότητας λάθους· οι βαθμολογίες της δηλώνουν τον αριθμό ν-γραμμάτων που δεν εμφανίζονται στις μεταφράσεις-αναφορά αλλά καμία πληροφορία σχετικά με το ν-γραμμα που δεν συμπίπτει. Όσο αφορά τις χαμηλές βαθμολογίες Bleu, η συσχέτιση είναι χαμηλότερη και φαίνεται ότι η μετρική ναι μεν αντικατοπτρίζει την ποιοτική μετάφραση σε αρκετά ικανοποιητικό βαθμό, αλλά μπορεί να αποδώσει χαμηλή βαθμολογία και σε μεταφράσεις οι οποίες δεν περιέχουν πολλά ή/και σοβαρά λάθη.

Ως προς τα εξαγόμενα του ίδιου συστήματος στην αντίστροφη κατεύθυνση, υπάρχει υψηλότερη συσχέτιση στις προτάσεις με χαμηλή βαθμολογία και χαμηλότερη στις προτάσεις με υψηλή βαθμολογία Bleu.

Βάσει των ανωτέρω συμπερασμάτων, θεωρούμε ότι η χρήση του Bleu κατά την ανάπτυξη συστήματος για συνεχείς δοκιμές είναι αποδοτική, καθώς μπορεί να αποτελέσει μια αρκετά έγκυρη ένδειξη της ποιότητας της μετάφρασης για αυτόν τον στόχο. Όταν η έρευνα απαιτεί λεπτές διακρίσεις και ποιοτικές κρίσεις εξακολουθεί να είναι απαραίτητη η μη αυτοματοποιημένη αξιολόγηση.

6 Μελλοντική εργασία

Κατά τη μη αυτοματοποιημένη αξιολόγηση της παρούσας εργασίας παράχθηκαν 331 μετεπεξεργασμένες προτάσεις, 229 στην κατεύθυνση αγγλικά-ελληνικά και 102 στην αντίστροφη. Παράλληλα δημιουργήθηκαν αρχεία με τα ταξινομημένα λάθη των προτάσεων που αξιολογήθηκαν. Συνολικά εντοπίστηκαν και διορθώθηκαν 2.158 λάθη, 1.434 στην ευθεία μετάφραση και 724 στην αντίστροφη. Η πληροφορία του είδους λάθους σε συνδυασμό με την πληροφορία της μετατροπής από τη λανθασμένη μορφή στη μετεπεξεργασμένη μπορούν να χρησιμοποιηθούν για τη δημιουργία ενός συστήματος αυτόματης μετεπεξεργασίας, αξιοποιώντας τη μελέτη που έχει αρχίσει να αναπτύσσεται τα τελευταία χρόνια στο αντικείμενο από τους Rudolf, R. [2014], Parton, K. κ.ά. [2012] και Mohaghegh, M. κ.ά. [2013]. Για την ανάπτυξη ενός τέτοιου συστήματος προτείνουμε την επιλογή της κατηγορίας λαθών Μορφολογία στην κατεύθυνση αγγλικά-ελληνικά, για την οποία διαθέτουμε 902 διορθωμένα λάθη από το σώμα μας, και την αυτόματη εξαγωγή κανόνων μετατροπής από το υπάρχον σώμα προτάσεων. Με αυτόν τον τρόπο μπορεί να βελτιωθεί σημαντικά η γραμματικότητα των μεταφρασμάτων στην ελληνική.

