

ΙΟΝΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ
ΣΧΟΛΗ ΙΣΤΟΡΙΑΣ ΚΑΙ ΜΕΤΑΦΡΑΣΗΣ-ΔΙΕΡΜΗΝΕΙΑΣ
ΤΜΗΜΑ ΞΕΝΩΝ ΓΛΩΣΣΩΝ, ΜΕΤΑΦΡΑΣΗΣ ΚΑΙ ΔΙΕΡΜΗΝΕΙΑΣ
ΠΜΣ «Επιστήμη της Μετάφρασης»



Μηχανική Μετάφραση και Μέσα Κοινωνικής Δικτύωσης:
Συγκριτική αξιολόγηση μηχανικά μεταφρασμένων
κειμένων με χρήση αυτόματων μετρικών

Μεταπτυχιακή Διπλωματική Εργασία

Συντάκτης: Κωνσταντίνος Δαγκλής ΑΜ: Ξ2019308

Επιβλέπουσα Καθηγήτρια: Κ. Κερμανίδου

Μέλη Επιτροπής: Β. Σωσώνη, Ε. Μαραγκουδάκης

Κέρκυρα, Μάρτιος 2021

Περίληψη

Αντικείμενο της παρούσας εργασίας είναι η μελέτη της σχέσης μεταξύ της μηχανικής μετάφρασης και των μέσων κοινωνικής δικτύωσης σε επίπεδο γλώσσας και θεματολογίας μέσα από τη συγκριτική αξιολόγηση μηχανικά μεταφρασμένων κειμένων. Για τον σκοπό αυτό επιλέχθηκαν κάποια κείμενα τα οποία μεταφράστηκαν με τη βοήθεια μηχανών αυτόματης μετάφρασης και στη συνέχεια αξιολογήθηκαν τόσο με αυτόματες όσο και με μη αυτόματες μεθόδους αξιολόγησης. Βασική προϋπόθεση για την επιλογή των κειμένων ήταν το να προέρχονται από τον κοινωνικό ιστό ή να αναφέρονται σε θεματολογία δημοφιλή στον κοινωνικό ιστό αλλά και να ανήκουν ταυτόχρονα σε παράλληλα σώματα κειμένων ώστε να μπορούν να αξιολογηθούν με τη βοήθεια μετρικών αυτόματης αξιολόγησης.

Εξετάστηκαν κείμενα για τα ζεύγη γλωσσών γερμανικά-αγγλικά, ιταλικά-αγγλικά και ιταλικά-ελληνικά ενώ για τη μηχανική μετάφραση χρησιμοποιήθηκαν το μεταφραστικό εργαλείο Google Translate και η μηχανή αυτόματης μετάφρασης του Facebook. Για τον υπολογισμό των μετρικών αυτόματης αξιολόγησης χρησιμοποιήθηκε το εργαλείο MTUOC (και πιο συγκεκριμένα το MTUOC-Eval) ενώ ειδικά για τα κείμενα με γλώσσα-στόχο τα ελληνικά, προτείνεται, μέσω της παρούσας εργασίας, η χρήση μιας παραλλαγής του γενικού tokenizer του MTUOC-Eval, διαμορφωμένης ειδικά για την ελληνική γλώσσα.

Η εργασία χωρίζεται σε 7 Κεφάλαια. Στα δύο πρώτα περιγράφεται το θεωρητικό πλαίσιο αναφορικά με τη μηχανική μετάφραση, τα μέσα κοινωνικής δικτύωσης και την αξιολόγηση ενώ στο 3^ο Κεφάλαιο παρουσιάζεται εν συντομία το εργαλείο MTUOC και οι μετρικές αυτόματης αξιολόγησης που δύναται να υπολογίσει. Στο 4^ο Κεφάλαιο παρουσιάζεται αναλυτικά η μελέτη και η αξιολόγηση των μηχανικά μεταφρασμένων κειμένων ενώ στο 5^ο παρατίθεται η πρόταση του εξειδικευμένου tokenizer για την ελληνική γλώσσα. Ακολουθεί μια ποιοτική ανάλυση των αποτελεσμάτων στο 6^ο Κεφάλαιο με στοχευμένες αναφορές στα προβλήματα της μηχανικής μετάφρασης στα μέσα κοινωνικής δικτύωσης αλλά και μια γενική συγκριτική αξιολόγηση των δύο μεταφραστικών μηχανών που χρησιμοποιήθηκαν. Στο 7^ο Κεφάλαιο, τέλος, σκιαγραφείται το θεωρητικό πλαίσιο της εκτίμησης ποιότητας, μιας μεθόδου που χρησιμοποιείται για να παρέχει αυτόματα ενδείξεις ποιότητας που αφορούν μηχανικά μεταφρασμένα κείμενα χωρίς να βασίζεται σε ανθρώπινες μεταφράσεις αναφοράς.

Λέξεις κλειδιά: Μηχανική μετάφραση, αξιολόγηση, μέσα κοινωνικής δικτύωσης, κοινωνικός ιστός, Google Translate, Facebook Translate, μετρικές αυτόματης αξιολόγησης, MTUOC, εκτίμηση ποιότητας

Abstract

The main purpose of this master's dissertation is the study of the relationship between machine translation (MT) and the social media, through the comparative evaluation of machine translated texts, with regard to the language used in the social networking websites and the social media most popular topics. In order to achieve this, a certain number of texts were selected and translated making use exclusively of machine translation software. The machine translated texts were later evaluated applying both automated and manual evaluation methods. In order to be eligible for selection, the texts had either to be extracted from social networking websites or to regard popular topics on the social media, while also being part of bilingual parallel corpora so that they could be evaluated utilizing automated MT evaluation metrics.

Three language pairs were studied: German to English, Italian to English and Italian to Greek. All texts were automatically translated twice, making use of Google Translate and Facebook Translate respectively, while the MTUOC tool (and more specifically MTUOC-Eval) was used to calculate the automated MT evaluation metrics. Specifically to Greek as a target language, a variant of the general tokenizer of MTUOC-Eval is proposed, especially designed for the Greek language.

This paper is divided into 7 Chapters. The first two describe the theoretical framework regarding machine translation, social media and MT evaluation, while Chapter 3 briefly presents the MTUOC tool and the automated MT evaluation metrics it supports. The 4th Chapter contains a detailed presentation and a documentation of the machine translated texts evaluation. The Greek tokenizer is introduced in the 5th Chapter, followed by Chapter 6 where a qualitative analysis of the results takes place, including targeted references to the social media machine translation problems and a general comparative evaluation between Google Translate and Facebook Translate. The 7th and final Chapter introduces the basics of quality estimation, a method used to automatically provide a quality indication for machine translation output without depending on human reference translations.

Keywords: Machine translation, evaluation, social media, social networking websites, Google Translate, Facebook Translate, machine translation evaluation metrics, MTUOC, quality estimation

Ευχαριστίες

Θα ήθελα να εκφράσω τις θερμές μου ευχαριστίες στις καθηγήτριές μου, την κυρία Βιλελμίνη Σωσώνη και την κυρία Κάτια Κερμανίδου, για την καθοδήγησή τους, τη βοήθειά τους και τις εύστοχες παρατηρήσεις τους, καθ' όλη τη διάρκεια συγγραφής της παρούσας εργασίας.

Περιεχόμενα

1. Μηχανική Μετάφραση.....	7
1.1 Σύντομη αναδρομή	7
1.2 Μέσα κοινωνικής δικτύωσης και μηχανική μετάφραση	9
2. Αξιολόγηση της Μηχανικής Μετάφρασης	11
2.1 Ανθρώπινη Αξιολόγηση.....	12
2.1.1 Κατάταξη	12
2.1.2 Ακρίβεια και Ευχέρεια.....	12
2.1.3 Κατηγοριοποίηση λαθών	13
2.1.4 Μετεπεξεργασία.....	13
2.2 Αυτόματη Αξιολόγηση.....	13
2.2.1 Κριτήρια ποιότητας των μετρικών αυτόματης αξιολόγησης.....	14
2.2.2 Κατηγορίες μετρικών αυτόματης αξιολόγησης	15
3. Το Εργαλείο MTUOC.....	16
3.1 MTUOC-eval	16
3.2 BLEU.....	17
3.3 NIST.....	19
3.4 EditDistance.....	20
3.5 WER	20
3.6 TER.....	21
4. Μελέτη Παράλληλων Σωμάτων Κειμένου	22
4.1 Επιλογή κειμένων	22
4.2 Επιλογή Μηχανών Αυτόματης Μετάφρασης.....	23
4.3 Σχολιασμός FooTweets	24
4.4 Σχολιασμός Telework	29
4.5 Σχολιασμός La Cina e il vino italiano	34
4.6 Σχολιασμός Eleven Project	39
4.7 Σχολιασμός του Twitter στην υπηρεσία της υγείας.....	44
5. Greek Tokenizer: Μια πρόταση για την ελληνική γλώσσα	49
6. Ποιοτική Ανάλυση Αποτελεσμάτων	53
6.1 Ιδιαιτερότητες της γλώσσας στα μέσα κοινωνικής δικτύωσης	53
6.2 Συνολική Σύγκριση Facebook Translate και Google Translate.....	56

7. Εκτίμηση Ποιότητας	61
7.1 Πλεονεκτήματα	62
7.2 Επίπεδα ανάλυσης εκτίμησης ποιότητας	62
7.3 Πλαίσια ανοιχτού κώδικα για εκτίμηση ποιότητας	63
Βιβλιογραφία	64
Σύνδεσμοι.....	68
Παράρτημα.....	70
FooTweets	71
Κείμενο Αφετηρίας.....	71
Κείμενο Αναφοράς	74
Facebook Translate – Αξιολόγηση ακρίβειας και ευχέρειας	76
Google Translate – Αξιολόγηση ακρίβειας και ευχέρειας	80
Facebook Translate – Κατηγοριοποίηση Λαθών.....	84
Google Translate – Κατηγοριοποίηση Λαθών.....	92
Telework.....	100
Κείμενο Αφετηρίας.....	100
Κείμενο Αναφοράς	102
Facebook Translate – Αξιολόγηση ακρίβειας και ευχέρειας	104
Google Translate – Αξιολόγηση ακρίβειας και ευχέρειας	108
Facebook Translate - Κατηγοριοποίηση λαθών.....	112
Google Translate - Κατηγοριοποίηση λαθών.....	118
La Cina e il vino italiano	123
Κείμενο Αφετηρίας.....	123
Κείμενο Αναφοράς	125
Facebook Translate – Αξιολόγηση ακρίβειας και ευχέρειας	127
Google Translate – Αξιολόγηση ακρίβειας και ευχέρειας	130
Facebook Translate - Κατηγοριοποίηση λαθών.....	133
Google Translate - Κατηγοριοποίηση λαθών.....	139
Eleven Project.....	144
Κείμενο Αφετηρίας.....	144
Κείμενο Αναφοράς	146
Facebook Translate – Αξιολόγηση ακρίβειας και ευχέρειας	148
Google Translate – Αξιολόγηση ακρίβειας και ευχέρειας	152
Facebook Translate - Κατηγοριοποίηση λαθών.....	156

Google Translate - Κατηγοριοποίηση λαθών	162
Το Twitter στην υπηρεσία της υγείας	166
Κείμενο Αφετηρίας.....	166
Κείμενο Αναφοράς	168
Facebook Translate – Αξιολόγηση ακρίβειας και ευχέρειας	170
Google Translate – Αξιολόγηση ακρίβειας και ευχέρειας	174
Facebook Translate - Κατηγοριοποίηση λαθών	178
Google Translate - Κατηγοριοποίηση λαθών	185
Κώδικας Tokenizer για την ελληνική γλώσσα	190

1.

Μηχανική Μετάφραση

Η αυτόματη μηχανική μετάφραση είναι μια από τις πρώτες προκλήσεις που αντιμετώπισαν τα υπολογιστικά συστήματα και μία από τις πρώτες εφαρμογές της υπολογιστικής γλωσσολογίας, ήδη από τη δεκαετία του 1950. Αποτελεί ίσως μια από τις πιο απαιτητικές προκλήσεις για τα συστήματα τεχνητής νοημοσύνης κυρίως λόγω του ζωντανού χαρακτήρα της ανθρώπινης γλώσσας, της ρευστότητας δηλαδή και της διαρκούς εξέλιξης που τη χαρακτηρίζει.

Συνίσταται στην αυτόματη μετατροπή ενός κειμένου από μια γλώσσα αφετηρίας σε μια γλώσσα στόχο. Σε μια εργασία μηχανικής μετάφρασης, την είσοδο του συστήματος αποτελεί μια ακολουθία συμβόλων σε μια φυσική γλώσσα την οποία ο υπολογιστής καλείται μέσα από διάφορες διαδικασίες να μετατρέψει σε μια ακολουθία συμβόλων σε μια άλλη φυσική γλώσσα (Goodfellow et al 2016: 98). Είναι αυτονόητο πως για ένα τμήμα κειμένου σε μια γλώσσα αφετηρίας δεν υπάρχει μία και μοναδική βέλτιστη μετάφραση σε μια γλώσσα στόχο. Αυτό οφείλεται στη φυσική αμφισημία και ευελιξία που έχουν οι ανθρώπινες γλώσσες και ως εκ τούτου η πρόκληση της μηχανικής μετάφρασης καθίσταται ιδιαίτερα απαιτητική, ίσως από τις πιο δύσκολες στον τομέα της τεχνητής νοημοσύνης. Είναι επίσης γεγονός ότι η ακριβής μετάφραση απαιτεί την ύπαρξη βασικής προϋπάρχουσας γνώσης ώστε να επιλυθούν τα ζητήματα αμφισημίας και να αποδοθεί ορθά το περιεχόμενο μιας πρότασης (Russell & Norvig 2009:21). Για τον λόγο αυτό, κατά την ανάπτυξη συστημάτων αυτόματης μετάφρασης είναι απαραίτητη η παρουσία ανθρώπων ειδικών τόσο από τον χώρο της επιστήμης των υπολογιστών όσο και από τον χώρο της μετάφρασης και της γλωσσολογίας.

1.1 Σύντομη αναδρομή

Οι πρώτες προσπάθειες ξεκίνησαν τη δεκαετία του 1950 με τον ανταγωνισμό ΗΠΑ και ΕΣΣΔ να αντικατοπτρίζεται και στο πεδίο των συστημάτων μηχανικής μετάφρασης. Αναπτύχθηκε έντονη δραστηριότητα καθώς τα οφέλη έμοιαζαν να είναι σημαντικά σε θέματα άντλησης πληροφοριών, κατασκοπίας κλπ, ειδικά εκείνη την εποχή, με τον Ψυχρό Πόλεμο να έχει ήδη αρχίσει και σταδιακά να κλιμακώνεται. Στην πράξη τα αποτελέσματα δεν ήταν και τόσο ευνοϊκά καθώς τα συστήματα που αναπτύχθηκαν ήταν αρκετά απλά και δεν έμοιαζαν ικανά να δώσουν κάποιο ουσιαστικό όφελος στους ενδιαφερόμενους με συνέπεια το ενδιαφέρον σιγά σιγά να ατονήσει. Χαρακτηριστική είναι η αναφορά ALPAC του 1966 η οποία οδήγησε στη δραματική περικοπή κονδυλίων για την ανάπτυξη τέτοιου είδους συστημάτων στις ΗΠΑ εγκαταλείποντας ουσιαστικά την ιδέα.

Παρόλα αυτά, από τα μέσα της δεκαετίας του 1960 άρχισαν σιγά σιγά να αναπτύσσονται τα συστήματα δεύτερης γενιάς που ήταν πιο περίπλοκα τόσο από γλωσσολογικής όσο και από υπολογιστικής άποψης με τις πρώτες επιτυχίες να εμφανίζονται στα μέσα της επόμενης δεκαετίας (Météo, Systran). Οι κλασικές μέθοδοι μηχανικής μετάφρασης συχνά περιλαμβάνουν κανόνες για τη μετατροπή ενός κειμένου από τη γλώσσα προέλευσης στη γλώσσα προορισμού. Οι κανόνες συνήθως αναπτύσσονται από γλωσσολόγους και ενδέχεται να λειτουργούν σε λεξικό, συντακτικό ή σημασιολογικό επίπεδο. Η εν λόγω προσέγγιση, επειδή

ακριβώς βασίζεται στους κανόνες, ονομάζεται «μηχανική μετάφραση βάσει κανόνων» (RBMT – Rule Based Machine Translation) και χαρακτηρίζεται από τον καθορισμό και τη σαφή χρήση γλωσσολογικά ενημερωμένων κανόνων και αναπαραστάσεων (Olive et al 2011:133). Βασικούς περιορισμούς των κλασικών προσεγγίσεων μηχανικής μετάφρασης αποτελούν τόσο η τεχνογνωσία που απαιτείται για την ανάπτυξη των κανόνων, όσο και ο μεγάλος αριθμός κανόνων και εξαιρέσεων που χαρακτηρίζουν τις φυσικές γλώσσες.

Πλησιάζοντας προς τη δεκαετία του 1990 το ενδιαφέρον άρχισε και πάλι να αναζωπυρώνεται με χαρακτηριστικά σημεία την εμφάνιση των πρώτων εμπορικών συστημάτων μηχανικής μετάφρασης για προσωπικούς υπολογιστές στην Ιαπωνία και την επανέναρξη της χρηματοδότησης της έρευνας στις ΗΠΑ. Ταυτόχρονα άρχισαν να εμφανίζονται νέες προσεγγίσεις με βάση τα δεδομένα (data-driven methods). Δεδομένου ότι η γλωσσική μετάφραση επιβαρύνεται με πάρα πολλούς κανόνες και εξαιρέσεις που πρέπει να τυποποιηθούν, άρχισε να εξετάζεται η ιδέα της μετάφρασης με βάση πρότυπα παραδείγματα μεταφράσεων. Έτσι ξεκίνησαν να αναπτύσσονται τα πρώτα συστήματα μηχανικής μετάφρασης βάσει παραδειγμάτων (Example-based Machine Translation), κυρίως στην Ιαπωνία. Τα εν λόγω συστήματα αναζητούν σε παράλληλα σώματα κειμένων μια πρόταση παρόμοια με την πρόταση που πρέπει να μεταφραστεί και προσπαθούν στη συνέχεια να πραγματοποιήσουν τις κατάλληλες αλλαγές στην αποθηκευμένη μετάφραση ώστε να αποδίδει όσο το δυνατόν πιο πιστά το νόημα της πρότασης που πρέπει να μεταφραστεί.

Η πιο χαρακτηριστική προσέγγιση με βάση τα δεδομένα είναι η στατιστική μηχανική μετάφραση (SMT – Statistical Machine Translation) και ήταν και αυτή που κυριάρχησε τις τελευταίες δεκαετίες. Βασίζεται στη χρήση στατιστικών μοντέλων των οποίων οι παράμετροι προκύπτουν από την ανάλυση παράλληλων σωμάτων κειμένου και αντιμετωπίζει τη μετάφραση ως υπολογιστικό πρόβλημα. Τα στατιστικά μοντέλα προσπαθούν να «εκτιμήσουν» την κατάλληλη μετάφραση με βάση τη θεωρία πιθανοτήτων αποδίδοντας υψηλές πιθανότητες στις «καλές» μεταφράσεις και χαμηλότερες στις «κακές» (Williams et al 2016:xiii). Δεν βασίζεται σε γλωσσολογικούς κανόνες και αυτό αποτελεί σημαντικό πλεονέκτημα καθώς πρόκειται για γενική μέθοδο η οποία μπορεί να εφαρμοστεί σε οποιοδήποτε ζεύγος γλωσσών. Απ' την άλλη πλευρά, απαιτεί αρκετούς πόρους σε επίπεδο σωμάτων κειμένου με αποτέλεσμα σε αρκετά ζεύγη γλωσσών να μην υπάρχουν οι απαραίτητοι γλωσσικοί πόροι. Σημαντικό μειονέκτημα επίσης αποτελεί η αδυναμία μοντελοποίησης σύνθετων γλωσσικών φαινομένων καθώς επικεντρώνεται κατά κύριο λόγο στη μετάφραση φράσεων.

Η πιο σύγχρονη προσέγγιση στον χώρο της μηχανικής μετάφρασης είναι η νευρωνική μηχανική μετάφραση (NMT – Neural Machine Translation). Βασίζεται στη χρήση νευρωνικών δικτύων και σε αντίθεση με τις άλλες προσεγγίσεις, τα συστήματα των οποίων επικεντρώνονται στη μετάφραση φράσεων, προσπαθεί να δημιουργήσει και να εκπαιδεύσει ένα ευρύ νευρωνικό δίκτυο ικανό να διαβάσει μια πρόταση και να παράγει μια σωστή μετάφραση (Bahdanau et al. 2014). Το βασικό μοντέλο στο οποίο στηρίζεται είναι το μοντέλο κωδικοποιητή-αποκωδικοποιητή (Encoder-Decoder Model). Η ακολουθία εισόδου που αποτελεί το προς μετάφραση κείμενο κωδικοποιείται με συγκεκριμένες διαδικασίες και στη συνέχεια υφίσταται επεξεργασία μέσω της οποίας παράγεται το κείμενο στη γλώσσα στόχο. Έχει μικρότερες απαιτήσεις σε μνήμη και χρόνο σε σχέση με τη στατιστική μηχανική μετάφραση, γεγονός που την καθιστά περισσότερο αποδοτική και πιο γρήγορη. Κατά συνέπεια, τα συ-

στήματα νευρωνικής μηχανικής μετάφρασης κερδίζουν όλο και περισσότερο έδαφος στο προσκήνιο της μηχανικής μετάφρασης με χαρακτηριστικό παράδειγμα το Google Translate το οποίο από το 2017 και έπειτα χρησιμοποιεί πλέον το νευρωνικό μοντέλο.

1.2 Μέσα κοινωνικής δικτύωσης και μηχανική μετάφραση

Ένα από τα πιο χαρακτηριστικά φαινόμενα της τελευταίας δεκαπενταετίας είναι η ταχεία εξάπλωση των μέσων κοινωνικής δικτύωσης και η ευρεία χρήση τους από πολλούς ανθρώπους σε ολόκληρο τον πλανήτη. Σύμφωνα με τα στοιχεία της υπηρεσίας DataReportal (<https://datareportal.com/reports/digital-2020-global-digital-overview>) που παρέχει στατιστικές πληροφορίες σχετικά με τη χρήση του διαδικτύου και των εφαρμογών του, ο συνολικός αριθμός χρηστών των κοινωνικών δικτύων το 2020 ανερχόταν σε περίπου 3,8 δισεκατομμύρια παγκοσμίως. Τα στοιχεία του 2021 ανεβάζουν τον εν λόγω αριθμό στα 3,96 δισεκατομμύρια που αντιστοιχούν περίπου στο μισό του συνολικού πληθυσμού της γης και πιο συγκεκριμένα στο 50,64%. Στις ηλικίες μάλιστα από 13 ετών και πάνω το εν λόγω ποσοστό ανέρχεται στο 58,11% και στη Βόρεια Αμερική εκτοξεύεται στο 82% (<https://backlinko.com/social-media-users>). Οι αριθμοί αυτοί αρκούν από μόνοι τους για να καταλάβει κανείς με ευκολία ότι οι κοινωνικές αλληλεπιδράσεις επηρεάζονται πλέον καθοριστικά από τα μέσα κοινωνικής δικτύωσης τα οποία αποτελούν πια αναπόσπαστο κομμάτι της ζωής μας.

Για να περιγράψει κανείς αναλυτικά τις επιπτώσεις, θετικές και αρνητικές, που έχει η χρήση των κοινωνικών δικτύων σε όλο το φάσμα της σύγχρονης ζωής θα χρειαζόταν τόμους ολόκληρους. Στην παρούσα εργασία το ενδιαφέρον επικεντρώνεται στην παγκόσμια διάσταση των μέσων κοινωνικής δικτύωσης και στη δυνατότητα που παρέχεται στους χρήστες τους να έρθουν σε επαφή και να αλληλεπιδράσουν με ανθρώπους από όλο τον πλανήτη. Μέσω αυτής της αλληλεπίδρασης έχει κάποιος την ευκαιρία να γνωρίσει για παράδειγμα νέες κουλτούρες, ακόμα και στα πιο απομακρυσμένα σημεία της γης, το πιο σημαντικό όμως είναι το γεγονός ότι τα μέσα κοινωνικής δικτύωσης μπορούν να χρησιμοποιηθούν για την ευαισθητοποίηση και, γιατί όχι, τη δράση σε θέματα που αφορούν όλους τους κατοίκους του πλανήτη ανεξαιρέτως, όπως η παγκόσμια ειρήνη, η παγκόσμια υγεία, η κλιματική αλλαγή ή η μόλυνση του περιβάλλοντος.

Τα παραπάνω προϋποθέτουν την ύπαρξη μιας κοινής γλώσσας με τη βοήθεια της οποίας θα έχουν τη δυνατότητα οι χρήστες των κοινωνικών δικτύων να επικοινωνούν μεταξύ τους. Στην εποχή μας τον ρόλο της κοινής γλώσσας τον έχει κατά κύριο λόγο η αγγλική, πολλοί άνθρωποι όμως δεν έχουν τη δυνατότητα να τη διδαχθούν λόγω των συνθηκών κάτω από τις οποίες ζουν, ούτε βεβαίως είναι υποχρεωμένοι να το κάνουν. Σε αυτό το κομμάτι έρχεται να συμβάλλει καθοριστικά η μηχανική μετάφραση η οποία μπορεί άμεσα να διευκολύνει την επικοινωνία και την αλληλεπίδραση μεταξύ ανθρώπων που μιλούν τελείως διαφορετικές γλώσσες.