Παράρτημα

Γλωσσάρι ελληνικά-αγγλικά

Μη αυτοματοποιημένη αξιολόγηση	Human/manual evaluation
Ακρίβεια	Precision
Αμφίδρομο σώμα κειμένων	Reciprocal corpus
Αναδιάταξη	Reordering
Ανάκληση	Recall
Ανάκτηση πληροφοριών	Information retrieval
Αξιολόγηση μέσω εκτέλεσης εργασίας	Task-based evaluation
Αποκωδικοποιητής	Decoder
Απόσταση διόρθωσης	Edit distance
Απόσταση κατά Levenshtein	Levenshtein distance
Αρμονικός μέσος όρος ακρίβειας και ανάκλησης	F-measure
Αυτόματη αξιολόγηση	Automatic evaluation
Βαθμολογία	Score
Βασιζόμενος σε κανόνες/που βασίζεται σε κανόνες/χρήση κανόνων	Rule-based
Γλωσσικό μητρώο	Register
Γλωσσικό μοντέλο	Language model
Γραμματικότητα	Fluency
Δείκτης σφαλμάτων λέξης	Word error rate
Διαγλώσσα	Interlingua
Εκτίμηση ποιότητας	Confidence/quality estimation
Ελλείπουσα λέξη	Missing word
Ενέργεια διόρθωσης	Edit
Εξαγόμενο	Output
Επιβλεπόμενη μηχανική μάθηση	Supervised machine learning
Επίλυση συναναφοράς	Coreference resolution
Επισημείωση	Annotation
Ετικέτα	Label
Ημι-αυτόματη αξιολόγηση	Semi-automated evaluation
Ισοδυναμία	Equivalence
Κατά λέξη μετάφραση	Literal translation
Κατάταξη	Ranking
Κειμενικό είδος	Genre
Κωδικοποίηση	Encoding

Λανθασμένη μετάφραση	Mistranslation
Λειτουργική λέξη	Function word
Λημματοποίηση	Lemmatisation
Μετάφραση-αναφορά	Reference translation
Μεταφραστικό μοντέλο	Translation model
Μετεπεξεργασία	Post-editing
Μετρική	Metric/metrics
Μετρική με χρήση μεταφράσεων-αναφορά/ μεταφράσεις-αναφορά	Reference-based metric
Μη γραμματικός	Disfluent
Μη επιβλεπόμενη μηχανική μάθηση	Unsupervised machine learning
Μήκος βέλτιστης σύμπτωσης	Best match length
Μηχανική/αυτόματη μετάφραση	Machine translation
Ν-γραμμα	N-gram
Ονοματισμένη οντότητα	Named entity
Παλινδρόμηση	Regression
Παραγωγή φυσικής γλώσσας	Natural language generation
Πιθανό λάθος	Issue
Πιστότητα	Accuracy
Ποινή συντομίας	Brevity penalty
Πολυλεκτικό σύνολο	Multiword expression
Προσεγγίσεις άμεσης μετάφρασης	Direct translation approaches
Προσεγγίσεις βασιζόμενες σε εμπειρική γνώση	Data-driven approaches
Ρύθμιση των παραμέτρων του συστήματος	Tuning
Στελέχωση	Stemming
Στοίχιση	Alignment
Συμβολοσειρά	String
Συναναφορά	Coreference
Συνδυασμός γλώσσας με γεωγραφική τοποθεσία	Locale
Συνεκτικότητα	Coherence
Συνοχή	Cohesion
Συστατικό	Constituent
Συστήματα με χρήση παραδειγμάτων	Example-based systems
Συστήματα μεταφοράς	Transfer systems
Σώμα εκπαίδευσης	Training corpus
Σώμα κειμένων	Corpus
Ταξινόμηση	Classification

Τρίγωνο του Vauqois	Vauqois triangle
Τροποποιημένη ακρίβεια	Modified precision
Υπερμετάφραση	Overtranslation
Υπομετάφραση	Undertranslation
Ύφος	Style
Ψευδόφιλη λέξη/μονάδα	False friend