Τα μέσα κοινωνικής δικτύωσης αντιλήφθηκαν από νωρίς αυτή την ανάγκη ενσωματώνοντας δυνατότητες μηχανικής μετάφρασης, με χαρακτηριστικά παραδείγματα το Facebook και το Twitter που από το 2011 και το 2014 αντίστοιχα ξεκίνησαν να παρέχουν υπηρεσίες αυτόματης μετάφρασης των δημοσιεύσεων σε συνεργασία με τον Bing Translator της Microsoft. Το

Facebook προχώρησε ένα βήμα παραπέρα έχοντας πλέον έχουν αναπτύξει το δικό του εργαλείο μηχανικής μετάφρασης που βρίσκεται σε λειτουργία από το 2017 ενώ τον Οκτώβριο του 2020 σύστησε για πρώτη φορά το M2M-100 (Fan 2020) ένα νέο σύστημα βασιζόμενο στο πολυγλωσσικό μοντέλο μηχανικής μετάφρασης Many-to-Many (Many-to-Many multilingual translation model). Το συγκεκριμένο μοντέλο βασίζεται στην απευθείας μετάφραση από μια γλώσσα σε μια άλλη χωρίς την παρεμβολή της αγγλικής γλώσσας όπως συμβαίνει στα περισσότερα μοντέλα λόγω του γεγονότος ότι η συντριπτική πλειοψηφία των δεδομένων για την εκπαίδευση των συστημάτων μηχανικής μετάφρασης αφορά την αγγλική γλώσσα. Αναφέρεται μάλιστα χαρακτηριστικά ότι στις δοκιμές συγκέντρωσε βαθμολογία BLEU έως και 10 μονάδες υψηλότερη σε σχέση με τα αγγλοκεντρικά συστήματα.

2.

Αξιολόγηση της Μηχανικής Μετάφρασης

Καθώς η μηχανική μετάφραση κερδίζει όλο και περισσότερο έδαφος στο χώρο της μετάφρασης, το ζήτημα της ποιότητάς της γίνεται όλο και πιο σημαντικό (Mauscec & Donaj: 2019) και η αξιολόγηση έχει ως στόχο τον προσδιορισμό της ποιότητας ενός μεταφράσματος που έχει παραχθεί από κάποια μηχανή αυτόματης μετάφρασης. Η αξιολόγηση των συστημάτων μηχανικής μετάφρασης είναι ένα πολύ σημαντικό πεδίο έρευνας, τόσο για την εκτίμηση της αποτελεσματικότητας των μηχανών αυτόματης μετάφρασης όσο και για τη βελτιστοποίηση της απόδοσής τους (Dorr et al. 2011: 801). Παρέχει χρήσιμες πληροφορίες για το αν ένα σύστημα λειτουργεί όπως πρέπει, για το ποια τμήματά του είναι αποτελεσματικά και για το ποια χρειάζονται βελτίωση ενώ πρέπει ταυτόχρονα να είναι ειλικρινής και αναπαράξιμη, και οι μέθοδοί της πρέπει να είναι όσο το δυνατόν πιο ενδεδειγμένες (Castilho 2019).

Οι εμπλεκόμενοι στη διαδικασία αξιολόγησης της μηχανικής μετάφρασης μπορεί να είναι πολλοί. Θα μπορούσε για παράδειγμα να είναι οι τελικοί χρήστες ενός εργαλείου μηχανικής μετάφρασης που δίνουν έμφαση στο νόημα και τη μετάδοση των πληροφοριών, μεταφραστές που αξιολογούν αν ένα μετάφρασμα χρειάζεται πλήρη ή ελαφριά μετεπεξεργασία, προγραμματιστές συστημάτων μηχανικής μετάφρασης που χρησιμοποιούν την αξιολόγηση για συντονισμό, διάγνωση προβλημάτων και βελτιώσεις στα συστήματά τους καθώς και άλλου είδους εφαρμογές που εξυπηρετούν συγκεκριμένους σκοπούς, όπως η ανάκτηση δεδομένων μεταξύ γλωσσών (Cross Language Information Retrieval) (Castilho 2019). Η αξιολόγηση είναι σημαντική για όλους τους εμπλεκόμενους. Οι ερευνητές πρέπει να γνωρίζουν εάν οι θεωρίες τους κάνουν τη διαφορά, οι επαγγελματίες προγραμματιστές θέλουν να εντυπωσιάσουν τους πελάτες τους ώστε να πουλήσουν το προϊόν τους και οι χρήστες πρέπει να αποφασίσουν ποιο σύστημα θα χρησιμοποιήσουν (Hovy et al. 2002).

Ο σκοπός για τον οποίο γίνεται μια μετάφραση παίζει επίσης σημαντικό ρόλο στην αξιολόγησή της. Το πόσο καλή ή όχι αξιολογείται μια μετάφραση εξαρτάται από τον σκοπό τον οποίο εξυπηρετεί ενώ όσοι εμπλέκονται στον χώρο της εμπορικής παραγωγής συστημάτων μηχανικής μετάφρασης λένε συχνά πως «*Ποιότητα είναι αυτό που ορίζει ο πελάτης*» (Koehn 2020: 42). Αν για παράδειγμα αντιμετωπίζουμε κάποιο πρόβλημα στον υπολογιστή μας και αναζητώντας λύση ψάχνοντας στις μηχανές αναζήτησης καταφέρουμε να βρούμε μόνο ένα κείμενο γραμμένο σε άγνωστη γλώσσα, πχ τα φινλανδικά, το οποίο όμως με τη χρήση ενός εργαλείου αυτόματης μετάφρασης θα μας οδηγήσει στη λύση του προβλήματος ακόμα και αν το αποτέλεσμα της μετάφρασης είναι γεμάτο από λάθη, μπορούμε να πούμε ότι η μετάφραση είναι ικανοποιητική (Koehn 2020: 42). Αυτού του είδους η αξιολόγηση ορίζεται ως αξιολόγηση με βάση την αποστολή (task-based translation evaluation).

Η αξιολόγηση της μηχανικής μετάφρασης μπορεί να γίνει είτε από ανθρώπους είτε από υπολογιστές. Μπορούμε λοιπόν να την χωρίσουμε σε δύο μεγάλες κατηγορίες, την ανθρώπινη αξιολόγηση και την αυτόματη αξιολόγηση.

2.1 Ανθρώπινη Αξιολόγηση

Η ανθρώπινη αξιολόγηση έχει το πλεονέκτημα ότι μπορεί να εκτιμήσει καλύτερα τα σύνθετα γλωσσικά φαινόμενα και άλλες ιδιαιτερότητες που μπορεί να έχει μια φυσική γλώσσα. Είναι πιο λεπτομερής σε σχέση με την αυτόματη αξιολόγηση ενώ δεν πρέπει να παραγνωρίζουμε το γεγονός ότι οι μεταφράσεις απευθύνονται σε ανθρώπους, συνεπώς η ανθρώπινη κρίση είναι το ιδανικό μέτρο αξιολόγησης για την ποιότητά της. Επιπλέον, η ανθρώπινη κατανόηση του πραγματικού κόσμου από έναν άνθρωπο-κριτή του δίνει τη δυνατότητα να εκτιμήσει καλύτερα την πρακτική σημασία των σφαλμάτων σε μια μετάφραση (Sanders et al. 2011: 807).

Οι πιο κοινές πρακτικές που χρησιμοποιούνται κατά την ανθρώπινη αξιολόγηση της μηχανικής μετάφρασης (Castilho 2017) είναι η κατάταξη (ranking), η αξιολόγηση ως προς την ακρίβεια (adequacy) και την ευχέρεια (fluency), η κατηγοριοποίηση λαθών (error categorization) και η μετεπεξεργασία (post-editing).

2.1.1 Κατάταξη

Η κατάταξη βασίζεται στη σύγκριση μεταφρασμάτων για το ίδιο κείμενο αφετηρίας που προέρχονται είτε από συστήματα μηχανικής μετάφρασης είτε έχουν γίνει από ανθρώπους. Στην προσπάθεια να κριθεί το ποια μετάφραση είναι καλύτερη κατατάσσονται από το καλύτερο προς το χειρότερο ή αντίστροφα και μπορούν να μας δώσουν χρήσιμες πληροφορίες. Κατατάσσοντας για παράδειγμα μεταφράσματα του ίδιου κειμένου αφετηρίας από το ίδιο σύστημα μηχανικής μετάφρασης σε διαφορετικά στάδια της ανάπτυξής του μπορούμε να βγάλουμε συμπεράσματα για το αν έχει σημειωθεί πρόοδος κατά την ανάπτυξη του συστήματος. Κατατάσσοντας μεταφράσματα που προέρχονται από διαφορετικά συστήματα μπορούμε να βγάλουμε χρήσιμα συμπεράσματα για το ποιο σύστημα είναι καταλληλότερο για το συγκεκριμένο είδος κειμένου.

2.1.2 Ακρίβεια και Ευχέρεια

Η αξιολόγηση ως προς την ακρίβεια και την ευχέρεια πραγματοποιείται χωρίζοντας το μεταφρασμένο κείμενο σε μικρότερα τμήματα και αποδίδοντας σε κάθε τμήμα μια βαθμολογία σε κλίμακα Likert. Οι βαθμίδες και το περιεχόμενο της κλίμακας Likert μπορούν να οριστούν από τον αξιολογητή, σύμφωνα όμως πάντα με τα πρότυπα του DQF (Dynamic Quality Framework), μιας πλατφόρμας που αναπτύχθηκε από την εταιρεία TAUS (Translation Automation User Society) το 2011 με σκοπό την τυποποίηση της αξιολόγησης της ποιότητας της μετάφρασης, μηχανικής και μη. Για τις ανάγκες της παρούσας εργασίας χρησιμοποιήθηκε κλίμακα Likert 5 βαθμίδων και πιο συγκεκριμένα των βαθμίδων Very Poor (Πολύ Φτωχή), Poor (Φτωχή), Fair (Μέτρια), Good (Καλή) και Excellent (Άριστη).

Η ακρίβεια ενός μεταφράσματος σχετίζεται με το κατά πόσο ένα μετάφρασμα αποδίδει επαρκώς το σημασιολογικό νόημα του κειμένου αφετηρίας. Ανάλογα με το βαθμό στον οποίο ο αξιολογητής κρίνει ότι το μεταφρασμένο τμήμα κειμένου ανταποκρίνεται στο νόημα του αντίστοιχου τμήματος στο κείμενο αφετηρίας, το κατατάσσει στην αντίστοιχη βαθμίδα της κλίμακας Likert. Επικεντρώνεται λοιπόν κατά κύριο λόγο στο κείμενο αφετηρίας μιας και έχει να κάνει με το νόημα της πρότασης και αγνοεί άλλα χαρακτηριστικά του κειμένου-

στόχου όπως το αν το αποτέλεσμα θα είναι ευανάγνωστο, χωρίς σημαντικά γραμματικά λάθη κτλ.

Η ευχέρεια ενός μεταφράσματος επικεντρώνεται ακριβώς στα χαρακτηριστικά που αναφέρονται στην παραπάνω παράγραφο και αγνοεί η ακρίβεια. Εξετάζει δηλαδή κατά κύριο λόγο το πόσο ευανάγνωστο είναι το μεταφρασμένο κείμενο, το κατά πόσο μοιάζει να είναι γραμμένο από άνθρωπο και όχι από μηχανή, αν έχει γραμματικά ή ορθογραφικά λάθη καθώς και άλλα χαρακτηριστικά που αφορούν την γλώσσα-στόχο ή ακόμα και την κουλτούρα των ομιλητών της. Με βάση λοιπόν αυτά τα χαρακτηριστικά, κάθε τμήμα του μεταφρασμένου κειμένου κατατάσσεται σε μια βαθμίδα της κλίμακας Likert.

Αξίζει να σημειωθεί ότι η ευχέρεια αφορά μόνο το μετάφρασμα και δεν λαμβάνει υπόψη το κείμενο αφετηρίας. Έτσι, μπορεί για παράδειγμα ένα τμήμα κειμένου να αξιολογηθεί άριστα ως προς την ευχέρειά του γιατί μπορεί να μοιάζει ότι γράφτηκε από άνθρωπο αλλά να έχει αποδώσει τελείως λανθασμένα το νόημα του αντίστοιχου τμήματος στο πρωτότυπο κείμενο ή, αντίθετα, να αποδίδει άριστα το νόημα του πρωτότυπου κειμένου λαμβάνοντας άριστη αξιολόγηση ως προς την ακρίβεια, περιέχοντας όμως γραμματικά και συντακτικά λάθη. Για τον λόγο αυτό, η αξιολόγηση ως προς την ακρίβεια και την ευχέρεια συνήθως γίνονται μαζί.

2.1.3 Κατηγοριοποίηση λαθών

Η κατηγοριοποίηση των λαθών γίνεται με βάση το πλαίσιο MQM (Multidimensional Quality Metrics) που αναπτύχθηκε από την πρωτοβουλία QTLaunchPad με τη χρηματοδότηση της Ευρωπαϊκής Ένωσης. Με βάση το εν λόγω πλαίσιο, τα διάφορα λάθη που εντοπίζονται σε ένα μετάφρασμα εντάσσονται σε κάποιες τυποποιημένες κατηγορίες. Η περιγραφή και η ανάλυση των διαφόρων κατηγοριών ξεφεύγει από τα πλαίσια της παρούσας εργασίας, μπορεί όμως κάποιος να τις δει αναλυτικά στον σχετικό δικτυακό τόπο (<http://www.qt21.eu/mqm-definition/definition-2015-12-30.html>, τελευταία προσπέλαση: 23/03/2021).

2.1.4 Μετεπεξεργασία

Η αξιολόγηση σε σχέση με τη μετεπεξεργασία αποτελεί κομμάτι της αξιολόγησης με βάση την αποστολή που αναφέρθηκε και παραπάνω. Αν η μετεπεξεργασία ενός μηχανικά μεταφρασμένου κειμένου απαιτεί λιγότερο κόπο και χρόνο για έναν μεταφραστή από το να μεταφράσει εξ αρχής ένα κείμενο, σημαίνει ταυτόχρονα και αύξηση της παραγωγικότητας του εν λόγω μεταφραστή αφού θα είναι πλέον σε θέση να «παράγει» μεγαλύτερο όγκο μεταφρασμένων κειμένων. Κάτι τέτοιο συνεπάγεται προφανώς και θετική αξιολόγηση για την ποιότητα της μηχανικής μετάφρασης ακόμα κι αν περιέχει σφάλματα (Koehe 2020: 44).

2.2 Αυτόματη Αξιολόγηση

Παρά τα αδιαμφισβήτητα πλεονεκτήματα που προσφέρει η αυτόματη αξιολόγηση υπάρχουν βεβαίως και μειονεκτήματα τα οποία έχουν να κάνουν κατά κύριο λόγο με το κόστος τόσο σε οικονομικό επίπεδο όσο και επίπεδο χρόνου. Οι άνθρωποι που λαμβάνουν μέρος σε μια διαδικασία αξιολόγησης πρέπει προφανώς να πληρωθούν για την εργασία τους ενώ

η ταχύτητα με την οποία δουλεύουν και ο όγκος των κειμένων που μπορούν να αξιολογηθούν σε συγκεκριμένο χρονικό διάστημα είναι κατά πολύ μικρότερα σε σχέση με τις δυνατότητες που παρέχουν οι υπολογιστές.

Για τον λόγο αυτό αναπτύχθηκαν κάποιες μετρικές αυτόματης αξιολόγησης οι οποίες είναι στην ουσία υπολογιστικά προγράμματα. Βασίζονται στη σύγκριση των κειμένων που παράγονται από μηχανές αυτόματης μετάφρασης με αντίστοιχες ανθρώπινες μεταφράσεις στα ίδια πρωτότυπα κείμενα που ονομάζονται κείμενα αναφοράς. Δέχονται ως είσοδο το αποτέλεσμα μιας μηχανικής μετάφρασης και κάποιο κείμενο αναφοράς και στη συνέχεια αφού καταγραφούν ομοιότητες και διαφορές μεταξύ των κειμένων, με κριτήρια που ορίζονται διαφορετικά για κάθε μετρική, παράγεται κάποια βαθμολογία η οποία είναι τελικά αυτή που δίνει και την εικόνα για την ποιότητα της μετάφρασης.

Οι μετρικές αυτόματης αξιολόγησης ποιότητας της μηχανικής μετάφρασης αναπτύχθηκαν λόγω του υψηλού κόστους, της έλλειψης επαναληψιμότητας, της υποκειμενικότητας και της βραδύτητας της ανθρώπινης αξιολόγησης καθώς και λόγω της ανάγκης για άμεση ενεργοποίηση και αυτόματο συντονισμό των παραμέτρων ενός συστήματος μηχανικής μετάφρασης (Och 2003). Κατά την ανάπτυξη ενός τέτοιου συστήματος, η συχνή αξιολόγηση της παραγόμενης μετάφρασης είναι επιτακτική ανάγκη. Ενδέχεται να χρειαστούν αρκετές αξιολογήσεις ακόμα και κατά τη διάρκεια μιας μέρας ώστε να εξεταστούν καλύτερα οι διαφορές πτυχές του συστήματος. Η πρόοδος που έχει συντελεστεί τα τελευταία χρόνια είναι πολύ σημαντική, έτσι ώστε ο σχεδιασμός και η πορεία ανάπτυξης ενός συστήματος μηχανικής μετάφρασης να επηρεάζεται σε σημαντικό βαθμό από τα αποτελέσματα της αυτόματης αξιολόγησης (Koehn 2020: 53).

2.2.1 Κριτήρια ποιότητας των μετρικών αυτόματης αξιολόγησης

Για να είναι αποδοτική η χρήση μετρικών αυτόματης αξιολόγησης πρέπει να ικανοποιούνται κάποια βασικά κριτήρια. Ένα από τα πιο σημαντικά κριτήρια είναι η υψηλή συσχέτιση με την ανθρώπινη κρίση σχετικά με την ποιότητα μιας μηχανικής μετάφρασης. Είναι επίσης πολύ σημαντικό μια μετρική να έχει υψηλή ευαισθησία και να αποτυπώνει τις διαφορές στην ποιότητα της μηχανικής μετάφρασης που παράγεται από δύο διαφορετικά συστήματα αλλά και από διαφορετικές εκδόσεις του ίδιου συστήματος (Banerjee, S. & Lavie, A.2005). Ένα τρίτο σημαντικό κριτήριο αποτελεί το κόστος. Μια αυτόματη μετρική πρέπει να έχει χαμηλό κόστος καθώς το υψηλό κόστος σε χρόνο και χρήμα είναι το κύριο μειονέκτημα των μετρικών αξιολόγησης που εμπλέκουν τον ανθρώπινο παράγοντα. Πρέπει επίσης να έχει νόημα, να είναι δηλαδή εύκολη η ερμηνεία της και όχι να παρέχει απλώς μια δυσνόητη βαθμολογία. Η συνέπεια είναι ένα ακόμα χαρακτηριστικό γνώρισμα που πρέπει να διαθέτει. Η χρήση της ίδιας μετρικής από διαφορετικά συστήματα πρέπει να οδηγεί στα ίδια συμπεράσματα. Επιπλέον, απαιτείται συνέπεια στο πώς αξιολογούνται δύο παρεμφερή δοκιμαστικά σώματα κειμένου από το ίδιο σύστημα. Σε περίπτωση που παρουσιαστούν μεγάλες διακυμάνσεις η μετρική δεν μπορεί να θεωρηθεί σταθερή, κάτι το οποίο σημαίνει ότι χρειάζονται μεγάλα δοκιμαστικά σώματα κειμένου ώστε να εξασφαλίζεται η αξιοπιστία των αποτελεσμάτων. Τέλος, σε κάθε περίπτωση πρέπει να εξασφαλίζεται η ορθότητα των αποτελεσμάτων και, καθώς η ευχέρεια και η ακρίβεια είναι αυτές που καθορίζουν σε σημαντι-

κό βαθμό την ποιότητα μιας μετάφρασης, οι αυτόματες μετρικές επικεντρώνονται στη συσχέτιση με τους δύο αυτούς παράγοντες (Koehn 2010).

Πέρα από τα παραπάνω, είναι πολύ σημαντικό για μια αυτόματη μετρική να μπορεί να εξάγει γρήγορα αποτελέσματα χωρίς σημαντική επιβάρυνση για τους πόρους ενός συστήματος. Μετρικές που απαιτούν επιπλέον πόρους όπως πχ γλωσσικά εργαλεία ή δεδομένα που δεν είναι ήδη διαθέσιμα για μια γλώσσα ή ακόμα και εκπαίδευση ενός συστήματος μπορεί να αυξήσουν σημαντικά το υπολογιστικό κόστος και να καταστούν ασύμφορες. (Koehn 2020). Αντίθετα πρέπει να μπορούν να ενσωματωθούν εύκολα σε διάφορα συστήματα ώστε η χρήση τους να είναι αποδοτική.

2.2.2 Κατηγορίες μετρικών αυτόματης αξιολόγησης

Οι μετρικές αυτόματης αξιολόγησης μπορούν να χωριστούν σε δύο μεγάλες κατηγορίες ανάλογα με την τεχνική την οποία χρησιμοποιούν. Έχουμε λοιπόν τις μετρικές που βασίζονται στην ακρίβεια (precision) και στην ανάκληση (recall) και τις μετρικές που βασίζονται στην απόσταση διόρθωσης (edit distance).

Η ακρίβεια αφορά τον λόγο του αριθμού των ν-γραμμάτων (των ακολουθιών δηλαδή ν μονάδων όπου οι μονάδες μπορεί να είναι χαρακτήρες, συλλαβές, λέξεις, σημεία στίξης κτλ) του μεταφράσματος που υπάρχουν και στο κείμενο αναφοράς προς το σύνολο των ν-γραμμάτων του μεταφράσματος. Η ανάκληση απ' την άλλη πλευρά αναφέρεται στον λόγο του αριθμού των ν-γραμμάτων του μεταφράσματος που εντοπίζονται και στο κείμενο αναφοράς προς το σύνολο των ν-γραμμάτων του κειμένου αναφοράς. Οι πιο σημαντικές μετρικές αυτόματης αξιολόγησης που βασίζονται στην ακρίβεια και την ανάκληση είναι η BLEU, η NIST, η ROUGE, η METEOR, η GTM, η WNM, η ParaEval, η BLANC και η LEPOR.

Η απόσταση διόρθωσης βασίζεται στην απόσταση κατά Levenshtein και ορίζεται ως ο ελάχιστος αριθμός ενεργειών που πρέπει να γίνουν σε μια συμβολοσειρά α ώστε να μετατραπεί σε μια συμβολοσειρά β . Ως ενέργειες διόρθωσης θεωρούνται η εισαγωγή, η διαγραφή και η αντικατάσταση ενός χαρακτήρα από κάποιον άλλο (Levenshtein, 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. Η πιο χαρακτηριστική μετρική αυτόματης αξιολόγησης που βασίζεται στην απόσταση διόρθωσης είναι η WER και οι διάφορες παραλλαγές της όπως η WERg, η TER, η invWER, η SER, η CDER, η HyTER και η PER.

Στην παρούσα εργασία χρησιμοποιήθηκαν οι μετρικές που υποστηρίζονται από το εργαλείο MTUOC και πιο συγκεκριμένα οι BLEU και NIST με βάση την ακρίβεια και την ανάκληση και οι WER και TER και EditDistance με βάση την απόσταση διόρθωσης. Οι εν λόγω μετρικές θα παρουσιαστούν πιο αναλυτικά παρακάτω.

3.