Γλωσσάρι αγγλικά-ελληνικά

Accuracy	Πιστότητα
Alignment	Στοίχιση
Annotation	Επισημείωση
Automatic evaluation	Αυτόματη αξιολόγηση
Best match length	Μήκος βέλτιστης σύμπτωσης
Brevity penalty	Ποινή συντομίας
Classification	Ταξινόμηση
Coherence	Συνεκτικότητα
Cohesion	Συνοχή
Confidence/quality estimation	Εκτίμηση ποιότητας
Constituent	Συστατικό
Coreference	Συναναφορά
Coreference resolution	Επίλυση συναναφοράς
Corpus	Σώμα κειμένων
Data-driven approaches	Προσεγγίσεις βασιζόμενες σε εμπειρική γνώση
Decoder	Αποκωδικοποιητής
Direct translation approaches	Προσεγγίσεις άμεσης μετάφρασης
Disfluent	Μη γραμματικός
Edit	Ενέργεια διόρθωσης
Edit distance	Απόσταση διόρθωσης
Encoding	Κωδικοποίηση
Equivalence	Ισοδυναμία
Example-based systems	Συστήματα με χρήση παραδειγμάτων
F-measure	Αρμονικός μέσος όρος ακρίβειας και ανάκλησης
False friend	Ψευδόφιλη λέξη/μονάδα
Fluency	Γραμματικότητα
Function word	Λειτουργική λέξη

Genre	Κειμενικό είδος
Human/manual evaluation	μη αυτοματοποιημένη αξιολόγηση
Information retrieval	Ανάκτηση πληροφοριών
Interlingua	Διαγλώσσα
Issue	Πιθανό λάθος
Label	Ετικέτα
Language model	Γλωσσικό μοντέλο
Lemmatisation	Λημματοποίηση
Levenshtein distance	Απόσταση κατά Levenshtein
Literal translation	Κατά λέξη μετάφραση
Locale	Συνδυασμός γλώσσας με γεωγραφική τοποθεσία
Machine translation	Μηχανική/αυτόματη μετάφραση
Metric/metrics	Μετρική
Missing word	Ελλείπουσα λέξη
Mistranslation	Λανθασμένη μετάφραση
Modified precision	Τροποποιημένη ακρίβεια
Multiword expression	Πολυλεκτικό σύνολο
N-gram	Ν-γραμμα
Named entity	Ονοματισμένη οντότητα
Natural language generation	Παραγωγή φυσικής γλώσσας
Output	Εξαγόμενο
Overtranslation	Υπερμετάφραση
Post-editing	Μετεπεξεργασία
Precision	Ακρίβεια
Ranking	Κατάταξη
Recall	Ανάκληση
Reciprocal corpus	Αμφίδρομο σώμα κειμένων
Reference translation	Μετάφραση-αναφορά
Reference-based metric	Μετρική με χρήση μεταφράσεων-αναφορά/ μεταφράσεις-αναφορά
Register	Γλωσσικό μητρώο
Regression	Παλινδρόμηση
Reordering	Αναδιάταξη
Rule-based	Βασιζόμενος σε κανόνες/που βασίζεται σε κανόνες/χρήση κανόνων
Score	Βαθμολογία
Semi-automated evaluation	Ημι-αυτόματη αξιολόγηση
Stemming	Στελέχωση

String	Συμβολοσειρά
Style	Ύφος
Supervised machine learning	Επιβλεπόμενη μηχανική μάθηση
Task-based evaluation	Αξιολόγηση μέσω εκτέλεσης εργασίας
Training corpus	Σώμα εκπαίδευσης
Transfer systems	Συστήματα μεταφοράς
Translation model	Μεταφραστικό μοντέλο
Tuning	Ρύθμιση των παραμέτρων του συστήματος
Undertranslation	Υπομετάφραση
Unsupervised machine learning	Μη επιβλεπόμενη μηχανική μάθηση
Vauqois triangle	Τρίγωνο του vauqois
Word error rate	Δείκτης σφαλμάτων λέξης

Βιβλιογραφία

Ξενόγλωσση

Akiba, Y., Imamura, K., and Sumita, E. (2001). “Using multiple edit distances to automatically rank machine translation output”. In: *Proceedings of MT Summit VIII of the European Association for Machine Translation.* Santiago de Compostela, Spain.

Amigo, E., Giménez, J., Gonzalo, J., and Màrquez, L. (2006). “MT evaluation: Human-like vs. human acceptable”. In: *Proceedings of COLING-ACL06, joint conference of the International Committee on Computational Linguistics and the Association for Computational Linguistics.* Sydney, Australia.

Babych, B. and Hartley, A. (2004). “Extending BLEU MT evaluation method with frequency weighting”. In: *Proceedings of ACL (Association for Computational Linguistics).* Barcelona, Spain.

Bakhtin, M. (1986). Speech Genres and Other Late Essays, Tr. Vern W. McGee, Eds. Caryl Emerson and Michael Holquist. University of Texas Press, Austin.

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.

Blatz, J., Fitzgerald, E., Foster, G., Gandraburn S., Goutte, C., Kulesza, A., Sanchis, A. and Ueffing, N. (2004). “Confidence estimation for machine translation”. In: *Proceedings of JHU / CLSP Summer Workshop.*

Bojar, O. (2011). “Analyzing Error Types in English-Czech Machine Translation”. In: *Prague Bulletin of Mathematical Linguistics*, 95:63–76. ISSN 0032-6585.

Bonnie, D., Przybocki, M., Snover, M., Le, A., Sanders, G., Bronsart, S., Strassel, S. and Glenn, M . (2011). “Chapter 5.1.1 Historical Background”. In: “Part 5: Machine Translation Evaluation”, Ed. Dorr, B.. In: *Handbook of Natural Language Processing and Machine Translation*, Eds. Olive, J., McCary, J. and Christianson, C.. DARPA Global Autonomous Language Exploitation.

Bonnie, D., Snover, M. and Madnani, N. (2011). “Chapter 5.1 Introduction”. In: “Part 5: Machine Translation Evaluation”, Ed. Dorr, B.. In: *Handbook of Natural Language Processing and Machine Translation*, Eds. Olive, J., McCary, J. and Christianson, C.. DARPA Global Autonomous Language Exploitation.

Burchardt, A., Gaspari, F., Harris, K., Lommel A., Popović, M., Toral A. and Uszkoreit, H. (2014). *Barriers for High-Quality Machine Translation*, deliverable D 1.3.1, QTLaunchpad.

Cabré, T. (1999). *Terminology: theory, methods and applications*, Amsterdam: John Benjamins.

Callison-Burch, Ch., Fordyce, C., Kohen, Ph., Monz, Ch. and Schroeder, J. (2007). "(Meta-) evaluation of machine translation". In: *Proceedings of the Second Workshop on Statistical Machine Translation '07*. Prague, Czech Republic, pp. 136-158.

Callison-Burch, Ch., Osborne, M., Koehn, P. (2006). "Re-evaluating the role of BLEU in machine translation research". In: *Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics (EACL)*. Trento, Italy, pp. 249-256.

Coughlin, D. (2003). "Correlating automated and human assessments of machine translation quality". In: *Proceedings of MT Summit IX*. New Orleans, LA, USA, pp. 63–70.

Crystal, D. (2008). *A Dictionary of Linguistics and Phonetics*. 6th edition. Blackwell Publishing.

Cui, L., Zhang, D., Li, M. and Zhou, M. (2011). "Function word generation in statistical machine translation systems". In: *Proceedings of Machine Translation Summit XIII*. Xiamen, China.

De Beaugrande, R.-A., Dressler, W. (1981). "Introduction to Text Linguistics". In: *XIV Congress of Linguists*. Berlin, Original 1981, digitally reformatted 2002.

Doddington, G. (2002). "Automatic Evaluation of Machine Translation Quality using N-gram Co-occurrence Statistics". In: *Proceedings of the 2nd Human Language Technologies Conference (HLT-02)*. San Diego, CA, USA, pp. 128-132.

Dreyer, M. and Marcu, D. (2012). "HyTER: Meaning-Equivalent Semantics for Translation Evaluation". In: *Proceedings of Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Montréal, Canada, pp. 162–171.