Το Εργαλείο MTUOC

Το MTUOC (Oliver 2020) είναι ένα έργο από το Τμήμα Τεχνών και Ανθρωπιστικών Επιστημών του Πανεπιστημίου Oberta de Catalunya (UOC) για τη διευκόλυνση της χρήσης και της ολοκλήρωσης συστημάτων νευρωνικής και στατιστικής μηχανικής μετάφρασης. Το μεγαλύτερο μέρος του λογισμικού που απαιτείται για την εκπαίδευση και τη χρήση τέτοιων συστημάτων διανέμεται με δωρεάν άδειες και, ως εκ τούτου, διατίθεται ελεύθερα για προσωπική, ακαδημαϊκή ή επαγγελματική χρήση. Προσφέρει μια σειρά από εργαλεία, σενάρια και προγράμματα που επιτρέπουν την εκπαίδευση συστημάτων νευρωνικής και στατιστικής μηχανικής μετάφρασης καθώς και την πραγματοποίηση μεταφράσεων με τη χρήση αυτών των συστημάτων ανεξάρτητα από το λειτουργικό σύστημα του χρήστη. Πιο συγκεκριμένα, προσφέρονται:

- Δομοστοιχεία Python που παρέχουν βασικές λειτουργίες όπως tokenization, truecasing κτλ
- Δέσμες ενεργειών Bash και Python καθώς και αρκετά αρχεία ρυθμίσεων που επιτρέπουν την προεπεξεργασία σωμάτων κειμένου και την εκπαίδευση συστημάτων μηχανικής μετάφρασης
- **MTUOC Server**: ο διακομιστής MTUOC λαμβάνει ένα τμήμα κειμένου για μετάφραση από έναν πελάτη (client) και το στέλνει στον MT Server ή διακομιστή μηχανικής μετάφρασης (Moses, Marian, ONMT ή ModernMT). Εκτελεί τις απαιτούμενες ενέργειες προεπεξεργασίας και όταν λάβει το μεταφρασμένο κείμενο, το μετεπεξεργάζεται και το στέλνει πίσω στον πελάτη. Ο εν λόγω διακομιστής μπορεί να προσομοιώσει διάφορα είδη διακομιστών μηχανικής μετάφρασης και να λειτουργήσει ως MTUOC, Moses, Marian, ONMT, NMTWizard ή ModernMT, γεγονός που τον καθιστά συμβατό με πολλά εργαλεία CAT.
- **MTUOC Client**: ο πελάτης MTUOC, λεγόμενος και μεταφραστής MTUOC, μπορεί να χειριστεί αρχεία κειμένου και XLIFF, να συνδεθεί με τον διακομιστή μηχανικής μετάφρασης και να μεταφράσει τα αρχεία. Μπορεί να συνδεθεί με διακομιστές MTUOC, Moses, Marian, ONMT, NMTWizard και ModernMT.
- **MTUOC-eval**: το εργαλείο για αυτόματη αξιολόγηση συστημάτων μηχανικής μετάφρασης το οποίο θα χρησιμοποιηθεί στην παρούσα εργασία. Επιτρέπει τον υπολογισμό των μετρικών BLEU, NIST, WER, EditDistance και TER.
- **Εικονικές μηχανές MTUOC**: οι περισσότερες σουίτες εργαλείων μηχανικής μετάφρασης λειτουργούν μόνο με Linux. Οι εικονικές μηχανές Virtual Box επιτρέπουν τη λειτουργία του διακομιστή MTUOC σε οποιοδήποτε λειτουργικό σύστημα.
- Προ-εκπαιδευμένα συστήματα νευρωνικής μηχανικής μετάφρασης έτοιμα προς χρήση με τον διακομιστή MTUOC

3.1 MTUOC-eval

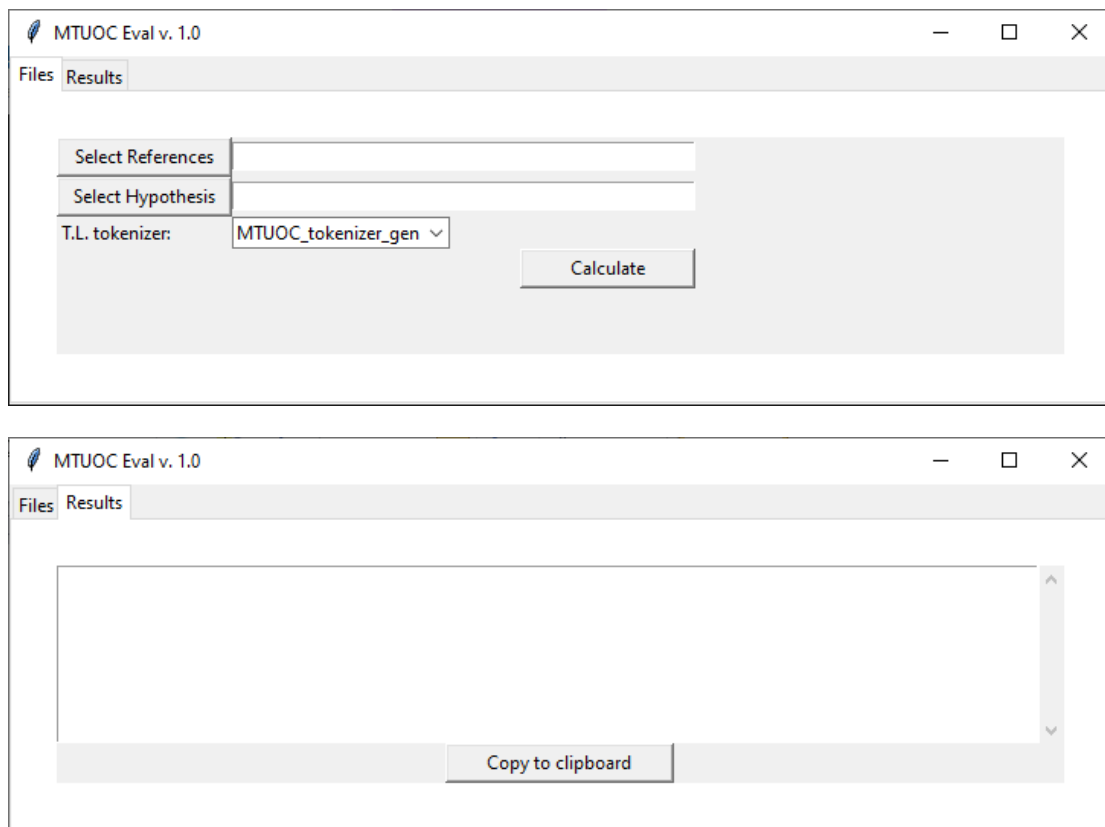
Το MTUOC-eval είναι ένα εργαλείο που επιτρέπει τον υπολογισμό μιας σειράς μετρικών αυτόματης αξιολόγησης για την αξιολόγηση μηχανικά μεταφρασμένων κειμένων. Είναι γραμμένο στη γλώσσα προγραμματισμού Python και προσφέρεται σε δύο εκδόσεις, μία για

λειτουργία σε τερματικό με γραμμή εντολών και μία με γραφική διεπαφή χρήστη. Διανέμεται σε κώδικα Python αλλά και ως εκτελέσιμο αρχείο των Windows.

Το MTUOC-eval επιτρέπει τον υπολογισμό των ακόλουθων μετρικών αυτόματης αξιολόγησης:

- BLEU
- NIST
- WER
- %EdDist
- TER

Για τις ανάγκες της παρούσας εργασίας χρησιμοποιήθηκε η έκδοση MTUOC-eval GUI που διαθέτει γραφική διεπαφή χρήστη.



3.2 BLEU

Η βαθμολογία BLEU (Papileni et al. 2002) είναι ένας αλγόριθμος αντιστοίχισης συμβολοσειρών που παρέχει βασικές μετρήσεις ποιότητας της μηχανικής μετάφρασης. Πρόκειται για την πιο ευρέως χρησιμοποιούμενη μετρική αυτόματης αξιολόγησης της μηχανικής μετάφρασης τα τελευταία 15 χρόνια. Αν και είναι σε μεγάλο βαθμό αποδεκτό ότι η BLEU έχει ελαττώματα, παραμένει μια από τις κύριες μετρικές για την αξιολόγηση των αποτελεσμάτων εξόδου ενός συστήματος μηχανικής μετάφρασης ακόμη και σήμερα, στην εποχή της ακμής της νευρωνικής μηχανικής μετάφρασης (Vashee, 2019).

Η BLEU απέκτησε δημοτικότητα επειδή ήταν μια από τις πρώτες μετρικές που παρουσίασε υψηλή συσχέτιση με την ανθρώπινη κρίση σχετικά με την ποιότητα της μηχανικής μετάφρασης, την πιο βασική από τις προϋποθέσεις που πρέπει να πληροί μια μετρική αυτόματης αξιολόγησης (Specia et al. 2010). Μελέτες έχουν δείξει ότι ο βαθμός συσχέτισης είναι όντως πολύ υψηλός, υπό την προϋπόθεση όμως ότι η BLEU χρησιμοποιείται σωστά. Παρά την αμφισβήτηση που έχει δεχτεί και παρά το γεγονός ότι η έρευνα σχετικά με τις αυτόματες μετρικές μετρά πάνω από μια δεκαετία ζωής, η BLEU παραμένει βασικός πυλώνας στην αυτόματη αξιολόγηση ποιότητας. Ο λόγος είναι το πρόβλημα αξιοπιστίας που παρουσιάζει η εμπιστοσύνη σε κάθε νέα καινούργια μετρική (Koehn 2020: 64). Πρακτικά δηλαδή, οι ερευνητές που αναφέρουν βαθμολογίες βασιζόμενες σε σχετικά νέες μετρικές ερωτώνται πάντοτε και για το αν και πώς οι βελτιώσεις που προτείνουν για τα συστήματά τους οδηγούν και σε υψηλότερες βαθμολογίες BLEU.

Η BLEU επιχειρεί να μετρήσει την αντιστοιχία μεταξύ των αποτελεσμάτων εξόδου ενός συστήματος μηχανικής μετάφρασης με τα αντίστοιχα αποτελέσματα μιας ανθρώπινης μετάφρασης. Η κεντρική ιδέα είναι ότι όσο πιο κοντά είναι μια αυτόματη μετάφραση σε μια επαγγελματική ανθρώπινη μετάφραση, τόσο καλύτερη είναι. Βασίζεται στην ακρίβεια (precision) και χρησιμοποιεί μια τροποποιημένη μορφή της για να συγκρίνει ν-γράμματα (για ν από 1 έως 4) μεταξύ του μηχανικού μεταφράσματος και του κειμένου αναφοράς. Η τροποποίηση γίνεται επειδή αρκετές φορές οι μηχανές αυτόματης μετάφρασης μπορεί να παράγουν υπερβολικά μεγάλο αριθμό ίδιων αποδεκτών λέξεων (που εντοπίζονται δηλαδή και στο κείμενο αναφοράς) με κίνδυνο να οδηγήσουν σε λανθασμένα συμπεράσματα.

$$\text{Ακρίβεια} = \frac{\# \text{ αποδεκτών } n - \text{ γραμμάτων}}{\# n - \text{ γραμμάτων του μεταφράσματος}}$$

$$\text{Τροποποιημένη Ακρίβεια} = \frac{\text{μέγιστος } \# \text{ αποδεκτών } n - \text{ γραμμάτων}}{\# n - \text{ γραμμάτων του μεταφράσματος}}$$

Χαρακτηριστικό παράδειγμα η υποψήφια μετάφραση *the the the the the the the* με κείμενα αναφοράς το *the cat is on the mat* και το *there is a cat on the mat* (Papileni et al. 2002). Συγκρίνοντας σε επίπεδο 1-γραμμάτων, και από τη στιγμή που η λέξη *the* είναι αποδεκτή, η τιμή της ακρίβειας είναι 7/7 για συγκρίσεις και με τα δύο κείμενα αναφοράς ενώ είναι προφανές ότι το μετάφρασμα δεν είναι σε καμία περίπτωση αυτό που πρέπει. Χρησιμοποιώντας την τροποποιημένη ακρίβεια η τιμή μεταβάλλεται σε 2/7 για το πρώτο κείμενο αναφοράς και 1/7 για το δεύτερο κείμενο αναφοράς, κάτι που είναι πολύ πιο κοντά στην πραγματικότητα.

Όσον αφορά την ανάκληση (recall), επειδή τα μηχανικά μεταφράσματα έχουν συνήθως μικρότερη έκταση σε σχέση με τις αντίστοιχες μεταφράσεις που γίνονται από ανθρώπους, είναι προφανές ότι ευνοεί τα μικρά κομμάτια κειμένου όπου οι διαφορές είναι μικρότερες ενώ πρέπει να ληφθεί υπόψη και το γεγονός ότι μπορεί να υπάρχουν περισσότερες από μία μεταφράσεις αναφοράς.

$$\text{Ανάκληση} = \frac{\# \text{ αποδεκτών } n - \text{ γραμμάτων}}{\# n - \text{ γραμμάτων του κειμένου αναφοράς}}$$

Για τον λόγο αυτό προβλέπεται η απόδοση μιας ποινής συντομίας (brevity penalty) στις μικρές προτάσεις. Η ποινή συντομίας ισούται με 1 σε περίπτωση που το μήκος του μηχανικού μεταφράσματος είναι μεγαλύτερο από το μήκος του κειμένου αναφοράς ενώ σε αντίθετη περίπτωση η ποινή υπολογίζεται με εκθετική μείωση.

Η μετρική BLEU βαθμολογεί μια μετάφραση σε κλίμακα από 0 έως 1. Όσο πιο κοντά βρίσκεται στο 1 η βαθμολογία, τόσο πιο κοντά βρίσκεται στην ανθρώπινη μετάφραση αναφοράς και τόσο καλύτερο θεωρείται αντίστοιχα το σύστημα μηχανικής μετάφρασης. Οι βαθμολογίες BLEU αναφέρονται συχνά και σε ποσοστιαία κλίμακα από 1 έως 100 για λόγους ευκολίας αλλά αυτό δεν πρέπει να συγχέεται με το ποσοστό ακρίβειας. Ένα μηχανικό μετάφρασμα θα σημείωνε βαθμολογία BLEU ίση με 1 μόνο εάν ήταν ολόιδιο με την ανθρώπινη μετάφραση αναφοράς. Ακόμα όμως και ικανοποιητικές ανθρώπινες μεταφράσεις του ίδιου κειμένου αφετηρίας θα μπορούσαν να βαθμολογηθούν στο εύρος του 0,6 ή του 0,7, καθώς είναι πιθανό να χρησιμοποιούν διαφορετικό λεξιλόγιο και φράσεις από τη μετάφραση αναφοράς. Πολύ υψηλές βαθμολογίες BLEU λοιπόν (άνω του 0,7) θα έπρεπε να εξετάζονται πιο προσεκτικά (Vashee 2019).

Η μετρική BLEU δίνει επίσης υψηλότερες βαθμολογίες για ακολουθίες λέξεων που απαντώνται τόσο στο μηχανικά μεταφρασμένο κείμενο όσο και στο κείμενο αναφοράς. Αν για παράδειγμα μια ακολουθία τεσσάρων λέξεων στο μηχανικό μετάφρασμα εντοπίζεται και στην ανθρώπινη μετάφραση αναφοράς, θα έχει μεγαλύτερο θετικό αντίκτυπο στη βαθμολογία BLEU από μια αντίστοιχη ακολουθία δύο λέξεων. Αυτό σημαίνει ότι μια ακριβής μετάφραση θα λάβει χαμηλότερη βαθμολογία εάν χρησιμοποιεί διαφορετικές, αλλά σωστές λέξεις ή τις ίδιες λέξεις, αλλά με διαφορετική σειρά.

3.3 NIST

Η NIST (Doddington 2002) είναι μια μετρική αυτόματης αξιολόγησης ποιότητας κειμένων που έχουν μεταφραστεί από συστήματα μηχανικής μετάφρασης. Το όνομά της προέρχεται από το Εθνικό Ινστιτούτο Προτύπων και Τεχνολογίας των ΗΠΑ (National Institute of Standards and Technology).

Βασίζεται όπως και η BLEU στην ακρίβεια με ορισμένες παραλλαγές. Ενώ η BLEU υπολογίζει την ακρίβεια ενός ν-γράμματος αποδίδοντας ίσο βάρος σε κάθε ένα από αυτά, η NIST υπολογίζει επιπλέον το πόση πληροφορία παρέχει ένα συγκεκριμένο ν-γράμμα. Όταν δηλαδή βρεθεί ένα αποδεκτό ν-γράμμα στο μετάφρασμα, όσο πιο σπάνια εμφανίζεται το ν-γράμμα, τόσο περισσότερο βάρος θα δοθεί σε αυτό. Για παράδειγμα, εάν το 2-γράμμα *on the* εντοπιστεί στο κείμενο, θα λάβει μικρότερο βάρος από το 2-γράμμα *interesting calculations*, καθώς το δεύτερο έχει πολύ λιγότερες πιθανότητες να εμφανιστεί στο κείμενο σε σχέση με το πρώτο.

Μια άλλη διαφορά της NIST από τη BLEU είναι ο τρόπος που υπολογίζουν την ποινή συντομίας. Ενώ στην BLEU μικρές παραλλαγές στο μήκος της μετάφρασης μπορούν να επηρεάσουν σημαντικά τη συνολική βαθμολογία, στη NIST η συνολική βαθμολογία επηρεάζεται σε χαμηλότερο βαθμό από μικρού εύρους παραλλαγές.

3.4 EditDistance

Η μετρική αυτόματης αξιολόγησης EditDistance που για λόγους συντομίας αναφέρεται και ως %EdDist αποτελεί υλοποίηση της απόστασης Levenshtein (Levenshtein, 1966). Η απόσταση Levenshtein μεταξύ δύο ακολουθιών δίνεται από τον ελάχιστο αριθμό λειτουργιών που απαιτούνται ώστε να μετασχηματιστεί η μία ακολουθία στην άλλη. Οι επιτρεπόμενες λειτουργίες είναι η εισαγωγή, η διαγραφή και η αντικατάσταση ενός χαρακτήρα.

Για παράδειγμα, η απόσταση μεταξύ των λέξεων *kitten* και *sitting* είναι 3 καθώς η μετατροπή της λέξης *kitten* στη λέξη *sitting* απαιτεί 3 λειτουργίες.

1. *kitten* -> *sitten* (αντικατάσταση *k* με *s*)
2. *sitten* -> *sittin* (αντικατάσταση *e* με *i*)
3. *sittin* -> *sitting* (εισαγωγή *g*)

Η μετρική %EdDist υπολογίζεται ως ο επί τοις εκατό ποσοστιαίος λόγος του συνόλου των λειτουργιών που απαιτούνται για τον μετασχηματισμό της ακολουθίας χαρακτήρων του μηχανικού μεταφράσματος στη μετάφραση αναφοράς προς τον συνολικό αριθμό χαρακτήρων του μηχανικού μεταφράσματος.

$$\%EdDist = 100 * \frac{edtotal}{chartotal}$$

όπου *edtotal* είναι το σύνολο των λειτουργιών που απαιτούνται για τον μετασχηματισμό του μηχανικά μεταφρασμένου κειμένου στο κείμενο αναφοράς και *chartotal* ο συνολικός αριθμός χαρακτήρων του μηχανικά μεταφρασμένου κειμένου.

3.5 WER

Η μετρική WER (Nießen et al., 2000) βασίζεται στην απόσταση Levenshtein (Levenshtein, 1966) - τον ελάχιστο αριθμό αντικαταστάσεων, διαγραφών και παρεμβολών που πρέπει να εκτελεστούν σε μια λέξη ώστε να ταυτιστεί πλήρως με μια άλλη λέξη. Στην περίπτωση της μηχανικής μετάφρασης, έχουμε να κάνουμε με τον ελάχιστο αριθμό αντικαταστάσεων, διαγραφών και παρεμβολών που πρέπει να εκτελεστούν σε ένα κείμενο που έχει παραχθεί από μια μηχανή αυτόματης μετάφρασης ώστε να μετατραπεί σε μια έγκυρη μετάφραση (ως έγκυρη ορίζεται η μετάφραση που έχει γίνει από ανθρώπινο μεταφραστή, η μετάφραση αναφοράς). Ο δείκτης σφαλμάτων λέξης (Word Error Rate), που είναι και αυτός που έχει δώσει το όνομά του στην εν λόγω μετρική, είναι ο λόγος του αθροίσματος των αντικαταστάσεων, διαγραφών και παρεμβολών που πρέπει να εκτελεστούν σε επίπεδο λέξεων προς το συνολικό αριθμό των λέξεων στο κείμενο αναφοράς.

$$WER = \frac{S + D + I}{N} = \frac{(S + D + I)}{S + D + C}$$

Όπου:

S είναι ο αριθμός των αντικαταστάσεων

D είναι ο αριθμός των διαγραφών

I είναι ο αριθμός των παρεμβολών

C είναι ο αριθμός των σωστών λέξεων

N είναι ο αριθμός των λέξεων στο κείμενο αναφοράς ($N = S + D + C$)

Πρέπει να σημειωθεί ότι η WER δεν επιτρέπει την αναδιάταξη των λέξεων.

3.6 TER

Η μετρική αυτόματης αξιολόγησης TER (Snover et al., 2006) αποτελεί παραλλαγή της WER. Βασίζεται στον αριθμό των ελάχιστων ενεργειών μετεπεξεργασίας που πρέπει να πραγματοποιηθούν σε ένα μηχανικά μεταφρασμένο κείμενο ώστε να ταυτιστεί με ένα κείμενο αναφοράς. Οι πιθανές ενέργειες μετεπεξεργασίας περιλαμβάνουν εισαγωγές, διαγραφές και αντικαταστάσεις μεμονωμένων λέξεων καθώς και μετατοπίσεις ακολουθιών λέξεων. Σε αυτό το σημείο εντοπίζεται και η βασική διαφορά με την WER καθώς η WER δεν επιτρέπει την αναδιάταξη των λέξεων και ολόκληρες φράσεις που είναι σωστές αλλά εντοπίζονται σε διαφορετικό σημείο μιας πρότασης σε σχέση με τη μετάφραση αναφοράς δεν λαμβάνονται υπόψη με συνέπεια την αύξηση της βαθμολογίας που υπολογίζει η μετρική.

Ο δείκτης σφαλμάτων μετάφρασης (Translation Edit Rate) υπολογίζει τον λόγο του ελάχιστου αριθμού ενεργειών μετεπεξεργασίας που χρειάζεται να πραγματοποιηθούν προς το σύνολο των λέξεων της μετάφρασης αναφοράς. Σε περίπτωση περισσότερων μεταφράσεων αναφοράς λαμβάνεται υπόψη ο μέσος όρος του αριθμού των λέξεων στις μεταφράσεις αναφοράς.

$$TER = \frac{\# \text{ ενεργειών μετεπεξεργασίας}}{M.O. \# \text{ λέξεων κειμένου αναφοράς}}$$

Όπως αναφέρθηκε και παραπάνω, οι πιθανές τροποποιήσεις περιλαμβάνουν την εισαγωγή, διαγραφή και αντικατάσταση μεμονωμένων λέξεων, καθώς και τη μετατόπιση ακολουθιών λέξεων. Μια μετατόπιση μετακινεί μια συνεχόμενη ακολουθία λέξεων εντός του εξεταζόμενου κειμένου σε ένα άλλο σημείο μέσα στο κείμενο. Όλες οι τροποποιήσεις, συμπεριλαμβανομένων των μετατοπίσεων οποιουδήποτε αριθμού λέξεων, από οποιαδήποτε απόσταση, έχουν ισοδύναμο κόστος. Επιπλέον, τα σημεία στίξης αντιμετωπίζονται ως κανονικές λέξεις και η εσφαλμένη χρήση κεφαλαίων ως κατάσταση που απαιτεί τροποποίηση.

4.

Μελέτη Παράλληλων Σωμάτων Κειμένου

Βασικό πυλώνα της παρούσας εργασίας αποτελεί η αξιολόγηση μηχανικά μεταφρασμένων κειμένων από τον χώρο των μέσων κοινωνικής δικτύωσης. Για να σχηματιστεί καλύτερη εικόνα και να βγουν πιο ολοκληρωμένα συμπεράσματα η αξιολόγηση επιλέχθηκε να πραγματοποιηθεί τόσο αυτόματα, με τη χρήση μετρικών αυτόματης αξιολόγησης, όσο και με ανθρώπινη παρέμβαση. Πιο συγκεκριμένα, χρησιμοποιήθηκε το εργαλείο MTUOC με τη βοήθεια του οποίου υπολογίστηκαν οι τιμές των μετρικών BLEU, NIST, %EdDist, WER και TER ενώ σε ανθρώπινο επίπεδο τα μηχανικά μεταφρασμένα κείμενα αξιολογήθηκαν ως προς την ακρίβεια (adequacy) και την ευχερεία (fluency) τους σύμφωνα με τα πρότυπα του DQF (Dynamic Quality Framework) και πραγματοποιήθηκε επίσης κατηγοριοποίηση των λαθών με βάση το πλαίσιο MQM (Multidimensional Quality Metrics).

4.1 Επιλογή κειμένων

Σχετικά με τα κείμενα προς μελέτη, η επιλογή ήταν τα παράλληλα σώματα κειμένου, κείμενα δηλαδή για τα οποία υπάρχουν διαθέσιμες μεταφράσεις και σε άλλες γλώσσες πέραν της πρωτότυπης. Το βασικό πλεονέκτημα της συγκεκριμένης επιλογής είναι φυσικά το γεγονός ότι οι διαθέσιμες μεταφράσεις των εν λόγω κειμένων μπορούν να χρησιμοποιηθούν ως κείμενα αναφοράς ώστε να μπορέσουν να λάβουν τιμές και να ενεργοποιηθούν οι μετρικές αυτόματης αξιολόγησης.

Αρχικός σκοπός ήταν η μελέτη κειμένων προερχόμενων αποκλειστικά από τα μέσα κοινωνικής δικτύωσης. Η σχετική αναζήτηση όμως έδειξε πως στην παρούσα φάση δεν υπάρχουν διαθέσιμα παρά ελάχιστα παράλληλα σώματα κειμένου που να έχουν συλλεχθεί από κοινωνικά δίκτυα και τα περισσότερα μάλιστα σε γλώσσες που η παρούσα εργασία δεν είχε τη δυνατότητα να υποστηρίξει. Εξαίρεση αποτέλεσε το *FooTweets* (Sluyter-Gäthje et al. 2018), μια συλλογή από 4000 tweets στην αγγλική γλώσσα που συλλέχθηκαν κατά τη διάρκεια του Παγκοσμίου Κυπέλλου Ποδοσφαίρου του 2014 και μεταφράστηκαν στα γερμανικά, ένα κομμάτι του οποίου μελετήθηκε και παρουσιάζεται στη συνέχεια.

Από κει και πέρα, έγινε προσπάθεια να γίνει επιλογή κειμένων με βάση τη θεματολογία που απασχολεί περισσότερο τους χρήστες των μέσων κοινωνικής δικτύωσης. Λόγω της ιδιαιτερότητας της εποχής και της πανδημίας που βρίσκεται σε εξέλιξη είναι προφανές ότι το πιο καυτό θέμα της συγκεκριμένης περιόδου είναι αυτή ακριβώς η κατάσταση για την οποία όμως δεν υπάρχουν διαθέσιμα παράλληλα σώματα κειμένου. Έτσι, τα κείμενα επιλέχθηκαν με βάση παλαιότερα στοιχεία που αφορούν εποχές «κανονικότητας» στην οποία όλοι ευελπιστούμε να επανέλθουμε σύντομα.

Στην παρακάτω εικόνα παρουσιάζονται τα πιο δημοφιλή θέματα στο Facebook και στο Twitter σύμφωνα με στοιχεία της Klout, του 2014.



Η επιλογή των κειμένων έγινε από δύο συλλογές παράλληλων σωμάτων κειμένου. Τα δύο κείμενα που αφορούν το ζεύγος γλωσσών ιταλικά-αγγλικά αντλήθηκαν από την πρώτη έκδοση του Paracrawl (Koehn et al. 2018), μια συλλογή από παράλληλα σώματα κειμένου στο διαδίκτυο που αφορούν τις ευρωπαϊκές γλώσσες και χρηματοδοτήθηκε από την Ευρωπαϊκή Ένωση (Espla-Gomis et al. 2019) και πιο συγκεκριμένα από τη Διευκόλυνση «Συνδέοντας την Ευρώπη» (Connecting Europe Facility). Τα δύο κείμενα που αφορούν το ζεύγος γλωσσών ιταλικά-ελληνικά αντλήθηκαν από το Parallel Global Voices (Prokorporidis et al. 2016), μια συλλογή παράλληλων σωμάτων κειμένου από την ομάδα ιστοσελίδων Global Voices, όπου εθελοντές μεταφράζουν και δημοσιεύουν ιστορίες πολιτών από όλο τον κόσμο από το 2004 και έπειτα.

4.2 Επιλογή Μηχανών Αυτόματης Μετάφρασης

Στην παρούσα εργασία χρησιμοποιήθηκαν δύο μηχανές αυτόματης μετάφρασης ώστε πέρα από την αξιολόγηση των μηχανικά μεταφρασμένων κειμένων να μπορέσει να γίνει και συγκριτική αξιολόγηση των δύο μηχανών. Οι μηχανές που επιλέχθηκαν ήταν το Google Translate το οποίο με βάση τη δημοτικότητά του είναι το πιο δημοφιλές εργαλείο αυτόματης μετάφρασης και το Facebook Translate, η μεταφραστική μηχανή του Facebook. Η επιλογή του εργαλείου του Facebook έγινε ακριβώς επειδή έχει αναπτυχθεί και λειτουργεί στο περιβάλλον ενός μέσου κοινωνικής δικτύωσης και, μάλιστα, του πιο δημοφιλούς παγκοσμίως.