España i Bonet, C. (2010). *Statistical Machine Translation. A practical tutorial*. MOLTO Kickoff Meeting, UPC, Barcelona, Spain.

Federmann, Ch. (2010). "Appraise: An Open-Source Toolkit for Manual Phrase-Based Evaluation of Translations". In: *Proceedings of the Seventh Conference on International Language Resources and Evaluation (LREC)*. Valletta, Malta.

Fernandes, L. (2006). "Corpora in Translation Studies: revisiting Baker's typology". In: *Fragmentos*, No 30, 087-095 Florianópolis.

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.

Giménez, J. and Màrquez, L. (2007). "Linguistic features for automatic evaluation of heterogenous MT systems". In: *Proceedings of the Second Workshop on Statistical Machine Translation '07.* Prague, Czech Republic.

Görög, A. (2014). "Quantifying and benchmarking quality: the TAUS Dynamic Quality Framework". In: *Revista Tradumàtica: tecnologies de la traducció, No 12, Traducció i qualitat*, 12/2014, ISSN: 1578-7559, http://revistes.uab.cat/tradumatica.

Han, A.L.F., Wong, D.F., and Chao, L.S. (2012). "LEPOR: A Robust Evaluation Metric for Machine Translation with Augmented Factors". In: *Proceedings of the 24th International Conference on Computational Linguistics (COLING 2012): Posters.* Mumbai, India, pp. 441–450.

Kanelli, O. (2010). "Notions of equivalence-Early interest among translation theorists". In: *Poliphilos*, Vol. 1 (2010), pp. 130-138.

Koby, Geoffrey S. and Alan K. Melby (2013). "Certification and Job Task Analysis (JTA): Establishing Validity of Translator Certification Examinations.". In: *The International Journal of Translation and Interpreting Research* (5)1, 174-210.

Koehn, Ph. (2006). *Statistical Machine Translation: the basic, the novel, and the speculative. SMT Tutorial.* The University of Edinburgh.

Koehn, Ph. (2010). *Statistical Machine Translation*. Cambridge University Press. http://www.statmt.org/book/ (04/11/2014).

Koehn, Ph., Hoang, H., Birch, A., Callison-Burch, Ch., Federico, M., Bertoldi, N., Cowan, B., Shen, W., Moran, Ch., Zens, R., Dyer, Ch., Bojar, O., Constantin, A., Herbst, E. (2007). "Moses: Open Source Toolkit for Statistical Machine Translation". In: *Annual Meeting of the Association for Computational Linguistics (ACL) demonstration session.* Prague, Czech Republic.

Koehn, Ph. and Monz, C. (2006). "Manual and automatic evaluation of machine translation between european languages". In: *Proceedings of the 2006 Workshop on Statistical Machine Translation.*

Lacruz, I., Denkowski, M. and Lavie, A. (2014). "Cognitive Demand and Cognitive Effort in Post-Editing". In: *Proceedings of the Third Workshop on Post-Editing Technology and Practice.* Eds. O'Brien, Sh., Simard, M. and Specia L., 11th Conference of the Association for Machine Translation in the Americas, Vancouver, BC, Canada.

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

Lawson, A. (2001). "Collecting, aligning and analysing parallel corpora". In: *Small Corpus*

Studies and ELT: Theory and Practice. Eds. Ghadessy, M., Henry, A. and Roseberry, R.L. Amsterdam: John Benjamins.

Lerat, P. 1995. *Les langues spécialisées*. Paris: PUF.

Leusch, G., Ueffing, N., and Ney, H. (2003). "A novel string-to-string distance measure with applications to machine translation evaluation". In: *Proceedings of MT Summit IX*. New Orleans, LA, USA.

Leusch, G., Ueffing, N., and Ney, H. (2006). "CDER: Efficient MT evaluation using block movements". In: *Proceedings of the 11th Conference of the European Chapter of the Association for Computational Linguistics*.

Levenshtein, V. (1966). "Binary codes capable of correcting deletions, insertions, and reversals". In: *Soviet Physics - Doklady* 10 (8): 707–710. Original in Russian 1965.