Για τις ανάγκες της εργασίας δημιουργήθηκαν και 2 δοκιμαστικοί λογαριασμοί στο Facebook ώστε να μπορέσει να χρησιμοποιηθεί η μεταφραστική μηχανή του, καθώς δεν μπορεί να λειτουργήσει αυτόνομα εκτός του συγκεκριμένου περιβάλλοντος. Τα προς μελέτη κείμενα ανέβηκαν ως δημοσιεύσεις στη γλώσσα αφετηρίας στο χρονολόγιο του πρώτου δοκιμαστικού χρήστη και μεταφράστηκαν αυτόματα στη γλώσσα στόχο εμφανιζόμενα ως νέα στην αρχική σελίδα του δεύτερου χρήστη. Η επιλογή της γλώσσας στόχου έγινε με επεξεργασία των ρυθμίσεων γλώσσας του λογαριασμού του δεύτερου χρήστη.

4.3 Σχολιασμός FooTweets

Το *FooTweets* (Sluyter-Gäthje et al. 2018) είναι το μοναδικό παράλληλο σώμα κειμένου, στην παρούσα εργασία, όπου τα κείμενα που εξετάζονται προέρχονται από κάποιο μέσο κοινωνικής δικτύωσης και πιο συγκεκριμένα από το Twitter. Πρόκειται για μια συλλογή από 4000 tweets στην αγγλική γλώσσα που συλλέχθηκαν κατά τη διάρκεια του Παγκοσμίου Κυπέλλου Ποδοσφαίρου του 2014 και μεταφράστηκαν από ανθρώπους στα γερμανικά δίνοντας έμφαση στη φύση των συναισθημάτων που εκφράζουν κατηγοριοποιώντας τα σε αρνητικά, ουδέτερα και θετικά. Δημοσιεύτηκε στα πρακτικά του Ενδέκατου Διεθνούς Συνεδρίου Γλωσσικών Πόρων και Αξιολόγησης (Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)).

Το παρόν απόσπασμα κειμένου αποτελείται από μια σειρά από tweets με αρνητική χροιά. Τα γερμανόφωνα πλέον tweets του παράλληλου σώματος κειμένου μεταφράστηκαν εκ νέου στα αγγλικά με τη χρήση τόσο του Facebook Translate όσο και του Google Translate και στη συνέχεια πέρασαν τη διαδικασία αυτόματης αξιολόγησης με το εργαλείο MTUOC χρησιμοποιώντας ως κείμενο αναφοράς τα αντίστοιχα αγγλόφωνα tweets του παράλληλου σώματος κειμένου. Ταυτόχρονα, τα μεταφράσματα υποβλήθηκαν και σε ανθρώπινη αξιολόγηση ως προς την ακρίβεια και την ευχέρειά τους και πραγματοποιήθηκε κατηγοριοποίηση των λαθών.

Στον παρακάτω πίνακα παρατίθενται τα αποτελέσματα της αυτόματης αξιολόγησης:

Πίνακας 1

Αυτόματη Μετρική	Μηχανή Αυτόματης Μετάφρασης	
	Facebook Translate	Google Translate
BLEU	0.2578	0.3907
NIST	4.751	5.881
WER	0.5897	0.4337
%EdDist	28.71	25.11
TER	0.3342	0.3037

Παρατηρούμε ότι η βαθμολογία BLEU είναι κάτω του 50% και για τις δύο μηχανές μετάφρασης, με το Google Translate όμως να σημειώνει αισθητά καλύτερο σκορ και το Facebook Translate να βρίσκεται κάτω από το όριο του 30% το οποίο συχνά θεωρείται σημείο κλειδί για το αν και κατά πόσο μια μηχανική μετάφραση είναι αποδεκτή (Lavie 2011:41). Η διαφορά αυτή αποτυπώνεται και στη βαθμολογία NIST, μειωμένη όμως κατά 2 ποσοστιαίες μονάδες περίπου.

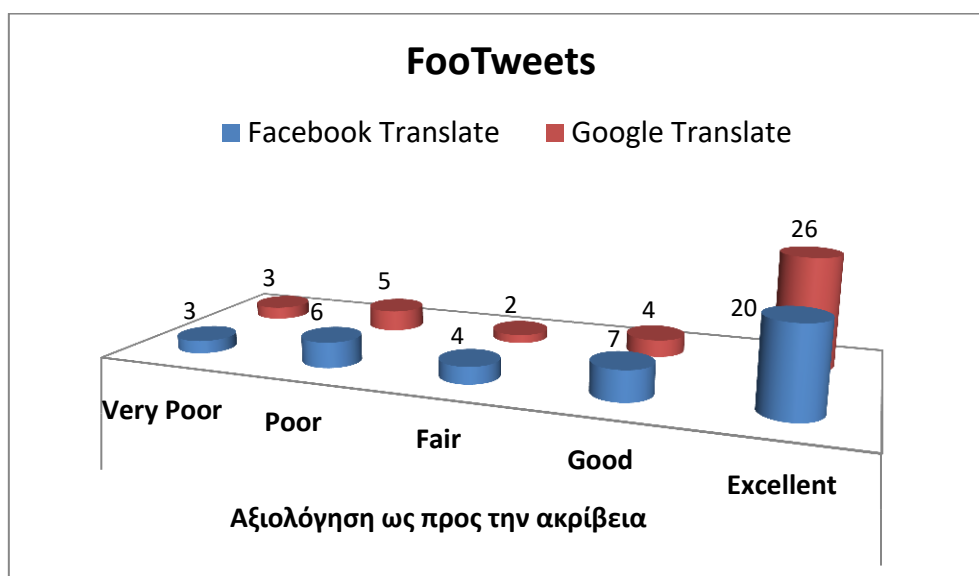
Όσον αφορά τις μετρικές που βασίζονται στην απόσταση διόρθωσης, κι εκεί παρατηρούμε ότι η μετάφραση μέσω Google Translate λαμβάνει καλύτερη αξιολόγηση με τον δείκτη σφαλμάτων λέξης (WER) αλλά και τον δείκτη σφαλμάτων μετάφρασης (TER) να είναι χαμηλότεροι σε σχέση με τους αντίστοιχους δείκτες για τη μετάφραση μέσω Facebook Translate. Η μετρική %EdDist που μετρά την απόσταση διόρθωσης είναι επίσης χαμηλότερη για το Google Translate επιβεβαιώνοντας όλες τις παραπάνω εκτιμήσεις.

Λαμβάνοντας υπόψη το γεγονός ότι το σώμα κειμένου προέρχεται από ένα μέσο κοινωνικής δικτύωσης, ακόμα κι αν αυτό δεν είναι το Facebook, θα περίμενε κανείς η μεταφραστική μηχανή του Facebook να σημειώνει καλύτερες επιδόσεις καθώς τα κείμενα στα μέσα κοινωνικής δικτύωσης παρουσιάζουν ιδιαιτερότητες όπως τα hashtags, τα cashtags, τα διάφορα εικονίδια κτλ. Ειδικά στο Twitter, από όπου προέρχεται το συγκεκριμένο απόσπασμα, υπάρχει και ο περιορισμός των 140 χαρακτήρων για κάθε tweet, γεγονός που οδηγεί πολλούς χρήστες στο να γράφουν χρησιμοποιώντας συντομογραφίες, όπως για παράδειγμα το γράμμα *u* αντί της λέξης *you*. Αξίζει να σημειωθεί ότι συντομογραφίες αυτού του τύπου χρησιμοποιούνται γενικότερα στα μέσα κοινωνικής δικτύωσης και όχι μόνο όταν υπάρχει περιορισμός στο όριο των χαρακτήρων. Η γλώσσα που χρησιμοποιείται επίσης είναι πιο κοντά στην καθημερινή καθομιλουμένη γλώσσα παρά στην επίσημη γλώσσα, γεγονός που προσδίδει ιδιαίτερο ενδιαφέρον στην αυτόματη μετάφρασή της.

Περνώντας στην ανθρώπινη αξιολόγηση ως προς την ακρίβεια και την ευχέρεια των δύο μεταφρασμάτων παρατίθενται συγκεντρωτικά κάποια αποτελέσματα στους παρακάτω δύο πίνακες όπου παρουσιάζεται το σύνολο των τμημάτων κειμένου και το πώς αξιολογήθηκαν:

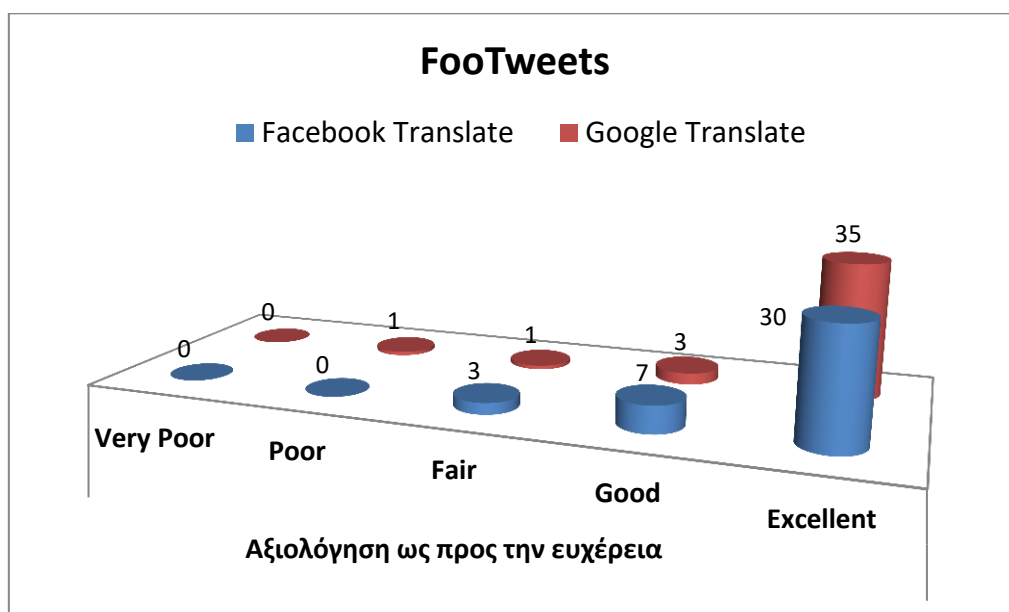
Πίνακας 2

Μηχανή Μετάφρασης	Αξιολόγηση ως προς την ακρίβεια				
	Very Poor	Poor	Fair	Good	Excellent
Facebook Translate	3	6	4	7	20
Google Translate	3	5	2	4	26



Πίνακας 3

Μηχανή Μετάφρασης	Αξιολόγηση ως προς την ευχέρεια				
	Very Poor	Poor	Fair	Good	Excellent
Facebook Translate	0	0	3	7	30
Google Translate	0	1	1	3	35



Πρέπει να σημειωθεί ότι κατά τη διαδικασία της ανθρώπινης αξιολόγησης λήφθηκαν υπόψη οι ιδιαιτερότητες του συγκεκριμένου κειμένου και περιπτώσεις προτάσεων που υπό άλλες συνθήκες θα θεωρούνταν λανθασμένες ή μη αποδεκτές, όπως πχ προτάσεις που ξεκινούν με μικρά γράμματα, στην εν λόγω περίπτωση κρίθηκαν ως αποδεκτές.

Ξεκινώντας από τον Πίνακα 3 όπου παρουσιάζεται η αξιολόγηση ως προς την ευχέρεια παρατηρούμε ότι σε γενικές γραμμές και οι δύο μεταφράσεις σε αυτό το κομμάτι κρίνονται πολύ καλές καθώς από τα 40 τμήματα κειμένου ως εξαιρετικά κρίνονται τα 30 για το Facebook Translate και τα 35 για το Google Translate. Όπως και στην αυτόματη αξιολόγηση, έτσι και σε αυτόν τον τομέα η μηχανή αυτόματης μετάφρασης της Google μοιάζει να υπερτερεί σε σχέση με την αντίστοιχη του Facebook. Παρόλα αυτά είναι αξιοσημείωτο το γεγονός ότι κανένα τμήμα κειμένου που μεταφράστηκε με το Facebook Translate δεν κρίθηκε φτωχό ως προς την ευχέρειά του, σε αντίθεση με το Google Translate όπου ένα τμήμα αξιολογήθηκε με τον συγκεκριμένο χαρακτηρισμό. Πρόκειται για το κομμάτι *warte was ? nicht , wenn er in den Ruhestand gehen muss # Gerrard # Worldcup* κατά τη μετάφραση του οποίου μέσω Google Translate το ερωτηματικό παραλείπεται με αποτέλεσμα να προκύπτει η συγκεκριμένη φράση *wait what not if he has to retire # Gerrard # Worldcup*. Κατά τα άλλα δεν παρουσιάστηκαν ιδιαίτερα προβλήματα.

Στο κομμάτι της ακρίβειας καταγράφηκαν περισσότερα λάθη ενώ και πάλι το Google Translate φαίνεται να υπερτερεί παρουσιάζοντας 26 τμήματα που κρίνονται ως εξαιρετικά ενώ

το Facebook παρουσιάζει αντίστοιχα 20 όπως φαίνεται στον Πίνακα 2. Σε αντίθεση με αυτά της ευχέρειας, τα σφάλματα που αφορούν την ακρίβεια ήταν πιο σοβαρά και σε κάποιες περιπτώσεις άλλαξαν ριζικά το νόημα του κειμένου. Και οι δύο μηχανές παρουσίασαν από 3 τμήματα κειμένου με πολύ φτωχή ακρίβεια. Στο Google Translate και οι 3 περιπτώσεις αφορούν την ποδοσφαιρική ορολογία, στις 2 μάλιστα πρόκειται για το ίδιο σφάλμα όπου η λέξη *Tor* μεταφράζεται εσφαλμένα ως *gate* με αποτέλεσμα να χαθεί το νόημα της πρότασης αφού θα έπρεπε να μεταφραστεί κανονικά ως *goal*. Η άλλη περίπτωση σφάλματος ορολογίας είναι η λέξη *Abseits* η οποία αντί για *offside* μεταφράζεται ως *apart* στερώντας έτσι από την αντίστοιχη πρόταση το νόημά της. Στο Facebook Translate μία εκ των περιπτώσεων αφορά την ορολογία και είναι η ίδια με αυτή του Google Translate με τις λέξεις *Tor* και *goal*. Οι άλλες δύο όμως αφορούν σοβαρά σφάλματα λανθασμένης μετάφρασης λέξεων. Η πρώτη αφορά τη φράση *Frankreich verschießt einen Elfmeter* που αποδίδεται εσφαλμένα ως *France shoots a penalty* μπερδεύοντας τη λέξη *schießen* που είναι η γερμανική αντίστοιχη της αγγλικής *shoot* με τη λέξη *verschießen* που εκφράζει απώλεια. Στην πραγματικότητα πρόκειται για απώλεια του πέναλτι και θα έπρεπε να αποδοθεί ως *France misses a penalty*. Η άλλη περίπτωση αφορά τη φράση *ich hätte nie gedacht, dass Spanien jemals in einem Spiel so unüberzeugend aussehen würden* που αποδίδεται ως *i never thought Spain would ever look so unsurpassing in a game* παρουσιάζοντας μια τελείως διαφορετική από την πραγματική εικόνα της Ισπανίας στο παιχνίδι. Η λέξη *unüberzeugend* έχει αρνητική χροιά και θα έπρεπε να μεταφραστεί ως *unconvincing*. Αποδίδεται αντίθετα ως *unsurpassing* δίνοντας την εικόνα μιας Ισπανίας που κυριαρχεί. Λαμβάνοντας μάλιστα υπόψη ότι στο συγκεκριμένο απόσπασμα όλα τα tweets έχουν αρνητική χροιά και η μηχανική μετάφραση δίνει ένα tweet με πολύ θετική χροιά, το συγκεκριμένο σφάλμα αναδεικνύεται ακόμα πιο σοβαρό.

Πέρα από τα παραπάνω, καταγράφηκαν και λιγότερο σημαντικά λάθη με το Facebook Translate να έχει 6 τμήματα κειμένου έναντι 5 του Google Translate όπου η ακρίβεια αξιολογήθηκε ως φτωχή. Το χαρακτηριστικό σε αυτήν την κατηγορία είναι ότι τα σφάλματα ταυτίζονται σε σημαντικό βαθμό αφού πρόκειται για τα ίδια 5 τμήματα κειμένου τα οποία και οι δύο μηχανές χειρίστηκαν με παρόμοιο τρόπο. Χαρακτηριστικό παράδειγμα η φράση *ich habe dieses Spiel nur geguckt, um mich über das Team, das verliert, lustig zu machen* που αποδόθηκε και από τις δύο μηχανές ως *I just watched this game to make fun of the team that was losing* ενώ η σωστή απόδοση θα ήταν ως *I only watched this game to make fun of the team that was losing* καθώς η χρήση του *just* αφορά περισσότερο το χρόνο κατά τον οποίο παρακολούθησε ο συντάκτης του εν λόγω tweet το παιχνίδι και όχι τον λόγο για τον οποίο το παρακολούθησε ο οποίος για να αναδειχτεί πρέπει να χρησιμοποιηθεί το *only* αντί του *just*.

Από κει και πέρα, προχωρώντας προς τα τμήματα του κειμένου με υψηλότερη αξιολόγηση ως προς την ακρίβεια, το Google Translate εμφανίζει περισσότερα τμήματα με εξαιρετική αξιολόγηση με το Facebook Translate να υπολείπεται ελαφρώς έχοντας αντίστοιχα περισσότερα με καλή και μέτρια αξιολόγηση. Τα σφάλματα σε αυτές τις κατηγορίες είναι λιγότερο σημαντικά και δεν ταυτίζονται στις δύο μηχανές αυτόματης μετάφρασης. Παράδειγμα αποτελεί η φράση *Totalschaden für Spanien* που το Facebook Translate την αποδίδει ως *total damage to Spain* και το Google Translate ως *total loss for Spain*. Αν και το νόημα αποδίδεται και στις δύο εκδοχές, είναι προφανές ότι απόδοση του Google Translate είναι πιο

ορθή, μιας και πρόκειται για ποδόσφαιρο, ενώ η αντίστοιχη του Facebook υστερεί. Υπάρχουν επίσης περιπτώσεις όπου οι διαφορές είναι ελάχιστες και βρίσκονται στις λεπτομέρειες. Χαρακτηριστικό παράδειγμα η λέξη *Schiri* που χρησιμοποιείται κυρίως στον προφορικό λόγο ως συντόμευση της λέξης *Schiedsrichter* που σημαίνει *διαιτητής*. Το Google Translate την αποδίδει ως *referee* ενώ το Facebook Translate ως *ref*. Και οι δύο είναι σωστές φυσικά, η *ref* όμως μοιάζει κατά κάποιο τρόπο «ορθότερη» καθώς πρόκειται για συντόμευση.

Ολοκληρώνοντας την ανθρώπινη αξιολόγηση και αντιπαραβάλλοντας τα αποτελέσματά της με αυτά της αυτόματης αξιολόγησης είναι φανερό ότι η υπεροχή του Google Translate έναντι του Facebook Translate όπως προκύπτει από τις αυτόματες μετρικές επιβεβαιώνεται. Αξίζει να σημειωθεί όμως ότι οι διαφορές δεν είναι τόσο μεγάλες όσο μοιάζουν να είναι, ειδικά όσον αφορά τη BLEU όπου το Facebook Translate εμφανίζει βαθμολογία χαμηλότερη του 30%. Η εξήγηση για το χαμηλό σκορ βρίσκεται στο κείμενο αναφοράς. Επειδή ακριβώς πρόκειται για συλλογή tweets, σε πολλά από αυτά υπάρχουν παραλείψεις τύπου *havent* αντί *haven't*, *Im* αντί *I'm*, συντομογραφίες τύπου *u* αντί *you*, ιδιωματικές λέξεις τύπου *lil* αντί για *little* κτλ. Όλες αυτές μεταφράστηκαν σε τυπική γλώσσα με προσέγγιση *informal to formal* (Sluyter-Gäthje et al. 2018: 2667) κατά την μετάφραση των πρωτότυπων αγγλόφωνων tweets στα γερμανικά. Η μηχανική μετάφραση των γερμανόφωνων πλέον tweets στα αγγλικά προφανώς και θα δώσει αποτελέσματα σε τυπική γλώσσα χωρίς να εμφανίζονται στο μεταφρασμένο κείμενο φαινόμενα όπως αυτά που περιγράφηκαν παραπάνω. Η αυτόματη αξιολόγηση των μηχανικών μεταφράσεων με τη χρήση των πρωτότυπων αγγλόφωνων tweets ως κείμενο αναφοράς θα καταγράψει αυτά τα φαινόμενα ως διαφορές και θα τα αποτυπώσει φυσικά στις βαθμολογίες δίνοντας μια εικόνα χαμηλότερης ποιότητας μετάφρασης σε σχέση με την πραγματική ποιότητα.

Μια λύση για περισσότερο αξιόπιστες βαθμολογίες στην αυτόματη αξιολόγηση θα ήταν η επεξεργασία του κειμένου αναφοράς και η μετατροπή του σε τυπική γλώσσα εξαλείφοντας τα συγκεκριμένα φαινόμενα. Στα πλαίσια της παρούσας εργασίας δεν κρίθηκε σκόπιμο να γίνει κάτι τέτοιο καθώς αντικείμενό της αποτελεί η σύγκριση των δύο μηχανών αυτόματης μετάφρασης. Ως προς αυτού του είδους τα φαινόμενα και οι δύο μηχανές έχουν την ίδια αντιμετώπιση οπότε το μόνο που θα άλλαζε ενδεχομένως θα ήταν το να παρουσιάσουν και οι δύο καλύτερα σκορ, χωρίς να αλλάζει κάτι στη σύγκριση μεταξύ τους.

Από τα παραπάνω όμως βγαίνει ένα σημαντικό συμπέρασμα που αξίζει να επισημανθεί και αφορά τη χρήση παράλληλων κειμένων για την εκπαίδευση συστημάτων μηχανικής μετάφρασης. Οι αυτόματες μετρικές άλλωστε χρησιμεύουν και αξιοποιούνται κατά κύριο λόγο κατά την ανάπτυξη των εν λόγω συστημάτων. Ενώ σε άλλα κειμενικά είδη η ύπαρξη παράλληλων κειμένων αρκεί ενδεχομένως ώστε να εκπαιδευτεί ένα σύστημα και προς τις δύο κατευθύνσεις, για παράδειγμα από τα αγγλικά προς τα γερμανικά αλλά και από τα αγγλικά προς τα γερμανικά, όταν τα κείμενα έχουν συλλεχθεί από μέσα κοινωνικής δικτύωσης υπάρχει διαφοροποίηση ανάλογα με την κατεύθυνση προς την οποία πραγματοποιείται η μετάφραση. Για τον λόγο αυτό, πρέπει οι ιδιαιτερότητες αυτού του είδους κειμένων να λαμβάνονται σοβαρά υπόψη κατά την ανάπτυξη των συστημάτων ώστε να αποφεύγονται τυχόν λανθασμένα συμπεράσματα και οι όποιες συνέπειες ενδέχεται να τα συνοδεύουν.

4.4 Σχολιασμός Telework

Το απόσπασμα κειμένου στο οποίο για τις ανάγκες της παρούσας εργασίας δόθηκε ο τίτλος *Telework* αντλήθηκε από την πρώτη έκδοση του ParaCrawl που δημοσιεύτηκε τον Ιανουάριο του 2018. Αν και αναφέρεται σε μια πρωτοβουλία που υλοποιήθηκε αρκετά χρόνια πριν, μεταξύ 2001 και 2003, επιλέχθηκε επειδή η θεματολογία του είναι επίκαιρη όσο ποτέ. Αναφέρεται φυσικά στην τηλεργασία, η οποία μπορεί την εποχή που γράφτηκε το εν λόγω κείμενο να ήταν ακόμα σε πολύ πρώιμο στάδιο, εν έτει 2020 όμως και εν μέσω της πανδημίας του κορωνοϊού αποτελεί μέρος της καθημερινότητας. Συγκεντρώνει λοιπόν τεράστιο ενδιαφέρον και απασχολεί πολλούς ανθρώπους οι οποίοι μπορούν να παρακολουθούν συζητήσεις, να ανταλλάσσουν απόψεις και να ενημερώνονται για ζητήματα σχετικά με την τηλεργασία και μέσα από τα μέσα κοινωνικής δικτύωσης. Μιας και πρόκειται για θέμα με παγκόσμιο ενδιαφέρον μπορούν να γίνουν συζητήσεις σε πολλές γλώσσες τις οποίες θα μπορούσε κάποιος να παρακολουθήσει και με τη βοήθεια των μηχανών αυτόματης μετάφρασης.

Η μελέτη του συγκεκριμένου αποσπάσματος κειμένου έχει ως στόχο την εξέταση αυτής ακριβώς της «γλώσσας της τηλεργασίας» και του τρόπου χειρισμού της από τις μεταφραστικές μηχανές. Η Ιταλία ήταν μία από τις χώρες που συμμετείχαν στην πρωτοβουλία και το ιταλόφωνο κείμενο μεταφράστηκε στα αγγλικά με τη βοήθεια τόσο του Facebook Translate όσο και του Google Translate. Τα μεταφράσματα αξιολογήθηκαν στη συνέχεια αυτόματα με τη βοήθεια του εργαλείου MTUOC χρησιμοποιώντας ως κείμενο αναφοράς το αντίστοιχο αγγλόφωνο απόσπασμα κειμένου από το ParaCrawl. Τα αποτελέσματα της αυτόματης αξιολόγησης παρατίθενται στον παρακάτω πίνακα:

Πίνακας 4

Αυτόματη Μετρική	Μηχανή Αυτόματης Μετάφρασης	
	Facebook Translate	Google Translate
BLEU	0.2989	0.3579
NIST	5.4597	5.778
WER	0.4999	0.49
%EdDist	40.92	38.62
TER	0.4681	0.4409

Παρατηρούμε ότι το Google Translate φαίνεται να έχει υπεροχή σε όλες τις κατηγορίες. Ειδικά στη βαθμολογία BLEU η διαφορά των δύο μεταφραστικών μηχανών σχεδόν αγγίζει τις 6 ποσοστιαίες μονάδες με 35,79% έναντι 29,89% αντίστοιχα ενώ η βαθμολογία του Facebook Translate όπως βλέπουμε βρίσκεται οριακά κάτω από το 30%. Η διαφορά των δύο μεταφραστικών μηχανών αποτυπώνεται και στη βαθμολογία NIST, μειωμένη όμως σχεδόν κατά το ήμισυ σε επίπεδο ποσοστιαίων μονάδων. Με 54,597% και 57,78% αντίστοιχα η διαφορά τους ανέρχεται σε κάτι παραπάνω από 3 ποσοστιαίες μονάδες..

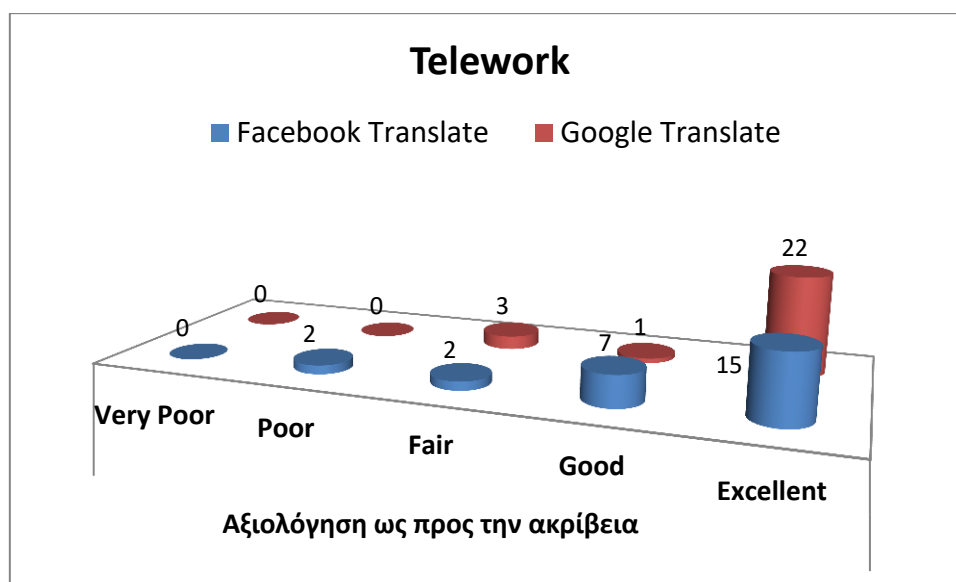
Περνώντας στις μετρικές που βασίζονται στην απόσταση διόρθωσης, παρατηρούμε ότι και σε αυτό το κομμάτι η μετάφραση μέσω Google Translate λαμβάνει καλύτερη αξιολόγηση. Η μετρική %EdDist που μετρά την απόσταση διόρθωσης είναι χαμηλότερη για το Google

Translate με τιμή 38,62 έναντι 40,92 επιβεβαιώνοντας την εκτίμηση των BLEU και NIST περί ποιοτικότερης μετάφρασης της εν λόγω μηχανής. Επιπλέον, ο δείκτης σφαλμάτων λέξης (WER) αλλά και ο δείκτης σφαλμάτων μετάφρασης (TER) είναι επίσης χαμηλότεροι σε σχέση με τους αντίστοιχους δείκτες για τη μετάφραση μέσω Facebook Translate. Με τιμές 49% και 49,99% για τον WER και 44,09% και 46,81% για τον TER για Google Translate και Facebook Translate αντίστοιχα, η μεταφραστική μηχανή της Google υπερτερεί και σε αυτήν την κατηγορία.

Περνώντας στην ανθρώπινη αξιολόγηση ως προς την ακρίβεια και την ευχέρεια των δύο μεταφρασμάτων θα γίνει προσπάθεια να διαπιστωθεί αν η υπεροχή που μοιάζει να έχει το Google Translate με βάση την αυτόματη αξιολόγηση βρίσκει σύμφωνη και την ανθρώπινη κρίση. Τα συγκεντρωτικά αποτελέσματα για το σύνολο των τμημάτων κειμένου και τον τρόπο με τον οποίο αξιολογήθηκαν παρατίθενται στους παρακάτω δύο πίνακες και στα αντίστοιχα γραφήματα:

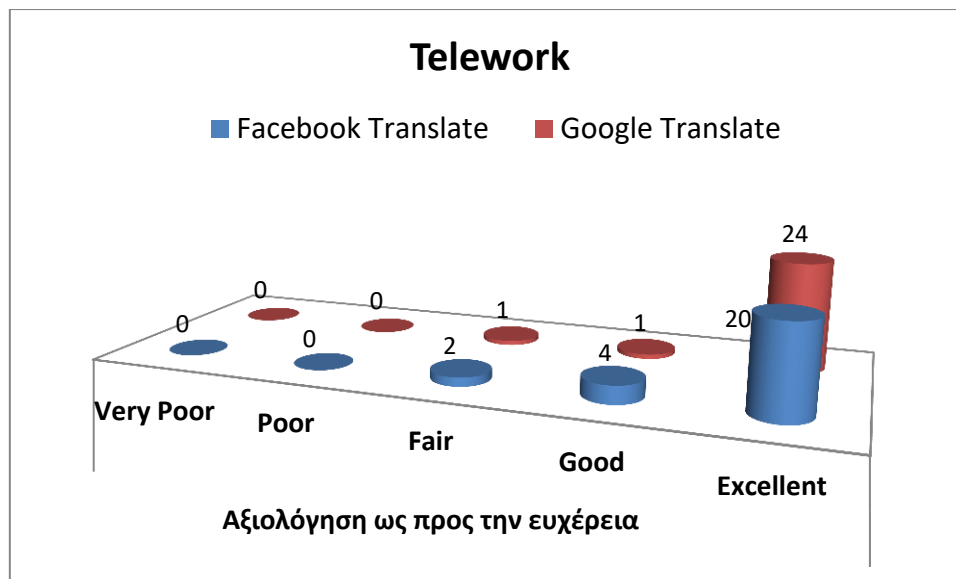
Πίνακας 5

Μηχανή Μετάφρασης	Αξιολόγηση ως προς την ακρίβεια				
	Very Poor	Poor	Fair	Good	Excellent
Facebook Translate	0	2	2	7	15
Google Translate	0	0	3	1	22



Πίνακας 6

Μηχανή Μετάφρασης	Αξιολόγηση ως προς την ευχέρεια				
	Very Poor	Poor	Fair	Good	Excellent
Facebook Translate	0	0	2	4	20
Google Translate	0	0	1	1	24



Στον Πίνακα 3, όπου παρουσιάζονται τα αποτελέσματα της αξιολόγησης ως προς την ευχέρεια, παρατηρούμε ότι σε κανένα από τα δύο μεταφράσματα δεν καταγράφονται τμήματα κειμένου με αρνητική αξιολόγηση, συνέπεια βεβαίως της διαρκούς βελτίωσης που παρουσιάζουν τα συστήματα νευρωνικής μετάφρασης στον συγκεκριμένο τομέα. Συγκριτικά, η υπεροχή του Google Translate επιβεβαιώνεται και στο κομμάτι της ευχέρειας καθώς σε σύνολο 26 τμημάτων καταγράφονται 24 με άριστη αξιολόγηση έναντι 20 του Facebook Translate το οποίο βεβαίως παρουσιάζει και αυτό πολύ καλή ανταπόκριση. Σε γενικές γραμμές δεν εμφανίζονται ιδιαίτερα προβλήματα καθώς το πιο «σοβαρό» είναι η εμφάνιση της φράσης *organization s'* στο μετάφρασμα του Facebook Translate όπου θα έπρεπε είτε η απόστροφος να βρίσκεται ανάμεσα στη λέξη *organization* και το γράμμα *s* ως *organization's* είτε το *organizations* να είναι μία λέξη με την απόστροφο στο τέλος της λέξης ως *organizations'*.

Στο κομμάτι της αξιολόγησης ως προς την ακρίβεια, τα αποτελέσματα της οποίας παρουσιάζονται στον Πίνακα 2, παρατηρούμε και πάλι υπεροχή του Google Translate οποίο εμφανίζει 22 τμήματα με εξαιρετική αξιολόγηση έναντι 15 του Facebook Translate. Στο σύνολο πάντως των τμημάτων με θετική αξιολόγηση η διαφορά δεν είναι τόσο μεγάλη καθώς στο Facebook Translate καταγράφονται 7 τμήματα κειμένου με καλή αξιολόγηση ως προς την ακρίβεια έναντι 1 του Google Translate. Έτσι, εμφανίζονται συνολικά 23 τμήματα με θετική αξιολόγηση για τη μεταφραστική μηχανική της Google έναντι 22 για την αντίστοιχη του Facebook.

Πέρα όμως από τα τμήματα με εξαιρετική αξιολόγηση που είναι περισσότερα για το μετάφρασμα του Google Translate, πολύ σημαντικό είναι και το γεγονός ότι για την εν λόγω μηχανή δεν καταγράφηκαν καθόλου τμήματα κειμένου με αρνητική αξιολόγηση ως προς την ακρίβεια της μετάφρασης. Στο μετάφρασμα του Facebook αντίθετα παρουσιάστηκαν δύο τμήματα με φτωχή αξιολόγηση. Αξίζει πάντως να σημειωθεί ότι και το Google Translate αντιμετώπισε πρόβλημα στη μετάφραση αυτών των τμημάτων, τα αντιμετώπισε όμως σχετικά καλύτερα και η μετάφρασή τους αξιολογήθηκε ως μέτρια.

Σχετικά με τα προβλήματα που αντιμετωπίστηκαν, και οι δύο μεταφραστικές μηχανές για παράδειγμα είχαν πρόβλημα στην απόδοση στην αγγλική γλώσσα κάποιων θεσμών όπως τα *Istituti Scolastici Superiori* και οι *Organizzazioni di Base degli operatori della formazione*. Πρόκειται για σχολικά ιδρύματα «ανώτερης δευτεροβάθμιας εκπαίδευσης» στην πρώτη περίπτωση (κάτι σαν το ελληνικό Λύκειο, ως «ανώτερη βαθμίδα» από το Γυμνάσιο, ή τα ΙΕΚ που στην Ελλάδα είναι αδιαβάθμητα) και οργανισμούς που έχουν ως βάση τους ανθρώπους που παρέχουν εκπαιδευτικές υπηρεσίες στη δεύτερη, εκπαιδευτικούς οργανισμούς δηλαδή στην ουσία. Στη δεύτερη περίπτωση και οι δύο μηχανές αποδίδουν τους οργανισμούς ως *Basic Organizations of Training Operators* που μοιάζει μάλλον περίπλοκη και χωρίς να αποδίδει ακριβώς το σωστό νόημα. Ο πρώτος θεσμός αποδίδεται ως *Institutes Superior Schools* από το Facebook Translate και ως *Schools Superiors* από το Google Translate. Καμία από τις δύο αποδόσεις φυσικά δεν μπορεί να θεωρηθεί ορθή.

Άλλο ένα σημείο με ιδιαίτερη δυσκολία ήταν η απόδοση της φράσης *momento educativo* που μεταφράστηκε και από τις δύο μηχανές αυτολεξεί ως *educational moment*. Η εν λόγω φράση έρχεται ως επεξήγηση στη διαφορά μεταξύ *formazione al telelavoro* και *educazione al telelavoro* που αποδίδονται ορθά και από τις δύο μηχανές ως *telework training* η πρώτη και ως *telework* ή *teleworking* (από Google και Facebook αντίστοιχα) *education* η δεύτερη. Ουσιαστικά προσπαθεί να αναδείξει τη διαφορά μεταξύ εκπαίδευσης και παιδείας σχετικά με την τηλεργασία και το *momento educativo* αναφέρεται ακριβώς στο σημείο που γίνεται αυτός ο διαχωρισμός καθώς η παιδεία είναι ευρύτερη έννοια. Μια πρόταση για ορθότερη μετάφραση θα ήταν το *educational character* ενώ στο κείμενο αναφοράς εμφανίζεται ως *educational event*.

Από κει και πέρα, το Facebook Translate παρουσίασε κάποια περαιτέρω προβλήματα αδυνατώντας για παράδειγμα να μεταφράσει κατάλληλα κάποιες λέξεις με παρόμοιο νόημα μεταξύ τους. Πιο συγκεκριμένα, η φράση *efficace ed efficiente* αποδόθηκε ως *efficient and efficient*. Αν η λέξη *efficace* ήταν μόνη της στο κείμενο δεν θα υπήρχε πρόβλημα να αποδοθεί ως *efficient*, στην προκειμένη περίπτωση όμως παρατηρούμε πλεονασμό αφού ακολουθείται από τη λέξη *efficiente* που ταυτίζεται απόλυτα με τη λέξη *efficient*. Η απόδοση από το Google Translate ως *effective and efficient* κρίνεται εξαιρετική. Παρόμοια περίπτωση είναι και η απόδοση της φράσης *sistema formativo ed educativo* ως *educational and educational system* από το Facebook Translate. Το συγκεκριμένο σφάλμα μπορεί να θεωρηθεί και πιο σοβαρό από το προηγούμενο αν λάβουμε υπόψη τη διαφορά μεταξύ εκπαίδευσης και παιδείας που προσπαθεί να αναδείξει το κείμενο και για την οποία έγινε λόγος στην προηγούμενη παράγραφο. Και σε αυτή την περίπτωση το Google Translate αποδίδει τη φράση ορθά ως *training and educational system*. Μια τρίτη (όχι φαρμακερή αλλά...) χαρακτηριστική περίπτωση είναι η φράση *supporto all'ideazione, progettazione, realizzazione di nuovi moduli formativi* που το Facebook Translate αποδίδει ως *support for the idea, planning, the creation of new training modules* και το Google Translate ως *support for the conception, design, implementation of new training modules*. Είναι προφανές ότι η μετάφραση του Facebook πάσχει, χωρίς να είναι πολύ μακριά στο θέμα της ακρίβειας, και σε καμία περίπτωση δεν μπορεί να χαρακτηριστεί ως θετική.

Παρά την υπεροχή του Google Translate θα πρέπει να σημειωθεί ότι υπήρξαν και περιπτώσεις τις οποίες χειρίστηκε καλύτερα το Facebook Translate. Πιο χαρακτηριστική είναι η

φράση *uno degli aspetti che le stesse organizzazioni devono curare* η οποία αποδόθηκε από το Google ως *one of the aspects that the same organizations must take care of* μιλώντας για ίδιους οργανισμούς που πρέπει να φροντίσουν κάποια θέματα ενώ η ουσία της πρότασης είναι ότι ο κάθε οργανισμός πρέπει ο ίδιος να φροντίσει τα θέματά του. Η συγκεκριμένη φράση αποδόθηκε πολύ ορθά από το Facebook Translate ως *one of the aspects that the organizations themselves need to take care of*.

Ανακεφαλαιώνοντας, μπορεί κανείς να πει ότι στο συγκεκριμένο απόσπασμα κειμένου οι εκτιμήσεις των αυτόματων μετρικών συμφωνούν σε μεγάλο βαθμό με τις ανθρώπινες εκτιμήσεις. Η μεταφραστική μηχανή της Google υπερέχει σε όλα τα επίπεδα σε σχέση με την αντίστοιχη του Facebook όπως αποτυπώνεται στους πίνακες τόσο της αυτόματης όσο και της ανθρώπινης αξιολόγησης. Θα περίμενε κανείς οι διαφορές να είναι μικρότερες μιας και πρόκειται για ένα θέμα αρκετά επίκαιρο. Λαμβάνοντας πάντως υπόψη το γεγονός ότι από τη μία οι μηχανές αυτόματης μετάφρασης εξελίσσονται και εκπαιδεύονται συνεχώς και από την άλλη το γεγονός ότι η τηλεργασία αποτελεί πλέον θέμα πρώτης γραμμής, θα παρουσίαζε ενδιαφέρον η εκ νέου μετάφραση και των δύο κειμένων από τις ίδιες μεταφραστικές μηχανές μετά από ένα εύλογο χρονικό διάστημα ώστε να διαπιστωθεί αν έχει βελτιωθεί η ποιότητα της μετάφρασης και αν έχουν αμβλυνθεί οι διαφορές μεταξύ τους.

4.5 Σχολιασμός *La Cina e il vino italiano*

Το απόσπασμα κειμένου στο οποίο για τις ανάγκες της παρούσας εργασίας αποδόθηκε ο τίτλος *La Cina e il vino italiano* αντλήθηκε από την πρώτη έκδοση του ParaCrawl που δημοσιεύτηκε τον Ιανουάριο του 2018. Η επιλογή του έγινε με βάση τη θεματολογία του που σχετίζεται με τη γαστρονομία μιας και η ιταλική κουζίνα είναι από τις πιο δημοφιλείς παγκοσμίως και εκατομμύρια άνθρωποι στον πλανήτη δηλώνουν λάτρεις της. Πρόκειται για μια συζήτηση με μια κοπέλα με σπουδές στην κινεζική γλώσσα και στους πολιτισμούς της Ανατολής η οποία εργάζεται στον χώρο του κρασιού. Μια από τις δραστηριότητές της είναι οι ξεναγήσεις σε οινοποιεία όπου παρέχει διάφορες πληροφορίες σε αλλοδαπούς επισκέπτες, προς το παρόν στην αγγλική γλώσσα, φιλοδοξώντας στην πορεία να εντάξει και την κινεζική γλώσσα στις υπηρεσίες της.

Η συζήτηση γίνεται στα ιταλικά, σε απλή, καθημερινή γλώσσα την οποία θα μπορούσε να παρακολουθήσει οποιοσδήποτε άνθρωπος στον κόσμο με τη βοήθεια μιας μηχανής αυτόματης μετάφρασης. Στην παρούσα εργασία χρησιμοποιήθηκαν η μεταφραστική μηχανή του Facebook και το Google Translate και τα αποτελέσματα της μηχανικής μετάφρασης υποβλήθηκαν καταρχάς σε αυτόματη αξιολόγηση μέσω του εργαλείου MTUOC. Ως κείμενο αναφοράς χρησιμοποιήθηκε το αντίστοιχο αγγλόφωνο απόσπασμα κειμένου από το ParaCrawl. Τα αποτελέσματα της εν λόγω αξιολόγησης παρατίθενται στον παρακάτω πίνακα:

Πίνακας 7

Αυτόματη Μετρική	Μηχανή Αυτόματης Μετάφρασης	
	Facebook Translate	Google Translate
BLEU	0.2713	0.2782
NIST	4.753	4.856
WER	0.6038	0.6045
%EdDist	43.4	42.9
TER	0.5447	0.5463

Παρατηρούμε ότι σε όλες τις κατηγορίες οι βαθμολογίες και για τις δύο μηχανές μετάφρασης είναι πολύ κοντά, με το Google Translate να σημειώνει ελαφρώς καλύτερο σκορ από το Facebook Translate όσον αφορά τις μετρικές BLEU και NIST. Σχετικά με την BLEU, και για τις δύο μηχανές η βαθμολογία βρίσκεται κάτω από το όριο του 30% με το Google Translate να υπερτερεί κατά μισή ποσοστιαία μονάδα περίπου συγκεντρώνοντας βαθμολογία 27,82% έναντι 27,13% του Facebook Translate. Η μικρή αυτή διαφορά αποτυπώνεται και στη βαθμολογία NIST, ελαφρώς αυξημένη καθώς ξεπερνά οριακά τη μία ποσοστιαία μονάδα παρουσιάζοντας σκορ 48,56% έναντι 47,53% αντίστοιχα.

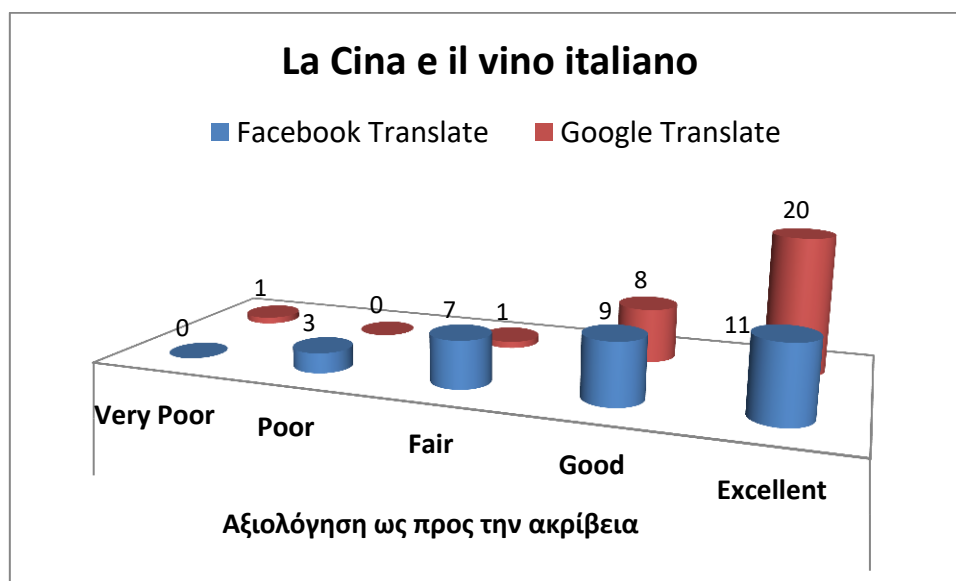
Όσον αφορά τις μετρικές που βασίζονται στην απόσταση διόρθωσης, η μετρική %EdDist είναι ελαφρώς χαμηλότερη για το Google Translate επιβεβαιώνοντας την οριακή υπεροχή που του αποδίδουν και οι μετρικές BLEU και NIST. Αυτό όμως που προκαλεί ιδιαίτερο ενδιαφέρον είναι οι τιμές στον δείκτη σφαλμάτων λέξης (WER) και στον δείκτη σφαλμάτων μετάφρασης (TER). Θα περίμενε κανείς να είναι χαμηλότερες για το Google Translate σε σχέση με τις αντίστοιχες τιμές για τη μετάφραση μέσω Facebook Translate. Στην πράξη όμως, και

όπως φαίνεται και στον Πίνακα 1, οι τιμές των δεικτών είναι χαμηλότερες για το Facebook Translate. Πρόκειται βεβαίως για ελάχιστες διαφορές, 0,07% για τη μετρική WER και 0,16% για την TER αντίστοιχα, οι οποίες παρόλα αυτά διαφοροποιούν ελαφρώς την εικόνα σε σχέση με τις υπόλοιπες μετρικές.

Για να διαπιστώσουμε τους λόγους για τους οποίους παρατηρείται η εν λόγω διαφοροποίηση θα πρέπει να ανατρέξουμε στο κείμενο αναφοράς. Πριν εξετάσουμε όμως τις λεπτομέρειες θα προχωρήσουμε πρώτα στην ανθρώπινη αξιολόγηση των δύο μεταφρασμάτων ως προς την ακρίβεια και την ευχέρεια ώστε να έχουμε μια καλύτερη και πιο ολοκληρωμένη εικόνα και για τα δύο. Τα συγκεντρωτικά αποτελέσματα της ανθρώπινης αξιολόγησης παρατίθενται στους παρακάτω δύο πίνακες και στα αντίστοιχα γραφήματα όπου παρουσιάζεται το σύνολο των τμημάτων κειμένου και το πώς αξιολογήθηκαν:

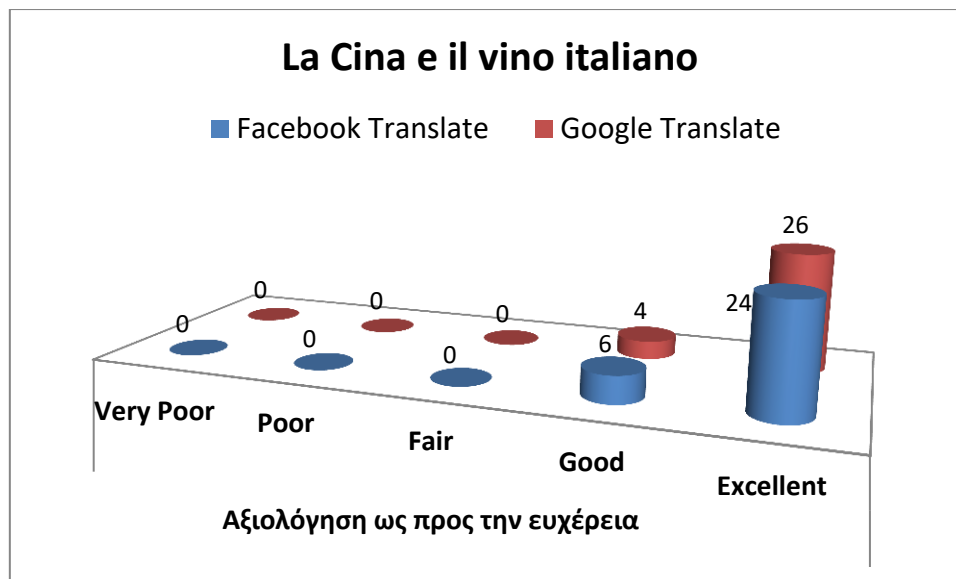
Πίνακας 8

Μηχανή Μετάφρασης	Αξιολόγηση ως προς την ακρίβεια				
	Very Poor	Poor	Fair	Good	Excellent
Facebook Translate	0	3	7	9	11
Google Translate	1	0	1	8	20



Πίνακας 9

Μηχανή Μετάφρασης	Αξιολόγηση ως προς την ευχέρεια				
	Very Poor	Poor	Fair	Good	Excellent
Facebook Translate	0	0	0	6	24
Google Translate	0	0	0	4	26



Ξεκινώντας από τον Πίνακα 3 όπου παρουσιάζεται η αξιολόγηση ως προς την ευχέρεια παρατηρούμε ότι στο συγκεκριμένο κομμάτι και οι δύο μηχανές μετάφρασης εμφανίζουν πολύ καλή ανταπόκριση. Είναι αξιοσημείωτο το γεγονός ότι, σε σύνολο 30 τμημάτων κειμένου, σε καμία από τις δύο δεν εμφανίζονται τμήματα με αρνητική ή μέτρια αξιολόγηση. Τα συστήματα νευρωνικής μετάφρασης άλλωστε, στα οποία ανήκουν τόσο το Google Translate όσο και το Facebook Translate, βελτιώνονται συνεχώς παρουσιάζοντας όλο και πιο άρτια αποτελέσματα ως προς την ευχέρεια και τα αποτελέσματα της παρούσας αξιολόγησης επιβεβαιώνουν για άλλη μια φορά την εν λόγω υπόθεση. Συγκρίνοντας τις δύο μηχανές μεταξύ τους παρατηρούμε ότι και σε αυτόν τον τομέα επιβεβαιώνεται η ελαφρά υπεροχή του Google Translate με 26 τμήματα κειμένου να λαμβάνουν εξαιρετική αξιολόγηση έναντι 24 του Facebook Translate.