Lin, Ch. Y. and Och, F. J. (2004). "Automatic Evaluation of Machine Translation Quality Using Longest Common Subsequence and Skip-Bigram Statistics". In: *Proceedings of the 42nd Meeting of the Association for Computational Linguistics (ACL'04)*, Main Volume. Barcelona, Spain, pp. 605-612.

Lita, L. V., Rogati, M., and Lavie, A. (2005). "BLANC: Learning Evaluation Metrics for MT". In: *Proceedings of Human Language Technology Conference and Conference on Empirical Methods in Natural Language Processing (HLT/EMNLP)*. Vancouver, pp. 740–747.

Manning, C. D., Raghavan, P., Schütze, H. (2008). *Introduction to Information Retrieval, Cambridge University Press*. http://informationretrieval.org/ (τελευταία προσπέλαση 11/11/2014).

Melamed, I., Green, R., and Turian, J. (2003). "Precision and recall of machine translation". In: *Proceedings of the HLT-NAACL 2003*. Edmonton, Canada.

Mohaghegh, M., Sarrafzadeh, A. and Mohammadi, M. (2013). "A Three-Layer Architecture for Automatic Post-Editing System Using Rule-Based Paradigm". In: *Proceedings of the 4th Workshop on South and Southeast Asian NLP (WSSANLP), International Joint Conference on Natural Language Processing*. Nagoya, Japan, pp. 17-24.

Nãdejde, M. (2011). *Syntactic Word Reordering for Statistical Machine Translation*. Master's thesis, Universität des Saarlandes.

Navarro, G. (2001). "A Guided Tour to Approximate String Matching". In: *ACM Computing Surveys* 33 (1): 31–88.

Newmark, P. (1988). *A Textbook of Translation*. Essex: Pearson Education Limited.

Niessen, S., Och, F., Leusch, G. and 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*.

Papineni, K., Roukos, S., Ward, T., Zhu, W. J. (2002). "BLEU: a method for automatic evaluation of machine translation". In: *ACL-2002: 40th Annual meeting of the Association for Computational Linguistics*. CiteSeerX: 10.1.1.19.9416, pp. 311–318.

Parton, K., Habash, N., McKeown, K., Iglesias, G. and De Gispert, A. (2012). "Can Automatic Post-Editing Make MT More Meaningful?". In: *Proceedings of European Association for Machine Translation*.

Popović, M., Ney, H. (2007). "Word Error Rates: Decomposition over POS Classes and Applications for Error Analysis". In: *Proceedings of the Second Workshop on Statistical Machine Translation, Association for Computational Linguistics*. Prague, pp. 48–55.

Przybocki, M., Le, A., Sanders, G., Bronsart, S., Strassel, S. and Glenn, M. (2011). "Chapter 5.4.3 Post-editing". In: "Part 5: Machine Translation Evaluation", Ed. Dorr, B.. In: *Handbook of Natural Language Processing and Machine Translation*, Eds. Olive, J., McCary, J. and Christianson, C., DARPA Global Autonomous Language Exploitation.

Przybocki, M., Peterson, K., Bronsart, S. and Sanders, G. (2009). *The NIST 2008 Metrics for Machine Translation Challenge – Overview, Methodology, Metrics, and Results*. Multimodal Information Group, National Institute of Standards and Technology, Gaithersburg MD, U.S.A.

Rieu, E. V. (1953). "Translation". In: Cassell's Encyclopedia of Literature, Vol. 1. Cassell, London.

Rudolf, R. (2014). "Depfix, a Tool for Automatic Rule-based Post-editing of SMT". In: *The Prague Bulletin of Mathematical Linguistics,* No. 102, pp. 47–56.

Sag, I. A., Baldwin, T., Bond, F., Copestake, A., Flickinger, D. (2002). "Multiword Expressions: A Pain in the Neck for NLP". In: *Computational Linguistics and Intelligent Text Processing. Lecture Notes in Computer Science*, Vol. 2276, Springer, pp. 1-15.