Περνώντας στην αξιολόγηση ως προς την ακρίβεια και στον Πίνακα 2, παρατηρούμε ότι το Google Translate υπερτερεί ξεκάθαρα παρουσιάζοντας 28 συνολικά τμήματα με θετική αξιολόγηση εκ των οποίων μάλιστα τα 20 χαρακτηρίζονται ως εξαιρετικά και τα 8 καλά. Το Facebook Translate παρουσιάζει συνολικά 20 συνολικά θετικά αξιολογημένα τμήματα, εκ των οποίων τα 11 χαρακτηρίζονται εξαιρετικά και τα 9 καλά.

Παρόλα αυτά, το μοναδικό τμήμα κειμένου που έλαβε πολύ φτωχή αξιολόγηση ανήκει στη μετάφραση μέσω Google Translate. Πρόκειται για το κομμάτι *Ci viene spontaneo chiederle se ha già fatto visite in cinese* που αποδίδεται ως *We naturally ask you if you have already made visits in Chinese*. Καταρχήν, η μεταφραστική μηχανή του Google συγχέει την προσωπική αντωνυμία *lei* που αναφέρεται σε τρίτο ενικό πρόσωπο γένους θηλυκού με την αντωνυμία *Lei* που είναι ο τύπος ευγενείας για το δεύτερο ενικό πρόσωπο. Αντί λοιπόν στη μετάφραση η ερώτηση να απευθύνεται προς την κοπέλα στην οποία αναφέρεται το κείμενο, απευθύνεται στον αναγνώστη. Το άλλο κομμάτι έχει να κάνει με τη φράση *se ha già fatto visite in cinese* η οποία αναφέρεται στις ξεναγήσεις που παρέχει η κοπέλα και το ερώτημα έχει να κάνει με το αν έχει προσφέρει ξεναγήσεις στην κινεζική γλώσσα. Μια ορθή απόδοση για τη φράση θα ήταν για παράδειγμα η *if she has already given tours in Chinese*. Στην πράξη όπως βλέπουμε αποδίδεται εντελώς εσφαλμένα βάζοντας τον αναγνώστη στον ρόλο του επισκέπτη και ρωτώντας τον αν ο ίδιος έχει πραγματοποιήσει επισκέψεις στα κινέζικα αλ-

λοιώνοντας τελείως το νόημα της πρότασης. Το Facebook Translate αποδίδει τη συγκεκριμένη φράση ως *It's spontaneous to ask her if she's already made visits to Chinese*. Σε αυτήν την περίπτωση, η χρήση της πρόθεσης *to* αλλάζει και το νόημα της λέξης *Chinese* καθώς δεν αντιπροσωπεύει πλέον γλώσσα αλλά ανθρώπους. Επίσης, ο ρόλος αυτού που κάνει την επίσκεψη αποδίδεται και πάλι στο άτομο που δέχεται την ερώτηση, τουλάχιστον όμως σε αυτήν την περίπτωση το άτομο αυτό είναι η κοπέλα και όχι ο αναγνώστης. Για τον λόγο αυτό διαφοροποιείται στον χαρακτηρισμό από τη μετάφραση του Google Translate και χαρακτηρίζεται ως φτωχή και όχι ως πολύ φτωχή.

Από κει και πέρα όμως, το Google δεν παρουσιάζει ιδιαίτερα προβλήματα με εξαίρεση τη φράση *un momento di grande formazione* την οποία και οι δύο μεταφραστικές μηχανές αποδίδουν ως *moment of great training* χάνοντας το νόημα της διδακτικής εμπειρίας (με διδακτικό χαρακτήρα δηλαδή) στην οποία αναφέρεται το κείμενο αφετηρίας. Το Facebook Translate αντίθετα εμφανίζει περισσότερα προβληματικά τμήματα κειμένου όπως για παράδειγμα με την ιδιωματική ιταλική φράση *chiodo fisso* που αποδίδεται κυριολεκτικά ως *fixed nail* χάνοντας το πραγματικό νόημά της που είναι αυτό της έμμονης ιδέας. Δεν λείπουν επίσης τα σφάλματα μετάφρασης στις αντωνυμίες όπως η φράση *le sue due grandi passioni* η οποία αν και αναφέρεται στην πρωταγωνίστρια της συζήτησης αποδίδεται ως *his two great passions*.

Ένα άλλο σφάλμα που εμφανίζεται μόνο στο Facebook Translate έχει να κάνει με την μετάφραση μετοχών οι οποίες στο πρωτότυπο κείμενο εκφράζουν χρονική αφετηρία που η μεταφραστική μηχανή παραλείπει να αποδώσει. Χαρακτηριστικά παραδείγματα οι φράσεις *e terminati gli studi* και *e rientrata in Italia* που αποδίδονται ως *and I have finished my studies* και *and I went back to Italy* αντίστοιχα. Το Google Translate τις αποδίδει ορθότερα ως *and after completing my studies* και *and when I returned to Italy*.

Σε αυτό το σημείο θα σταθούμε λίγο περισσότερο συγκρίνοντας τα δύο μεταφράσματα με το κείμενο αναφοράς. Στο κείμενο αναφοράς οι δύο φράσεις αναφέρονται ως *and once I finished my studies* και *and once back in Italy* αντίστοιχα. Παρατηρούμε ότι ενώ η μετάφραση του Facebook Translate αποτυγχάνει να αποδώσει τη χρονική αφετηρία που εκφράζουν οι εν λόγω φράσεις, σε επίπεδο λέξεων είναι πιο κοντά στο κείμενο αναφοράς. Η φράση *I finished my studies* του Facebook Translate για παράδειγμα ταυτίζεται απόλυτα με την αντίστοιχη του κειμένου αναφοράς ενώ στη μετάφραση του Google Translate αποδίδεται ως *I completed my studies* που σε επίπεδο νοήματος δεν έχει φυσικά καμία διαφορά. Απομονώνοντας τα τμήματα κειμένου των δύο μεταφρασμάτων όπου περιέχεται η εν λόγω φράση και με τη βοήθεια του εργαλείου MTUOC, χρησιμοποιώντας ως αναφορά το αντίστοιχο τμήμα του κειμένου αναφοράς, υπολογίστηκαν οι μετρικές WER και TER και τα αποτελέσματα παρουσιάζονται στον παρακάτω πίνακα:

Πίνακας 10

Αυτόματη Μετρική	Μηχανή Αυτόματης Μετάφρασης	
	Facebook Translate	Google Translate
WER	0.4483	0.6207
TER	0.4483	0.5517

Επαναλαμβάνοντας τη διαδικασία και για τα τμήματα κειμένου όπου περιέχεται το μετάφρασμα της φράσης *e rientrata in Italia* πήραμε τα παρακάτω αποτελέσματα:

Πίνακας 11

Αυτόματη Μετρική	Μηχανή Αυτόματης Μετάφρασης	
	Facebook Translate	Google Translate
WER	0.4894	0.5106
TER	0.4681	0.4894

Και στις δύο περιπτώσεις παρατηρούμε ότι οι δείκτες WER και TER έχουν υψηλότερες τιμές για το Google Translate υποδηλώνοντας και χαμηλότερης ποιότητας μετάφραση, επειδή ακριβώς σε επίπεδο λέξεων η μετάφραση του Facebook Translate είναι πιο κοντά στο κείμενο αναφοράς. Ο δείκτης TER επηρεάζεται αντίστοιχα από τον WER καθώς αποτελεί παραλλαγή του και για αυτό εμφανίζει υψηλότερες τιμές στην περίπτωση του Google. Στην πράξη, και έχοντας υποβάλει τα κείμενα και σε ανθρώπινη αξιολόγηση, γνωρίζουμε ότι συμβαίνει το αντίθετο όσον αφορά την ποιότητα των μεταφράσεων καθώς αυτές του Google Translate είναι ορθότερες. Οι τιμές των μετρικών BLEU και NIST επιβεβαιώνουν την υπόθεσή μας.

Τμήματα κειμένου όμως όπως τα παραπάνω είναι ικανά να προκαλέσουν διαφοροποίηση στις τιμές των μετρικών που βασίζονται στην απόσταση διόρθωσης και να καταγραφούν αποκλίσεις στα συμπεράσματα όπως αυτά που παρατηρήθηκαν στον Πίνακα 1. Κάνοντας μια πρόχειρη σύγκριση μεταξύ των μεταφρασμάτων και του κειμένου αναφοράς μπορεί κάποιος να βρει και άλλα τμήματα που παρουσιάζουν αντίστοιχα φαινόμενα όπως πχ τη φράση *I stayed in touch with UNISG* της μετάφρασης του Facebook Translate που βρίσκεται αυτούσια στο κείμενο αναφοράς ενώ σε αυτήν του Google Translate αναφέρεται ως *I kept in touch with UNISG* ή την *an emerging sector with a great future* που από το Google αναφέρεται ως *an emerging sector that has a great future*.

Το συμπέρασμα που πρέπει τελικά να κρατήσουμε από την μελέτη του παρόντος αποσπάσματος κειμένου είναι ότι αυτόματες μετρικές που βασίζονται στην απόσταση διόρθωσης όπως η WER και η TER μπορεί σε κάποιες περιπτώσεις να μας οδηγήσουν σε εσφαλμένες εκτιμήσεις όσον αφορά την ποιότητα μιας μετάφρασης και δεν θα πρέπει να λαμβάνονται υπόψη χωρίς να παράλληλα να εξετάζονται και άλλες μετρικές. Για να έχουμε μια σαφέστερη και πιο ξεκάθαρη άποψη για τη μετάφραση θα πρέπει να λάβουμε υπόψη και άλλες μετρικές όπως για παράδειγμα η BLEU και η NIST και να εξετάσουμε τη συνολική εικόνα τους προτού προβούμε σε κρίσεις και εξαγωγή συμπερασμάτων.

4.6 Σχολιασμός Eleven Project

Το απόσπασμα κειμένου στο οποίο για τις ανάγκες της παρούσας εργασίας δόθηκε ο τίτλος *Eleven Project* αντλήθηκε από το Parallel Global Voices (Prokopidis et al. 2016), μια συλλογή παράλληλων σωμάτων κειμένου από την ομάδα ιστοσελίδων Global Voices, όπου εθελοντές μεταφράζουν και δημοσιεύουν ιστορίες πολιτών από όλο τον κόσμο. Πρόκειται για μια πρωτοβουλία που έλαβε χώρα στις 11 Νοεμβρίου του 2011 (11-11-11) με σκοπό να φέρει άτομα από όλο τον κόσμο πιο κοντά στα πλαίσια της προώθησης της παγκόσμιας ειρήνης και των αναπτυξιακών στόχων του ΟΗΕ για τη χιλιετία. Απευθυνόταν σε όλους τους ανθρώπους ανεξαιρέτως και τους καλούσε να καταγράψουν κάποιο κομμάτι της ζωής τους εκείνη την ημέρα με τη χρήση φωτογραφιών, βίντεο ή μουσικής αναδεικνύοντας απλές καθημερινές στιγμές που ενώνουν τους πάντες, ανεξαρτήτως φυλής και χρώματος. Αναφέρεται μάλιστα χαρακτηριστικά στην πρόσκληση ότι η εν λόγω καταγραφή μπορεί να γίνει σε οποιαδήποτε γλώσσα. Το γεγονός αυτό σε συνδυασμό με το γεγονός ότι τέτοιου είδους πρωτοβουλίες και καλέσματα προωθούνται κατά κύριο λόγο χρησιμοποιώντας τα μέσα κοινωνικής δικτύωσης καθιστά το συγκεκριμένο απόσπασμα κειμένου ιδιαίτερα ενδιαφέρον προς μελέτη σχετικά με τον τρόπο με τον οποίο το αντιμετωπίζουν οι μηχανές αυτόματης μετάφρασης. Μια πρόσκληση άλλωστε σε ανθρώπους από όλο τον κόσμο και σε όλες τις γλώσσες δεν θα μπορούσε να γραφτεί από ανθρώπους σε κάθε γλώσσα ξεχωριστά.

Το κείμενο μεταφράστηκε από τα ιταλικά στα ελληνικά τόσο με τη χρήση του Facebook Translate όσο και με τη χρήση του Google Translate. Τα μεταφράσματα αξιολογήθηκαν στη συνέχεια αυτόματα ως προς την ποιότητά τους με τη βοήθεια του εργαλείου MTUOC το οποίο έδωσε αποτελέσματα για τις μετρικές BLEU, NIST, WER, %EdDist και TER. Ως κείμενο αναφοράς χρησιμοποιήθηκε το αντίστοιχο ελληνόφωνο κομμάτι του παράλληλου σώματος κειμένων και τα αποτελέσματα της αυτόματης αξιολόγησης εμφανίζονται στον παρακάτω πίνακα:

Πίνακας 12

Αυτόματη Μετρική	Μηχανή Αυτόματης Μετάφρασης	
	Facebook Translate	Google Translate
BLEU	0.1672	0.2246
NIST	3.779	4.238
WER	0.7024	0.6518
%EdDist	63.13	57.86
TER	0.6655	0.6151

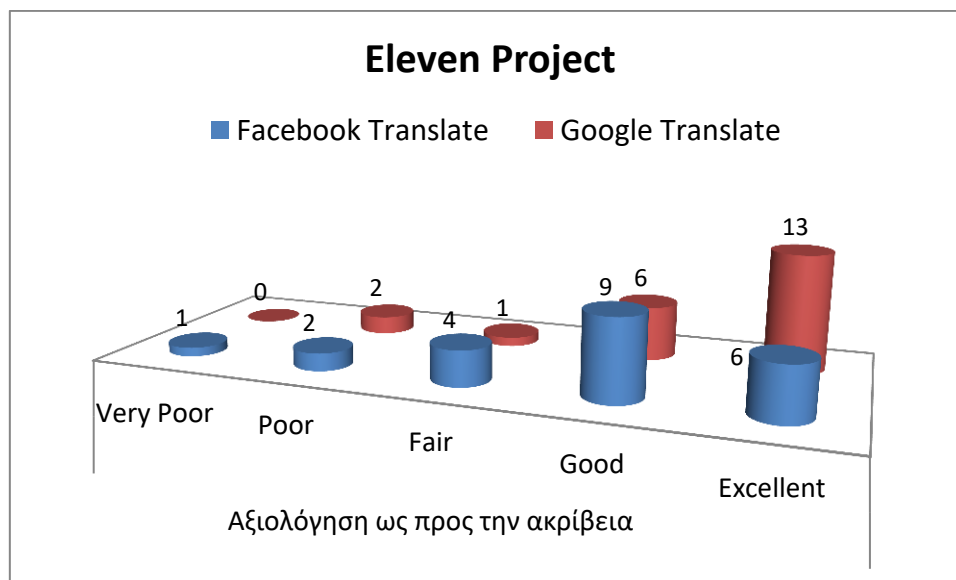
Παρατηρούμε ότι και για τις δύο μηχανές αυτόματης μετάφρασης η βαθμολογία BLEU είναι χαμηλή, αρκετά κάτω από το 30%. Συγκρίνοντας τις δύο μηχανές μεταξύ τους βλέπουμε ότι το Google Translate εμφανίζει αισθητά καλύτερη βαθμολογία BLEU, υψηλότερη κατά περίπου 5,7 ποσοστιαίες μονάδες σε σχέση με αυτήν του Facebook Translate με 22,46% έναντι 16,72% αντίστοιχα. Η διαφορά αυτή αποτυπώνεται και στη βαθμολογία NIST, μειωμένη κατά περίπου 1 ποσοστιαία μονάδα με σκορ 42,38% για τη μηχανή της Google έναντι 37,79% για την αντίστοιχη του Facebook.

Σχετικά με τις μετρικές που βασίζονται στην απόσταση διόρθωσης η μετρική %EdDist είναι εμφανώς χαμηλότερη για το Google Translate επιβεβαιώνοντας την υπεροχή που του αποδίδουν και οι μετρικές BLEU και NIST. Όπως βλέπουμε στον Πίνακα 1, η τιμή της %EdDist για τη μηχανή της Google ανέρχεται σε 57,86 έναντι του 63,13 για το Facebook Translate. Ο δείκτης σφαλμάτων λέξης (WER) επίσης έχει χαμηλότερη τιμή για το Google Translate με 65,18% έναντι 70,24% του Facebook Translate. Η διαφορά αυτή των 5 ποσοστιαίων μονάδων υπέρ της μηχανής της Google αντικατοπτρίζεται και στον δείκτη σφαλμάτων μετάφρασης (TER) με την εν λόγω μετρική να εμφανίζει τιμή 61,51% έναντι του 66,55% που εμφανίζει η αντίστοιχη μεταφραστική μηχανή του Facebook.

Όλα τα παραπάνω στοιχεία καταδεικνύουν σαφή υπεροχή του Google Translate όσον αφορά την ποιότητα της μετάφρασης του συγκεκριμένου κειμένου. Παρουσιάζει αρκετό ενδιαφέρον λοιπόν η σύγκριση των αυτόματων εκτιμήσεων με τις ανθρώπινες εκτιμήσεις που προκύπτουν μετά από την αξιολόγηση των δύο μεταφρασμάτων ως προς την ακρίβεια και ως προς την ευχέρειά τους. Τα συγκεντρωτικά αποτελέσματα για το σύνολο των τμημάτων κειμένου και τον τρόπο με τον οποίο αξιολογήθηκαν παρατίθενται στους παρακάτω δύο πίνακες και στα αντίστοιχα γραφήματα:

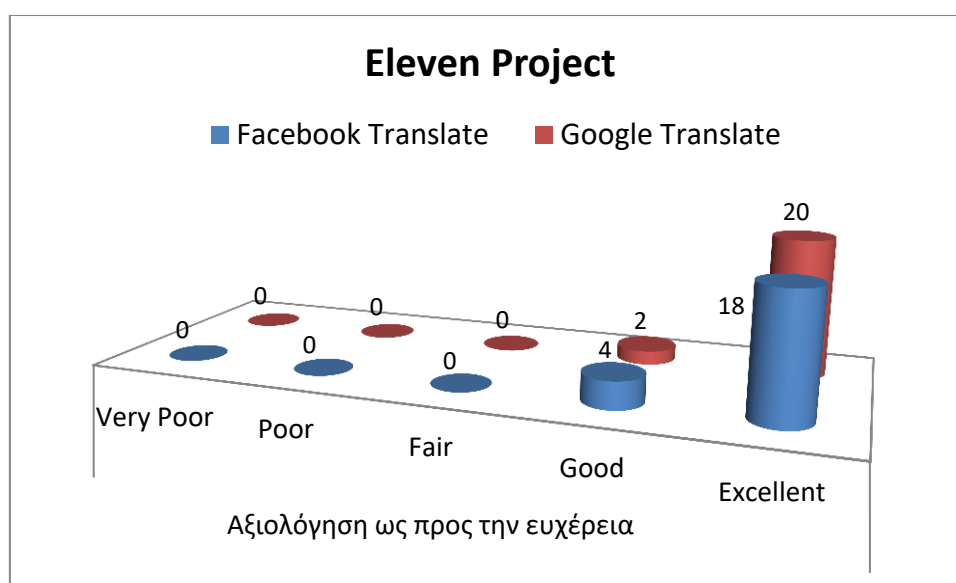
Πίνακας 13

Μηχανή Μετάφρασης	Αξιολόγηση ως προς την ακρίβεια				
	Very Poor	Poor	Fair	Good	Excellent
Facebook Translate	1	2	4	9	6
Google Translate	0	2	1	6	13



Πίνακας 14

Μηχανή Μετάφρασης	Αξιολόγηση ως προς την ευχέρεια				
	Very Poor	Poor	Fair	Good	Excellent
Facebook Translate	0	0	0	4	18
Google Translate	0	0	0	2	20



Ξεκινώντας από τον Πίνακα 3, όπου παρουσιάζονται τα αποτελέσματα της αξιολόγησης ως προς την ευχέρεια, παρατηρούμε ότι, όπως συμβαίνει συνήθως στα αποτελέσματα της νευρωνικής μηχανικής μετάφρασης, δεν εμφανίζονται ιδιαίτερα προβλήματα. Σε κανένα από τα δύο μεταφράσματα δεν καταγράφονται τμήματα κειμένου με αρνητική αξιολόγηση. Σε γενικές γραμμές και τα δύο μεταφράσματα αξιολογούνται ως πολύ ποιοτικά με το Google Translate να υπερέχει ελαφρά εμφανίζοντας 20 τμήματα κειμένου με άριστη και 2 με καλή αξιολόγηση έναντι 18 και 4 του Facebook Translate αντίστοιχα. Τα σφάλματα που παρατηρήθηκαν έχουν να κάνουν κυρίως με παραλείψεις άρθρων όπως πχ στη φράση *ti verrà chiesto di presentare una foto, un video o un file audio* που αποδίδεται από το Facebook Translate ως *θα σας ζητηθεί να υποβάλετε μια φωτογραφία, βίντεο ή αρχείο ήχου* ή κάποιες πτώσεις που δεν αποδίδονται όπως πρέπει, χωρίς όμως σοβαρές επιπτώσεις στην ποιότητα του κειμένου. Χαρακτηριστικό παράδειγμα η φράση *a far parte di tre progetti speciali: un libro fotografico, una compilation musicale e un film documentario* που αποδίδεται από το Facebook Translate ως *που θα γίνει μέρος τριών ειδικών έργων: ένα φωτογραφικό βιβλίο, μια μουσική συλλογή και μια ταινία ντοκιμαντέρ* και από το Google Translate ως *που θα γίνει μέρος τριών ειδικών έργων: ένα άλμπουμ φωτογραφιών, μια συλλογή μουσικής και μια ταινία ντοκιμαντέρ* ενώ το φωτογραφικό άλμπουμ, η μουσική συλλογή και η ταινία ντοκιμαντέρ θα έπρεπε να αναφέρονται σε γενική πτώση.

Όσον αφορά την ακρίβεια, η ανθρώπινη αξιολόγηση κατέγραψε σαφώς περισσότερα σφάλματα και, όπως φαίνεται και στον Πίνακα 2, υπήρξαν και τμήματα κειμένου με μέτρια και αρνητική αξιολόγηση και για τις δύο μεταφραστικές μηχανές. Ξεκινώντας από αυτά με θετική αξιολόγηση παρατηρούμε ότι και πάλι το Google Translate σημειώνει καλύτερες επιδόσεις εμφανίζοντας 13 τμήματα με άριστη και 6 με καλή αξιολόγηση έναντι 6 και 9 του Facebook Translate αντίστοιχα. Τα σφάλματα που διαφοροποίησαν την αξιολόγηση σε καλή και άριστη δεν ήταν ιδιαίτερα σοβαρά. Χαρακτηριστικά παραδείγματα οι φράσεις *Il mondo che vorremmo avere tra 100 anni* που αποδίδεται από το Google Translate ως *Ο κόσμος που θέλουμε να είχαμε σε 100 χρόνια* ή η μετάφραση από το Facebook Translate του *11Eleven Project* ως *11 Έντεκα Πρότζεκτ* σε κάποια τμήματα κειμένου και η αυτολεξεί αμετάφραστη παράθεσή του σε κάποια άλλα τμήματα στερώντας από το κείμενο μια σταθερή εικόνα.

Στα πιο σοβαρά προβλήματα που προέκυψαν από τη μηχανική μετάφραση θα σταθούμε καταρχήν σε ένα κομμάτι κειμένου του οποίου τη μετάφραση για κάποιον ανεξήγητο λόγο το Facebook Translate δεν ολοκλήρωσε ποτέ. Πρόκειται για τη φράση *Per documentarti meglio sulle 8 sfide per lo sviluppo e le misure adottate per conseguirle entro il 2015, consulta la nostra Sezione Speciale* η οποία αποδίδεται ως *Για καλύτερη τεκμηρίωση των 8 αναπτυξιακών προκλήσεων και μέτρων που λαμβάνονται για την επίτευξή τους μέχρι το 2015, δείτε το Ειδ.* Ουσιαστικά δηλαδή η μετάφραση διακόπτεται απότομα. Το εν λόγω σφάλμα παρατηρήθηκε αρχικά στη μετάφραση ολόκληρου του κειμένου και δεν πρόκειται για τμήμα στο τέλος του κειμένου το οποίο θα μπορούσε πχ να παραληφθεί εάν υπήρχαν περιορισμοί στην έκταση. Στη συνέχεια, απομονώθηκε η συγκεκριμένη φράση και το αποτέλεσμα της μηχανικής μετάφρασης ήταν και πάλι το ίδιο όπως αναγράφεται παραπάνω. Σε μια τρίτη προσπάθεια απομονώθηκε μόνο το κομμάτι *consulta la nostra Sezione Speciale* για το οποίο τελικά το Facebook Translate έδωσε ολοκληρωμένη απόδοση ως *δείτε το ειδικό Τμήμα μας*. Τέτοιες παραλείψεις φυσικά δεν είναι σε καμία περίπτωση αποδεκτές και το εν λόγω τμήμα κειμένου δεν θα μπορούσε παρά να χαρακτηριστεί ως πολύ φτωχό ως προς την ακρίβεια.

Και οι δύο μεταφραστικές μηχανές εμφάνισαν από δύο τμήματα με φτωχή αξιολόγηση ως προς την ακρίβεια. Αξίζει να σημειωθεί ότι πρόκειται για διαφορετικά τμήματα, γεγονός που καταδεικνύει και τη διαφορετική προσέγγιση των μηχανών απέναντι στο κείμενο. Πιο συγκεκριμένα, το Facebook Translate αποτυγχάνει να αποδώσει το νόημα της φράσης *inoltre video, fotografie o musica* μεταφράζοντάς το ως *προωθητικά βίντεο, φωτογραφίες ή μουσική* ενώ στην πραγματικότητα πρόκειται για προστακτική και η ορθή απόδοση θα ήταν *προωθήστε μας βίντεο, φωτογραφίες ή μουσική*. Στη φράση *l'importante è che Ti rappresenti* η μεταφραστική μηχανή του Facebook αντί να μεταφράσει το αυτοπαθές ρήμα *rappresentarsi* μεταφράζει το ενεργητικό ρήμα *rappresentare* παραλείποντας να μεταφράσει τη λέξη *Ti* η οποία μάλιστα αναγράφεται με κεφαλαία γράμματα για να δώσει περισσότερη έμφαση στο γεγονός ότι ο κάθε συμμετέχων εκπροσωπείται. Αποδίδεται λοιπόν ως *το σημαντικό είναι ότι εκπροσωπείτε* ενώ η ορθή απόδοση θα ήταν *εκπροσωπείστε* (γιατί όχι και *ΕΚΠΡΟΣΩΠΕΙΣΤΕ*) ή *εκπροσωπείτε τον εαυτό σας* (ή *ΤΟΝ ΕΑΥΤΟ ΣΑΣ*).

Το Google Translate εμφανίζει σοβαρά προβλήματα σε άλλα τμήματα κειμένου όπως πχ το *Video di 11Eleven Project su Youtube* που μεταφράζεται ανεξήγητα ως *11 Έντεχνο βίντεο στο Youtube*. Σε άλλη περίπτωση, στη φράση *accendi il tuo computer, cellulare, macchina foto-*

grafica o registratore e documenta, in qualsiasi lingua, una parte significativa della tua giornata η λέξη *documenta*, ενώ πρόκειται για προστατική του ρήματος *documentare*, μεταφράζεται ως το ουσιαστικό *documento* (έγγραφο) αλλοιώνοντας τελείως το νόημα της φράσης. Έτσι αντί να προτρέπει σε καταγραφή των γεγονότων της ημέρας με τη βοήθεια ενός υπολογιστή, ενός κινητού κτλ η μηχανική μετάφραση εμφανίζει το κείμενο *ενεργοποιήστε τον υπολογιστή, το κινητό τηλέφωνο, τη φωτογραφική μηχανή ή τη συσκευή εγγραφής και το έγγραφο, σε οποιαδήποτε γλώσσα, σημαντικό μέρος της ημέρας σας*.

Υπήρξαν επίσης και τμήματα κειμένου που αξιολογήθηκαν ως μέτρια ως προς την ακρίβειά τους, 4 για το Facebook Translate και 1 για το Google Translate. Σε αυτό το σημείο αξίζει να σταθούμε στο ένα και μοναδικό τμήμα κειμένου στη συγκεκριμένη κατηγορία που είναι κοινό και για τις δύο μεταφραστικές μηχανές. Το αντιμετωπίζουν με εντελώς διαφορετικό τρόπο αποσπώντας τελικά τον ίδιο χαρακτηρισμό ως προς την ποιότητα μεν, για εντελώς διαφορετικούς λόγους δε. Πρόκειται για το κομμάτι *11-11-11 Menti Creative Unite - 3 mesi alla partenza!* την οποία το Facebook Translate αποδίδει ως *11-11-11 Ενωμένα Δημιουργικά Μυαλά - 3 μήνες για αρχή!* με τη φράση *3 μήνες για αρχή* να ξενίζει καθώς δεν αποδίδει ξεκάθαρα το νόημα της αναμονής για την έναρξη αφού θα μπορούσε να αναφέρεται και σε κάποιο προσωρινό διάστημα διάρκειας 3 μηνών. Το Google Translate τη μεταφράζει ως *11-11-11 Creative Minds Unite - 3 μήνες ακόμη!* αποδίδοντας κάπως καλύτερα το διάστημα της τρίμηνης αναμονής αν και πάλι η συγκεκριμένη φράση θα μπορούσε να αναφέρεται και σε τρίμηνη επέκταση ενός χρονικού διαστήματος. Σε κάθε περίπτωση όμως είναι πιο ξεκάθαρη από τη φράση *3 μήνες για αρχή*. Το κομμάτι *Menti Creative Unite* αντίθετα μεταφράζεται στα αγγλικά ενώ η μηχανή προφανώς έχει εντολές να το μεταφράσει στα ελληνικά και δεν πρόκειται για κάποια φράση χαρακτηριστική που χρησιμοποιείται πχ αυτολεξεί στα αγγλικά ακόμα και στην ελληνική γλώσσα ώστε να δικαιολογείται μια τέτοια ενέργεια.

Σε γενικές γραμμές παρατηρούμε ότι για το παρόν κείμενο εμφανίστηκαν αρκετά προβλήματα και στα δύο μεταφράσματα επιβεβαιώνοντας με αυτόν τον τρόπο και τις χαμηλές βαθμολογίες των αυτόματων μετρικών που έδωσε το εργαλείο MTUOC. Αναδείχτηκαν επίσης και οι διαφορές στην προσέγγιση των δύο μεταφραστικών μηχανών σε συγκεκριμένα τμήματα του κειμένου αλλά και ως προς το συνολικό κείμενο γενικότερα. Αξίζει τέλος να σημειωθεί η σαφής υπεροχή του Google Translate παρά το γεγονός ότι πρωτοβουλίες όπως το 11Eleven Project προωθούνται κυρίως μέσα από τα μέσα κοινωνικής δικτύωσης και θα περίμενε κάποιος καλύτερες επιδόσεις από μια μεταφραστική μηχανή που φτιάχτηκε αποκλειστικά για χρήση σε ένα μέσο κοινωνικής δικτύωσης όπως το Facebook.

4.7 Σχολιασμός του Twitter στην υπηρεσία της υγείας

Το απόσπασμα κειμένου στο οποίο για τις ανάγκες της παρούσας εργασίας δόθηκε ο τίτλος *Το Twitter στην υπηρεσία της υγείας* αντλήθηκε από το Parallel Global Voices (Prokorporidis et al. 2016), μια συλλογή παράλληλων σωμάτων κειμένου από την ομάδα ιστοσελίδων Global Voices, όπου εθελοντές μεταφράζουν και δημοσιεύουν ιστορίες πολιτών από όλο τον κόσμο. Αναφέρεται σε μια πρωτοβουλία που λήφθηκε από κάποιους ανθρώπους στην Αίγυπτο σε μια προσπάθεια να αξιοποιήσουν τις δυνατότητες που παρέχει ένα μέσο κοινωνικής δικτύωσης όπως το Twitter ώστε να ευαισθητοποιήσουν τους πολίτες σε θέματα που αφορούν την υγεία. Σε μια εποχή όπως η σημερινή, με μια πανδημία να βρίσκεται σε έξαρση σε παγκόσμιο επίπεδο, πρωτοβουλίες όπως η συγκεκριμένη μπορούν να αποδειχτούν ιδιαίτερα χρήσιμες θέτοντας τα μέσα κοινωνικής δικτύωσης στην υπηρεσία της υγείας. Μπορούν και πρέπει να λειτουργήσουν σε πολλές γλώσσες μιας και δεν πρόκειται για τοπικό αλλά για παγκόσμιο φαινόμενο και οι μηχανές αυτόματης μετάφρασης μπορούν να συνεισφέρουν σημαντικά σε ένα τέτοιου είδους εγχείρημα.

Το συγκεκριμένο κείμενο μεταφράστηκε δύο φορές από τα ιταλικά στα ελληνικά με τη βοήθεια των μεταφραστικών μηχανών Facebook Translate και Google Translate. Χρησιμοποιώντας ως κείμενο αναφοράς το αντίστοιχο ελληνόφωνο τμήμα του παράλληλου σώματος κειμένων, τα δύο μεταφράσματα υποβλήθηκαν στη διαδικασία αυτόματης αξιολόγησης με τη βοήθεια του εργαλείου MTUOC και λήφθηκαν τιμές για τις μετρικές BLEU, NIST, WER, %EdDist και TER. Οι τιμές των εν λόγω μετρικών παρατίθενται στον παρακάτω πίνακα:

Πίνακας 15

Αυτόματη Μετρική	Μηχανή Αυτόματης Μετάφρασης	
	Facebook Translate	Google Translate
BLEU	0.1414	0.1386
NIST	3.88	3.762
WER	0.7113	0.6911
%EdDist	63.33	62.82
TER	0.6562	0.6771

Παρατηρούμε ότι η βαθμολογία BLEU είναι αρκετά χαμηλή και για τα δύο μεταφραστικά εργαλεία, αρκετά πιο χαμηλά από το άτυπο όριο του 30%, με το Facebook Translate να εμφανίζει ελαφρώς καλύτερο σκορ με 14,14% έναντι του 13,86% του Google Translate. Παρόμοια εικόνα παρουσιάζει και η NIST με τη διαφορά για τα δύο εργαλεία να παρουσιάζεται ελαφρώς αυξημένη με 38,8% και 37,62% για Facebook Translate και Google Translate αντίστοιχα.

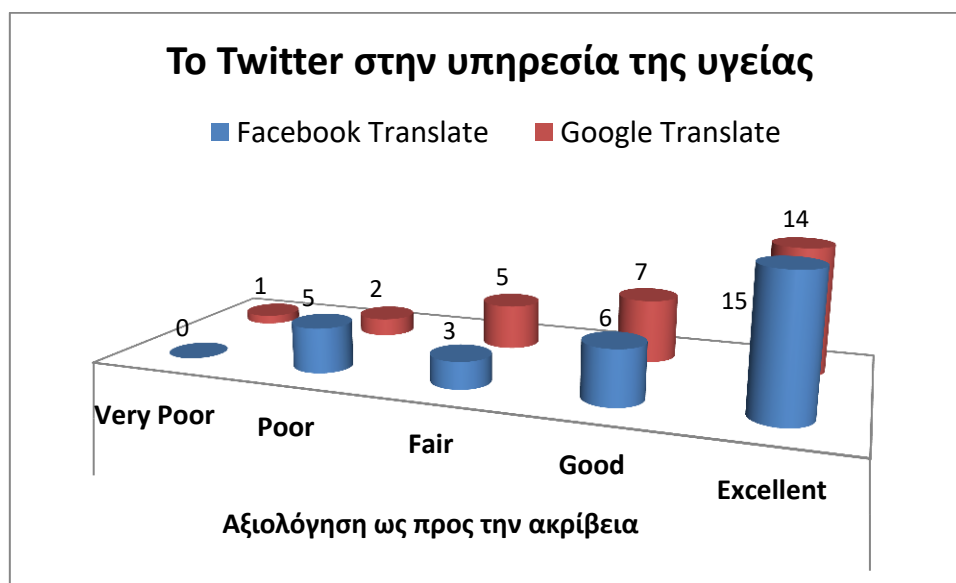
Ιδιαίτερο ενδιαφέρον παρουσιάζουν τα αποτελέσματα που έδωσε το εργαλείο MTUOC για τις μετρικές που βασίζονται στην απόσταση διόρθωσης. Με βάση τις τιμές BLEU και NIST θα περίμενε κανείς μια, έστω και ελαφριά, υπεροχή για την μεταφραστική μηχανή του Facebook η οποία και θα συνεπαγόταν χαμηλότερες τιμές τόσο για τη μετρική %EdDist όσο και για τις μετρικές WER και TER. Παρόλα αυτά, όπως βλέπουμε στον Πίνακα 1, η συγκεκριμένη υπόθεση δικαιώνεται μόνο όσον αφορά τον δείκτη σφαλμάτων μετάφρασης (TER) ο οποίος

έχει τιμή 65,62% για το Facebook Translate έναντι 67,71% για τη μηχανή της Google. Ο δείκτης σφαλμάτων λέξης (WER), αντίθετα, είναι χαμηλότερος για το Google Translate με τιμή 69,11% έναντι 71,13% αντίστοιχα. Το ίδιο συμβαίνει και με τη μετρική %EdDist η οποία εμφανίζεται με χαμηλότερη τιμή για το εργαλείο της Google με 62,82 έναντι 63,33 του εργαλείου της Facebook.

Περνώντας στην ανθρώπινη αξιολόγηση ως προς την ακρίβεια και την ευχέρεια των δύο μεταφρασμάτων θα γίνει προσπάθεια να διαπιστωθεί αν η υπεροχή που μοιάζει να έχει το Facebook Translate με βάση την BLEU, τη NIST και την TER βρίσκει σύμφωνη και την ανθρώπινη κρίση. Τα συγκεντρωτικά αποτελέσματα για το σύνολο των τμημάτων κειμένου και τον τρόπο με τον οποίο αξιολογήθηκαν παρατίθενται στους παρακάτω δύο πίνακες και στα αντίστοιχα γραφήματα:

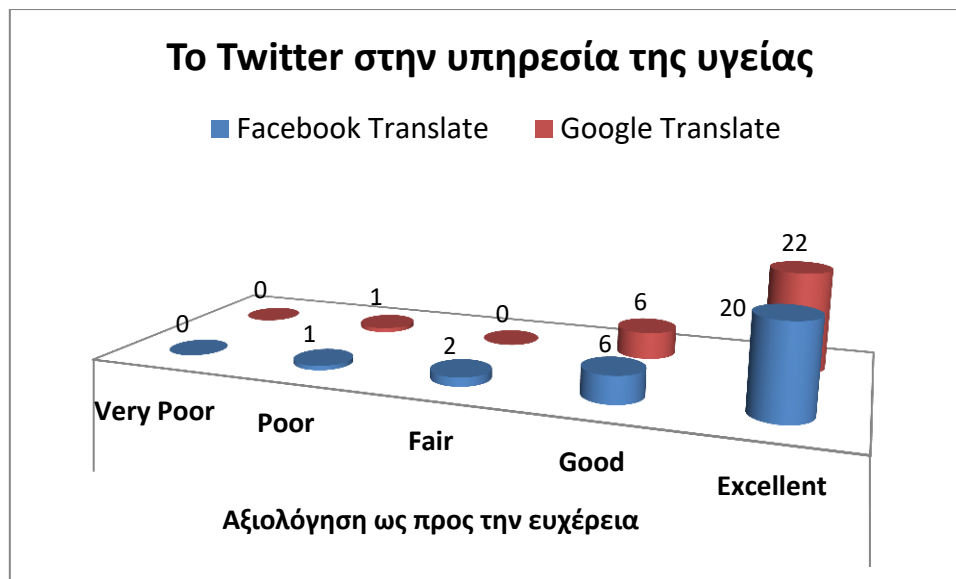
Πίνακας 16

Μηχανή Μετάφρασης	Αξιολόγηση ως προς την ακρίβεια				
	Very Poor	Poor	Fair	Good	Excellent
Facebook Translate	0	5	3	6	15
Google Translate	1	2	5	7	14



Πίνακας 17

Μηχανή Μετάφρασης	Αξιολόγηση ως προς την ευχέρεια				
	Very Poor	Poor	Fair	Good	Excellent
Facebook Translate	0	1	2	6	20
Google Translate	0	1	0	6	22



Ξεκινώντας από τον Πίνακα 3 και την αξιολόγηση ως προς την ευχέρεια, παρατηρούμε ότι καταγράφεται μόλις ένα τμήμα κειμένου με αρνητική αξιολόγηση και για τα δύο μεταφράσματα ενώ η γενική εικόνα είναι πολύ καλή, κάτι που λίγο έως πολύ παρατηρήθηκε και στα υπόλοιπα κείμενα που εξετάστηκαν στην παρούσα εργασία. Όπως έχει ήδη αναφερθεί, τα συστήματα νευρωνικής μετάφρασης έχουν παρουσιάσει εξαιρετική πρόοδο στον συγκεκριμένο τομέα και το παρόν απόσπασμα κειμένου αποτελεί ακόμα μία επιβεβαίωση.

Συγκρίνοντας τις δύο μεταφραστικές μηχανές, παρατηρείται σχετική ισορροπία με μια πολύ ελαφριά υπεροχή του Google Translate καθώς σε σύνολο 29 τμημάτων κειμένου καταγράφονται 22 τμήματα με άριστη αξιολόγηση και 6 με καλή αξιολόγηση έναντι 20 και 6 του Facebook Translate αντίστοιχα. Η ισορροπία επιβεβαιώνεται και από το τμήμα κειμένου με φτωχή αξιολόγηση που είναι το ίδιο και για τα δύο εργαλεία. Πρόκειται για τη φράση *Siamo un gruppo di volontari, tutto è iniziato con me, Kareem Mossad, e mia moglie, Iman Ismail; poi si è unito un amico medico, Dr. Ahmed Ibrahim* η οποία αποδίδεται στα ελληνικά ως *Είμαστε ομάδα εθελοντών, όλα ξεκίνησαν από εμένα, Kareem Mossad, και τη σύζυγό μου, Iman Ismail, μετά ιατρικός φίλος, Dr. Αχμέτ Ιμπραήμ* από το Facebook Translate και ως *Είμαστε μια ομάδα εθελοντών, όλα ξεκίνησαν μαζί μου, η Kareem Mossad, και η σύζυγός μου, Iman Ismail. στη συνέχεια προσχώρησε σε έναν ιατρικό φίλο, τον Δρ. Ahmed Ibrahim*. Παρατηρούμε ότι εμφανίζονται σφάλματα στα γένη και στις πτώσεις, παραλείψεις άρθρων, σφάλματα σε σημεία στίξης και είναι και οι δύο μεταφράσεις αρκετά προβληματικές ως προς την ευχέρεια.

Στο κομμάτι της ακρίβειας καταγράφεται επίσης σχετική ισορροπία η οποία μάλιστα για τις θετικές αξιολογήσεις είναι σχεδόν απόλυτη. Και για τις δύο μεταφραστικές μηχανές καταγράφονται από 21 τμήματα κειμένου με θετική αξιολόγηση με το Facebook Translate να έχει 15 με θετική και 6 με καλή αξιολόγηση και το Google Translate 14 και 7 αντίστοιχα. Το Google Translate εμφανίζει 2 παραπάνω τμήματα με μέτρια αξιολόγηση με 5 εναντίον 3 του Facebook Translate ενώ διαφοροποιημένη εικόνα έχουμε και για τα τμήματα με φτωχή αξιολόγηση όπου το εργαλείο του Facebook παρουσιάζει 5 τμήματα έναντι 2 του εργαλείου της Google. Το Google Translate εμφανίζει όμως και ένα τμήμα με πολύ φτωχή αξιολόγηση και αφορά το κομμάτι που αναφέρεται στην παραπάνω παράγραφο και έλαβε φτωχή αξιολόγηση.

λόγηση και ως προς την ευχέρεια. Παρατηρούμε ότι η απόδοση είναι αρκετά προβληματική και ως προς την ακρίβεια, ειδικά στην απόδοση από το Google Translate. Ελαφρώς καλύτερη εικόνα παρουσιάζει η απόδοση από το Facebook Translate, η οποία και δικαιολογεί εν τέλει τη διαφοροποίηση στον χαρακτηρισμό.

Από κει και πέρα, αξίζει να σταθούμε σε μερικά ακόμα σημεία που έχουν να κάνουν με τα σφάλματα που παρουσιάστηκαν ξεκινώντας από ένα όμοιο σφάλμα που όμως παρουσιάστηκε σε δύο διαφορετικά τμήματα κειμένου. Πρόκειται για την ερώτηση σχετικά με το αν η πρωτοβουλία στην οποία αναφέρεται το κείμενο χρηματοδοτείται από κάποιους και τη σχετική απάντηση. Την ερώτηση λοιπόν που στο ιταλόφωνο κείμενο εμφανίζεται ως *L'account è sponsorizzato da un ente commerciale o da una ONG?* το Google Translate την αποδίδει ορθά, χωρίς κανένα πρόβλημα ως προς την ακρίβεια της, μεταφράζοντάς την ως *Ο λογαριασμός χρηματοδοτείται από εμπορική οντότητα ή ΜΚΟ;* Το εργαλείο του Facebook όμως δίνει τη μετάφραση *Ο λογαριασμός είναι χορηγός εμπορικής οντότητας ή ΜΚΟ;* βάζοντας εσφαλμένα την ίδια την πρωτοβουλία στον ρόλο του χορηγού. Τα πράγματα αντιστρέφονται λίγο παρακάτω όταν οι πρωταγωνιστές του κειμένου δίνουν τη σχετική απάντηση *Non siamo ufficialmente sponsorizzati da nessun ente commerciale o ONG*. Σε αντίθεση με το προηγούμενο τμήμα κειμένου, το Facebook Translate δίνει σε αυτήν την περίπτωση την καθόλα ορθή απόδοση *Δεν είμαστε επισήμως χορηγούμενοι από κανένα εμπορικό φορέα ή ΜΚΟ*. Απ' την άλλη πλευρά, η μεταφραστική μηχανή της Google αποδίδει την εν λόγω φράση ως *Δεν χρηματοδοτούμε επίσημα καμία εμπορική οντότητα ή ΜΚΟ* παρουσιάζοντας το ίδιο ακριβώς πρόβλημα με την αντίστοιχη μηχανή του Facebook νωρίτερα. Κατά μία περίπτωση έννοια λοιπόν, έχουμε ισορροπία και σε αυτό το κομμάτι.

Ένα άλλο σημείο στο οποίο αξίζει να σταθούμε είναι ο χειρισμός ενός εκ παραδρομής σφάλματος στο κείμενο αφετηρίας. Πιο συγκεκριμένα, έχει παραλειφθεί η απόστροφος στη φράση *I' uso* και αναγράφεται ως μία λέξη, δηλαδή ως *Iuso* που δεν έχει κανένα απολύτως νόημα στην ιταλική γλώσσα. Το Facebook Translate δεν την αφήνει τελείως αμετάφραστη, κάτι που δεν θα ήταν παράξενο για μια ανύπαρκτη στην ουσία λέξη, αλλά την αποδίδει με λατινική γραφή ως *Iuse*, μια λέξη που επίσης δεν έχει κανένα νόημα ούτε στην ιταλική ούτε στην ελληνική γλώσσα. Το Google Translate αντίθετα αναγνωρίζει το σφάλμα και την αποδίδει στα ελληνικά σωστά ως *η χρήση*. Ενδιαφέρον παρουσιάζει επίσης και ο χειρισμός της ερώτησης *Da quanto tempo operate online?* κατά τη μετάφραση της οποίας και οι δύο μεταφραστικές μηχανές παρουσίασαν προβλήματα. Το Facebook Translate τη χειρίστηκε καλύτερα αποδίδοντάς την ως *Πόσο καιρό λειτουργείς online;* κάνοντας λάθος ως προς το πρόσωπο που απευθύνεται η ερώτηση το οποίο θα έπρεπε να είναι το δεύτερο πληθυντικό. Το εργαλείο της Google, αντίθετα, την αποδίδει ως *Πόσο καιρό ήσασταν συνδεδεμένοι;* μην κάνοντας λάθος ως προς το πρόσωπο μεν, χρησιμοποιώντας όμως λανθασμένα παρατατικό χρόνο και μεταφράζοντας μάλλον άστοχα στη συγκεκριμένη περίπτωση το *online* ως *συνδεδεμένοι*. Αξίζει τέλος να σημειωθεί η αλλοπρόσαλλη συμπεριφορά του Facebook Translate ως προς κάποια ονόματα τομέα (domain names). Η φράση *healthegypt.org e healthegypt.net* αποδίδεται ως *healthegypt.org και υγιή. Δίχτυ* αναγνωρίζοντας μόνο για το πρώτο κομμάτι από τα δύο ότι πρόκειται για όνομα τομέα και αφήνοντάς το αμετάφραστο, σε αντίθεση με το δεύτερο για το οποίο γίνεται προσπάθεια να μεταφραστεί στα ελληνικά δίνοντας τελικά αυτήν την περιεργή μετάφραση.

Συνοψίζοντας, ίσως φανεί κάπως παράξενο αλλά μπορούμε να πούμε ότι οι αντικρουόμενες εκτιμήσεις των αυτόματων μετρικών όπως παρουσιάζονται στον Πίνακα 1 επιβεβαιώνονται και από τις ανθρώπινες εκτιμήσεις που παρουσιάζονται στους Πίνακες 2 και 3. Τα δύο μεταφράσματα είναι πολύ κοντά ως προς την ποιότητά τους χωρίς να είναι ξεκάθαρο αν κάποιο από τα δύο υπερτερεί τελικά έναντι του άλλου. Προσπαθώντας για παράδειγμα να κρίνουμε από τα τμήματα κειμένου με άριστη αξιολόγηση βλέπουμε ότι ως προς την ευχέρεια υπερέχει ελαφρώς το Google Translate ενώ ως προς την ακρίβεια το Facebook Translate. Απ' την άλλη πλευρά, το εργαλείο της Google έχει λιγότερα τμήματα με αρνητική αξιολόγηση ως προς την ακρίβεια, ένα από αυτά όμως έχει πολύ φτωχή αξιολόγηση, κάτι που δεν συμβαίνει σε κανένα από τα τμήματα κειμένου που μεταφράστηκαν με το εργαλείο του Facebook. Σε κάθε περίπτωση, παρουσιάστηκαν σημαντικά σφάλματα και οι χαμηλές βαθμολογίες BLEU και NIST και οι υψηλές τιμές αντίστοιχα για WER, %EdDist και TER και για τα δύο εργαλεία κρίνονται απόλυτα αντιπροσωπευτικές παρά τις όποιες αποκλίσεις.

5.