Sánchez-Gijón, P. and Torres-Hostench, O. (2014). "MT Post-editing into the mother tongue or into a foreign language? Spanish-to-English MT translation output post-edited by translation trainees". In: *Proceedings of the Third Workshop on Post-Editing Technology and Practice*. Eds. O'Brien, Sh., Simard, M. and Specia L.. 11th Conference of the Association for Machine Translation in the Americas, Vancouver, BC, Canada.

Sanders, G., Przybocki, M., Madnani, N. and Snover M. (2011). "Chapter 5.1.2 Human Subjective Judgments". In: "Part 5: Machine Translation Evaluation", Ed. Dorr, B.. In: *Handbook of Natural Language Processing and Machine Translation*, Eds. Olive, J., McCary, J. and Christianson, C.. DARPA Global Autonomous Language Exploitation.

Schwartz, L. (2014). “Monolingual Post-Editing by a Domain Expert is Highly Effective for Translation Triage”. In: *Proceedings of the Third Workshop on Post-Editing Technology and Practice*, Eds. O'Brien, Sh., Simard, M. and Specia L., 11th Conference of the Association for Machine Translation in the Americas, Vancouver, BC, Canada.

Silva, J., Coheur, L., Costa, Â. and Trancoso, I. (2012). “Dealing with unknown words in statistical machine translation”. In: *Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC'12), European Language Resources Association (ELRA)*. ISBN 978-2-9517408-7-7.

Sinclair, J. (1991). *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

Sinclair, J. (1996). *EAGLES Preliminary Recommendations on Corpus Typology.* EAGLES Document EAG-TCWG-CTYP/P (Version of May).

Snover, M., Dorr, B., Schwartz, R., Micciulla, L. and Makhoul, J. (2006). “A study of translation edit rate with targeted human annotation”. In: *Proceedings of the 7th Conference of the Association for Machine Translation in the Americas*. Boston Marriott, Cambridge, Massachusetts, USA.

Specia, L., Raj, D., Turchi, M. (2010). *Machine translation evaluation versus quality estimation*. Springer Science+Business Media B.V., Mach Translat (2010) 24:39–50 DOI 10.1007/s10590-010-9077-2.

Specia, L., Shah, K., De Souza, J. G. C. and Cohn, T. (2013). “QuEst - A translation quality estimation framework”. In: *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics*. Sofia, Bulgaria, pp. 79–84.

Steinberger, R., Ebrahim, M., Poulis, A., Carrasco-Benitez, M., Schlüter, P., Przybyszewski, M., Gilbro, S. (2014). “An overview of the European Union's highly multilingual parallel corpora”. In: *Lan Resources & Evaluation*. DOI 10.1007/s10579-014-9277-0. European Union. Springer

Teubert, W. (1996). “Comparable or Parallel Corpora?”. In: *International Journal of Lexicography* 9 (3).

Tiedemann, J. (2009). “News from OPUS - A Collection of Multilingual Parallel Corpora with Tools and Interfaces". In: *Recent Advances in Natural Language Processing*, Eds. N. Nicolov and K. Bontcheva and G. Angelova and R. Mitkov, Vol. V. Amsterdam/Philadelphia: John Benjamins, pp. 237-248.

Tomás, J., Mas, J. A., and Casacuberta, F. (2003). “A quantitative method for machine translation evaluation”. In: *Proceedings of the EACL 2003 Workshop on Evaluation Initiatives in Natural Language Processing: are evaluation methods, metrics and resources reusable?*

Tytler, A. F. (1790). *Essay on the principles of translation*. Dent, London, 1912.

Vermeer, H. J. (2000). “Skopos and Commission in Translational Action”. In: *The Translation Studies Reader. Ed.* Lawrence Venuti. London. New York: Routledge, pp. 221–32.

Vinay, J. P. and Darbelnet, J. (1958). *Stylistique comparée du français et de l’anglais. Méthode de Traduction*. Nouvelle édition revue et corrigée, [Paris]: Didier 1977.

Widmer, F. (1959). *Fug und Unfug des Übersetzens*. Kiepenhauer & Witsch, Cologne and Berlin.

Zhou, L., Lin, C.-Y., and Hovy, E. (2006). “Re-evaluating machine translation results with paraphrase support”. In: *Proceedings of the 2006 Conference on Empirical Methods in Natural Language Processing (EMNLP)*.

Ελληνόγλωσση

Αναστασιάδη-Συμεωνίδη, Α. και Βλέτση, Ε. (2009). «Ψευδόφιλες μονάδες: οι διαχρονικοί άσπονδοι φίλοι στην ελληνική» στα Πρακτικά του Πανελλήνιου Συνέδριου «Η Διδασκαλία της Ελληνικής Γλώσσας (ως πρώτης/μητρικής, δεύτερης/ξένης)», Νυμφαίο Φλώρινας.

Βώρος, Φ. Κ. (2004). Οι αξιολογικές κρίσεις. Η χρήση και η σημασία τους στην εκπαίδευση, αναδημοσίευση από το Δοκίμια για την Παιδεία, Αθήνα, http://www.voros.gr/anadimosieuseis/anadimosieuseiseis05.html (29/04/2015), σελ. 194-203.

Γαβριηλίδου, Μ., Κουτσομπόγερα, Μ., Πατρικάκος Α. και Πιπερίδης, Σ. (2012). Η ελληνική γλώσσα στην ψηφιακή εποχή. Σειρά Λευκών Βίβλων. Επιμελητές: Uszkoreit, H., Rehm, G. META-NET. Springer.

Γεωργακοπούλου, Α. και Γούτσος, Δ. (1999). Κείμενο και Επικοινωνία. Αθήνα: Ελληνικά Γράμματα.

Κελάνδριας, Π. (2013). “Η ορολογική σύγκλιση ως μεταφραστικό εργαλείο της ειδικής μετάφρασης” στις Ανακοινώσεις του 9ου Συνεδρίου «Ελληνική Γλώσσα και Ορολογία», Ελληνική Εταιρεία Ορολογίας, Αθήνα.

Λάιονς, Τ. (2006). Γλωσσολογική σημασιολογία, μετάφραση Γιώργου Ανδρουλάκη, Αθήνα: Πατάκης.

Μπαμπινιώτης, Γ. (1998). Θεωρητική Γλωσσολογία. Εισαγωγή στη Σύγχρονη Γλωσσολογία, Αθήνα.

Μπατσαλιά, Φρ., Σελλά-Μάζη, Ε. (2004). Γλωσσολογική προσέγγιση στη θεωρία και τη διδακτική της μετάφρασης, Αθήνα: Εκδόσεις «Έλλην».

Σοφιανόπουλος, Σ. (2009). Μοντελοποίηση Γλώσσας για συστήματα Μηχανικής Μετάφρασης με μονόγλωσσο σώμα κειμένων. Διδακτορική Διατριβή, Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών ΕΜΠ.

Φωτοπούλου, Α. (2003). "Πολυλεκτικοί συνδυασμοί και ορολογία" στα Πρακτικά του 4ου συνεδρίου «Ελληνική γλώσσα και ορολογία» της ΕΛΕΤΟ, Αθήνα.

Ricoeur, P. (2010). Για τη μετάφραση, Μετάφραση: Αυγουστής, Γ., Πρωτότυπο: Sur la traduction (2008), Αθήνα: Εκδόσεις Πατάκη.

Η ολοκλήρωση της παρούσας διπλωματικής εργασίας συγχρηματοδοτήθηκε μέσω του Έργου «Υποτροφίες ΙΚΥ» από πόρους του ΕΠ «Εκπαίδευση και Δια Βίου Μάθηση», του Ευρωπαϊκού Κοινωνικού Ταμείου (ΕΚΤ) του ΕΣΠΑ, 2007-2013.