Greek Tokenizer: Μια πρόταση για την ελληνική γλώσσα

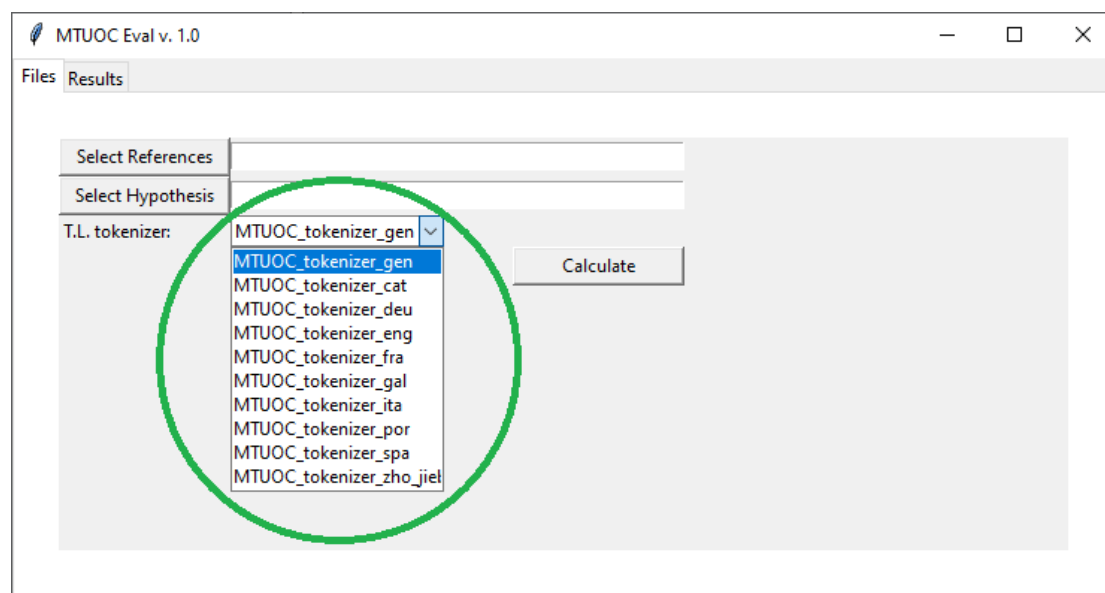
Μια από τις βασικές εργασίες κατά την επεξεργασία φυσικής γλώσσας μέσω υπολογιστή είναι η αναγνώριση των λεκτικών μονάδων ή tokenization όπως λέγεται στα αγγλικά. Πρόκειται για μια διαδικασία μέσω της οποίας το ακατέργαστο κείμενο χωρίζεται σε μικρότερα τμήματα που έχουν κάποιο νόημα. Τα μικρότερα αυτά τμήματα ονομάζονται tokens και συνήθως είναι οι λέξεις του κειμένου. Στις γλώσσες όπου οι λέξεις στις προτάσεις γράφονται με κενά ανάμεσά τους (όπως η ελληνική ή οι γλώσσες που χρησιμοποιούν το λατινικό αλφάβητο), η αναγνώριση των λεκτικών μονάδων είναι κυρίως ζήτημα διαχωρισμού των σημείων στίξης (Koehn 2010:34). Σε γλώσσες περισσότερο δύσκολες ή πολύπλοκες, όπως η κινεζική, εφαρμόζονται ειδικοί αλγόριθμοι ώστε να γίνει με επιτυχία ο εν λόγω διαχωρισμός. Η αντίστροφη διαδικασία, κατά την οποία πραγματοποιείται η ανασύνθεση μιας πρότασης ενώνοντας τις λεκτικές μονάδες από τις οποίες αποτελείται, ονομάζεται στα αγγλικά detokenization ενώ στα ελληνικά δεν της έχει αποδοθεί ακόμα επίσημα κάποιος όρος.

Πέρα από τον βασικό διαχωρισμό, η αναγνώριση των λεκτικών μονάδων έχει να διαχειριστεί και άλλα θέματα όπως για παράδειγμα λέξεις που χωρίζονται με αποστροφούς ή παύλες. Πολλές φορές δεν έχει νόημα ο διαχωρισμός των τμημάτων που βρίσκονται δεξιά και αριστερά των σημείων στίξης όπως στη γαλλική λέξη *p'tit* ή στην αγγλική *co-operate*, ούτε ο διαχωρισμός κάποιων σημείων στίξης ως ξεχωριστά tokens. Όταν πχ έχουμε τη φράση *Joe's* στην αγγλική γλώσσα, ο διαχωρισμός σε 3 tokens (*Joe*, *'* και *s*) δεν είναι τόσο ξεκάθαρος όσο ο διαχωρισμός σε 2 (*Joe* και *'s*) όπου χάρη στο token *'s* διακρίνεται ξεκάθαρα η γενική κτητική.

Η αναγνώριση λεκτικών μονάδων στα πλαίσια της μηχανικής μετάφρασης έχει ως στόχο τον διαχωρισμό του κειμένου σε όσο το δυνατόν λιγότερα tokens. Δεν είναι επιθυμητή η ύπαρξη διαφορετικών tokens για την ίδια λέξη επειδή κάποιες φορές μπορεί να γράφεται με κεφαλαία γράμματα ή ακολουθείται από κάποιο σημείο στίξης (Koehn 2010:35). Δεν θέλουμε για παράδειγμα να έχουμε διαφορετικά tokens για τη λέξη *house* χωρίς να ακολουθείται από κάποιο σημείο στίξης, τη λέξη *house*, ακολουθούμενη από κόμμα ή τη λέξη *"house"* σε εισαγωγικά. Ούτε για τη λέξη *House* με το πρώτο γράμμα κεφαλαίο ή *HOUSE* με όλα τα γράμματα κεφαλαία αντίστοιχα. Πρόκειται για την ίδια λέξη σε όλες τις περιπτώσεις. Για τον λόγο αυτό, κατά τη διαδικασία της αναγνώρισης λεκτικών μονάδων όλες οι λέξεις συνήθως μετατρέπονται σε μορφή όπου όλα τα γράμματά τους είναι μικρά (lowercasing) με εξαίρεση περιπτώσεις όπου έχει σημασία η χρήση ενός κεφαλαίου γράμματος οπότε αυτό διατηρείται (truecasing). Ο κύριος *Μυλωνάς* για παράδειγμα δεν έχει απαραίτητα σχέση με το επάγγελμα του *μυλωνά* και οι δύο λέξεις πρέπει να αποτελέσουν ξεχωριστά tokens.

Το εργαλείο MTUOC-Eval διαθέτει έναν γενικό (general) tokenizer που βασίζεται στις βιβλιοθήκες της πλατφόρμας NLTK (Natural Language Toolkit, μια πλατφόρμα επεξεργασίας της φυσικής γλώσσας με τη βοήθεια της γλώσσας προγραμματισμού Python) και μπορεί να εφαρμοστεί ανεξάρτητα από την ανθρώπινη γλώσσα στην οποία είναι γραμμένο το κείμενο αναφοράς και το μηχανικά μεταφρασμένο κείμενο. Κατά τη διαδικασία αναγνώρισης λεκτικών μονάδων διαχωρίζει κάθε λέξη αλλά και κάθε σημείο στίξης (με εξαίρεση την τελεία)

σε ξεχωριστά tokens. Για να διαχειριστεί θέματα όπως αυτά που αναφέρονται στις δύο παραπάνω παραγράφους, ο δημιουργός του εργαλείου δίνει τη δυνατότητα χρήσης ξεχωριστού tokenizer για κάποιες από τις πιο δημοφιλείς ανθρώπινες γλώσσες. Όλοι οι tokenizers βασίζονται στον γενικό tokenizer με κάποιες προσθήκες στον κώδικά τους, ανάλογα με τις ιδιαιτερότητες της εκάστοτε γλώσσας. Ο χρήστης μπορεί να επιλέξει αυτόν που επιθυμεί από το αναδυόμενο μενού όπως φαίνεται και στην παρακάτω εικόνα:

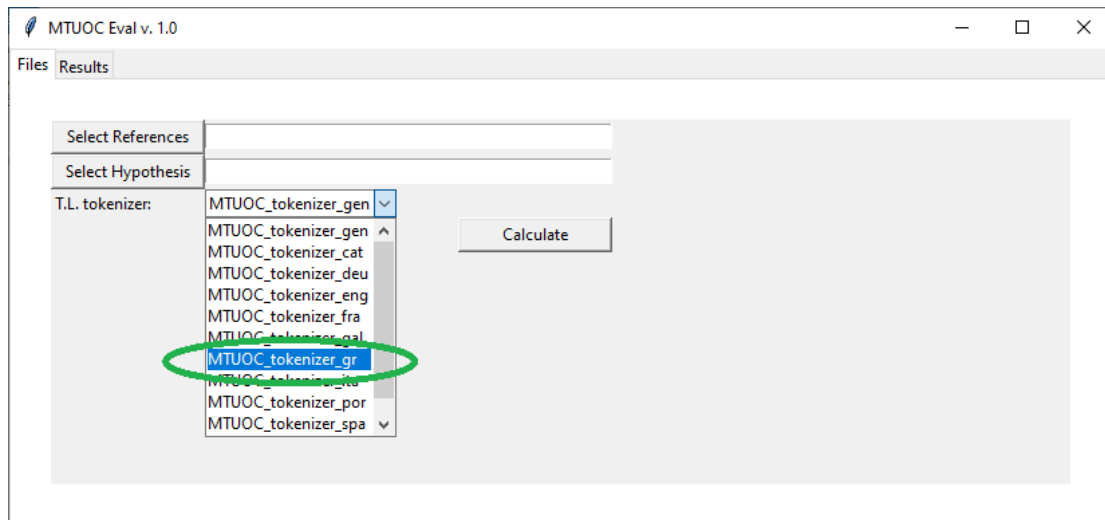


Πέρα από τον γενικό tokenizer, ο οποίος βρίσκεται στην κορυφή του μενού, παρέχονται επιλογές για tokenizer στην καταλανική, τη γερμανική, την αγγλική, τη γαλλική, τη γαλικιανή, την ιταλική, την πορτογαλική, την ισπανική και την κινεζική γλώσσα. Στην πράξη, μέχρι τον Ιανουάριο του 2021, είχε γραφτεί επιπλέον κώδικας μόνο για τα καταλανικά (μητρική γλώσσα του δημιουργού της εφαρμογής), τα αγγλικά, τα γαλλικά και τα ιταλικά ενώ οι υπόλοιποι tokenizers έδιναν τα ίδια αποτελέσματα με τον γενικό tokenizer.

Στα κείμενα που μελετήθηκαν στην παρούσα εργασία χρησιμοποιήθηκε ο αγγλικός tokenizer για τα κείμενα *FooTweets*, *Telework* και *La Cina e il vino italiano* μιας και γλώσσα-στόχος σε αυτά ήταν τα αγγλικά. Στα δύο τελευταία όπου γλώσσα-στόχος ήταν τα ελληνικά χρησιμοποιήθηκε ο γενικός tokenizer, γεννήθηκε όμως η ιδέα μιας προσπάθειας για ανάπτυξη ενός tokenizer για την ελληνική γλώσσα. Βασική φιλοσοφία ήταν η διαχείριση περιπτώσεων όπου οι λέξεις χωρίζονται με απόστροφο και κρίθηκε σκόπιμο η απόστροφος να μην αποτελέσει ξεχωριστό token. Για παράδειγμα, η φράση *κατ' επέκταση* με τον γενικό tokenizer θα χωριζόταν σε 3 tokens (*κατ*, *'* και *επέκταση*) ενώ με τον ελληνικό σε 2 (*κατ'* και *επέκταση*). Ο κώδικας του ελληνικού tokenizer παρατίθεται στο παράρτημα της παρούσας εργασίας με τα σημεία που αναφέρονται στις ιδιαιτερότητες της ελληνικής γλώσσας να έχουν επισημανθεί με έντονη γραφή.

Από την ηλεκτρονική διεύθυνση

<https://drive.google.com/file/d/1DgEBuBivSLm9eSwZbeF1G-EOmcoJ15EN/view> μπορεί κάποιος να κατεβάσει την προτεινόμενη έκδοση του MTUOC-Eval με την προσθήκη του tokenizer για την ελληνική γλώσσα και να το επιλέξει από το πτυσσόμενο μενού όπως φαίνεται στην παρακάτω εικόνα:



Τα δύο κείμενα στα οποία γλώσσα-στόχος ήταν τα ελληνικά αξιολογήθηκαν πειραματικά εκ νέου με τη χρήση του ελληνικού tokenizer και τα αποτελέσματα της αξιολόγησης παρατίθενται στους παρακάτω δύο παρακάτω πίνακες. Οι διαφορές επισημαίνονται με έντονη γραφή και διαφορετικό χρώμα φόντου.

Eleven Project

Αυτόματη Μετρική	Μηχανή Αυτόματης Μετάφρασης			
	Facebook Translate		Google Translate	
	General Tokenizer	Greek Tokenizer	General Tokenizer	Greek Tokenizer
BLEU	0.1672	0.1672	0.2246	0.2246
NIST	3.779	3.782	4.238	4.238
WER	0.7024	0.6996	0.6518	0.6518
%EdDist	63.13	63.13	57.86	57.86
TER	0.6655	0.6637	0.6151	0.6151

Το Twitter στην υπηρεσία της υγείας

Αυτόματη Μετρική	Μηχανή Αυτόματης Μετάφρασης			
	Facebook Translate		Google Translate	
	General Tokenizer	Greek Tokenizer	General Tokenizer	Greek Tokenizer
BLEU	0.1414	0.1443	0.1386	0.1389
NIST	3.88	3.885	3.762	3.765
WER	0.7113	0.7109	0.6911	0.6908
%EdDist	63.33	63.33	62.82	62.82
TER	0.6562	0.6556	0.6771	0.6766

Αρχικά παρατηρούμε ότι στο *Eleven Project* δεν υπάρχει καμία διαφοροποίηση στις τιμές των μετρικών για τη μετάφραση από το Google Translate, ανεξάρτητα από το ποιος tokenizer χρησιμοποιείται. Αυτό συμβαίνει προφανώς διότι δεν υπάρχουν λέξεις που χωρίζονται με αποστροφούς, ούτε στο κείμενο αναφοράς ούτε και στο μεταφρασμένο κείμενο.

Όσον αφορά τη μετάφραση από Facebook Translate, όταν χρησιμοποιείται ο ελληνικός tokenizer, παρατηρούμε ελαφρώς καλύτερη εικόνα για τη μετάφραση με τις τιμές των BLEU και NIST να είναι ελαφρώς ανεβασμένες και αυτές των WER και TER ελαφρώς πιο χαμηλές σε σχέση με τις αντίστοιχες τιμές που προέκυψαν από τη χρήση του γενικού tokenizer. Η %EdDist δεν μεταβάλλεται σε καμία περίπτωση.

Στο *Twitter στην υπηρεσία της υγείας* η %EdDist παραμένει και πάλι αμετάβλητη σε όλες τις περιπτώσεις. Για όλες τις υπόλοιπες μετρικές παρουσιάζεται ελαφρώς καλύτερη εικόνα και για τα δύο μηχανικά μεταφρασμένα κείμενα όταν χρησιμοποιείται ο ελληνικός tokenizer. Όπως και στο *Eleven Project*, οι τιμές για τη BLEU και τη NIST εμφανίζονται αυξημένες και οι TER και WER αντίστοιχα μειωμένες.

Είναι προφανές ότι η γενική εικόνα των μετρικών αυτόματης αξιολόγησης δεν μεταβάλλεται σε τέτοιο βαθμό ώστε να επηρεάζονται όσα γράφτηκαν κατά τον σχολιασμό των κειμένων στα προηγούμενα κεφάλαια. Η έστω και ελαφριά διαφοροποίηση όμως στις τιμές, καταδεικνύει ότι η χρήση ενός πιο εξειδικευμένου tokenizer θα μπορούσε να προσθέσει λεπτομέρεια στο αποτέλεσμα η οποία, ανάλογα και με τη φύση του εξεταζόμενου κειμένου, θα μπορούσε να είναι αρκετά χρήσιμη.

6.

Ποιοτική Ανάλυση Αποτελεσμάτων

Ολοκληρώνοντας την παρούσα μελέτη αξίζει να σταθούμε στις δυσκολίες και στις ιδιαιτερότητες που παρουσιάζει η γλώσσα που χρησιμοποιείται στα μέσα κοινωνικής δικτύωσης, και κατ' επέκταση η μηχανική μετάφρασή της, και πώς αυτές εντοπίστηκαν στα κείμενα που εξετάστηκαν. Πρέπει να σημειωθεί ότι επειδή ακριβώς πρόκειται για μέσα κοινωνικής δικτύωσης τα οποία διαβάζει και χρησιμοποιεί πολύς κόσμος, τυχόν λανθασμένες μεταφράσεις περιεχομένου ή δημοσιεύσεων μπορεί ενίοτε να έχουν πολύ δυσάρεστες συνέπειες. Για τον λόγο αυτό είναι απαραίτητη η διαρκής αξιολόγηση των μεταφραστικών μηχανών και η βελτίωση των παραγόμενων μεταφράσεων.

Χαρακτηριστική είναι η περίπτωση Παλαιστίνιου εργάτη στη Δυτική Όχθη ο οποίος το 2017 δημοσίευσε στο προφίλ του στο Facebook μια φωτογραφία του μπροστά στη μπουλντόζα με την οποία εργαζόταν και το μήνυμα *yusbiuhum* (yusbiuhum) που στην αραβική γλώσσα σημαίνει *καλημέρα*. Η μεταφραστική μηχανή του Facebook απέδωσε το μήνυμά του τελείως λανθασμένα ως *επιτεθείτε τους* στην εβραϊκή γλώσσα ή *πληγώστε τους* στην αγγλική οδηγώντας στη σύλληψή του από τις αρχές καθώς υπέθεσαν ότι πρόκειται για τρομοκράτη ο οποίος σχεδίαζε να επιτεθεί σε ανυποψίαστους πολίτες με τη μπουλντόζα. Τελικά η παρεξήγηση λύθηκε και το Facebook αναγκάστηκε να ζητήσει δημόσια συγγνώμη για το σφάλμα της μεταφραστικής του μηχανής (Hern 2017b).

Μια άλλη χαρακτηριστική και πιο πρόσφατη περίπτωση είναι αυτή της λανθασμένης μετάφρασης της ανακοίνωσης για την τελετή γενεθλίων του βασιλιά της Ταϊλάνδης όπου η αγγλική φράση *[Live] Candle-lighting ceremony to celebrate the birthday of HM the King on July 28, 2020 at 6.45 PM* μεταφράστηκε αυτόματα στην ταϊλανδέζικη γλώσσα με τρόπο προσβλητικό προς τη χώρα και τον μονάρχη της αναγκάζοντας και πάλι το Facebook να ζητήσει δημόσια συγγνώμη και να αναστείλει προσωρινά τις υπηρεσίες μηχανικής μετάφρασης στην ταϊλανδέζικη έκδοση τόσο του Facebook όσο και του Instagram (Tortermvasana 2020). Βεβαίως, τέτοιου είδους παρανοήσεις δεν έχουν συμβεί μόνο στο Facebook αλλά και σε άλλα μέσα κοινωνικής δικτύωσης όπως για παράδειγμα η κινέζικη εφαρμογή ανταλλαγής μηνυμάτων WeChat όπου η ουδέτερη φράση *hei laowai*, που στα αγγλικά κυριολεκτικά σημαίνει *black foreigner*, αποδόθηκε από τη μεταφραστική μηχανή της εφαρμογής με την προσβλητική και μη αποδεκτή λέξη *nigger* προκαλώντας σφοδρές αντιδράσεις (Hern 2017a).

6.1 Ιδιαιτερότητες της γλώσσας στα μέσα κοινωνικής δικτύωσης

Η γλώσσα που χρησιμοποιείται στα μέσα κοινωνικής δικτύωσης παρουσιάζει κάποιες ιδιαιτερότητες οι οποίες δημιουργούν δυσκολίες κατά τη διαδικασία της μηχανικής μετάφρασής της. Καταρχήν, περιέχει κάποια στοιχεία που απουσιάζουν από την επίσημη γραπτή γλώσσα. Πιο χαρακτηριστικά είναι τα σύμβολα που χρησιμοποιούνται για ειδικές λειτουργίες όπως για παράδειγμα τα hashtags που συμβολίζονται με τη δίεση (#) ακολουθούμενα από το εκάστοτε θέμα στο οποίο αναφέρονται, τα cashtags (\$) που αναφέρονται σε επιχειρήσεις και φυσικά το σύμβολο AT (@) που αναφέρεται σε άλλους χρήστες του μέσου κοινωνικής

δικτύωσης. Ο τρόπος που θα χειριστεί η μεταφραστική μηχανή τα κείμενα που ακολουθούν τα εν λόγω σύμβολα είναι πολύ σημαντικός καθώς πιθανές μεταφράσεις μπορούν να οδηγήσουν σε σφάλματα ή ακόμα και να απενεργοποιήσουν τελείως τις ειδικές λειτουργίες τις οποίες εξυπηρετούν. Στα κείμενα που εξετάστηκαν, δεν παρατηρήθηκαν προβλήματα στο συγκεκριμένο κομμάτι, καθώς τα συστήματα αυτόματης μετάφρασης έχουν εκπαιδευτεί κατάλληλα ώστε να αναγνωρίζουν τα ειδικά σύμβολα και να μην μεταφράζουν κομμάτια του κειμένου που δεν πρέπει να μεταφραστούν. Προβλήματα θα μπορούσαν να δημιουργηθούν σε περιπτώσεις όπου ένα ειδικό σύμβολο έχει και κάποια άλλη λειτουργία όπως για παράδειγμα η δίσση (#) στη μουσική ορολογία. Η νότα *Nτο#* ας πούμε, ή *Do#* στα ιταλικά, στην αγγλική μουσική γλώσσα αναγράφεται ως *C#*. Αν εμφανιζόταν σε μια δημοσίευση σε μέσο κοινωνικής δικτύωσης ενδεχομένως να μην αποδίδονταν όπως έπρεπε από τη μεταφραστική μηχανή. Πρόκειται όμως για αρκετά εξειδικευμένο σενάριο.

Ένα άλλο χαρακτηριστικό της γλώσσας των μέσων κοινωνικής δικτύωσης είναι τα emoticons. Πρόκειται για ακολουθίες συμβόλων οι οποίες χρησιμοποιούνται για να εκφράσουν κατά κύριο λόγο συναισθήματα. Μετεξέλιξή τους αποτελούν τα emoji τα οποία είναι εικονίδια που μπορούν να συμπεριλαμβάνονται στα κείμενα και να εκφράζουν συναισθήματα αλλά και αντικείμενα. Ούτε σε αυτό το κομμάτι παρουσιάστηκαν προβλήματα καθώς πρόκειται κατά κύριο λόγο για σύμβολα χωρίς κείμενο τα οποία απλώς μεταφέρονται αυτούσια και στο μηχανικό μετάφρασμα και αντιμετωπίζονται με τον ίδιο τρόπο αφού οι μεταφραστικές μηχανές έχουν εκπαιδευτεί ώστε να αναγνωρίζουν τις διάφορες ακολουθίες. Σε κάποιο ακραίο σενάριο θα μπορούσαν βεβαίως να υπάρξουν θέματα όπως για παράδειγμα στη φράση *Είμαι καλά ;-)* όπου τα σύμβολα *;-)* εκφράζουν το κλείσιμο του ματιού (wink) και η όλη φράση έχει θετική χροιά. Αν εκ παραδρομής μπει κενό μετά το ελληνικό ερωτηματικό η μεταφραστική μηχανή μπορεί να δει την εν λόγω φράση ως ερώτηση και να την αποδώσει λανθασμένα.

Το πιο σημαντικό όμως χαρακτηριστικό της γλώσσας στα κοινωνικά δίκτυα είναι ότι δεν πρόκειται για επίσημη γλώσσα αφού ο καθένας μπορεί να γράψει και να εκφραστεί χωρίς να είναι «υποχρεωμένος» να ακολουθήσει ακόμα και τους βασικούς κανόνες γραμματικής, συντακτικού κτλ. Αυτό έχει σαν αποτέλεσμα να παρατηρούμε πολλές φορές φαινόμενα όπως η λανθασμένη ή και καθόλου χρήση των κεφαλαίων γραμμάτων αλλά και προτάσεις με λανθασμένα ή ακόμα και χωρίς σημεία στίξης.

Στα κείμενα που εξετάστηκαν, το εν λόγω φαινόμενο παρατηρήθηκε πολύ έντονα στο *FooTweets* μιας και ήταν και το μοναδικό κείμενο που αντλήθηκε απευθείας από μέσο κοινωνικής δικτύωσης και πιο συγκεκριμένα από το Twitter. Όσον αφορά τη χρήση των κεφαλαίων γραμμάτων η συμπεριφορά και των δύο μεταφραστικών μηχανών ήταν μάλλον «απροσδιόριστη» καθώς κάποιες προτάσεις-τιτιβίσματα που ξεκινούσαν με μικρό γράμμα στο κείμενο αφητηρίας αποδόθηκαν στο μετάφρασμα ξεκινώντας με κεφαλαίο γράμμα ενώ κάποιες άλλες όχι, χωρίς να προκύπτει κάποιο προφανές στοιχείο που να δικαιολογεί αυτήν τη συμπεριφορά. Αντιθέτως, δεν υπήρξε πρόβλημα στις προτάσεις που ξεκινούσαν με κεφαλαίο γράμμα και στο κείμενο αφητηρίας. Σε κάθε περίπτωση όμως, δεν παρατηρήθηκαν προβλήματα στην απόδοση του νοήματος εξαιτίας των κεφαλαίων.

Αντίθετα, καταγράφηκε πρόβλημα σχετικό με τα σημεία στίξης στην πρόταση *warte was ? nicht , wenn er in den Ruhestand gehen muss # Gerrard # Worldcup* στο *FooTweets* όπου κατά τη μετάφραση μέσω Google Translate το ερωτηματικό παραλήφθηκε με αποτέλεσμα να προκύψει η συγκεχυμένη φράση *wait what not if he has to retire # Gerrard # Worldcup*. Επίσης, ένα άλλο πρόβλημα σχετικό με τα σημεία στίξης καταγράφηκε στο κείμενο του *Twitter* στην υπηρεσία της υγείας όπου η μεταφραστική μηχανή του Facebook απόδωσε τη φράση *healthegypt.org e healthyegypt.net* ως *healthegypt.org* και υγιή. Δίχτυ αναγνωρίζοντας μόνο για το πρώτο κομμάτι από τα δύο ότι πρόκειται για όνομα τομέα (domain name) και αφήνοντάς το αμετάφραστο. Στο δεύτερο κομμάτι, αντίθετα, εξέλαβε την τελεία ως το τέλος μιας πρότασης και προσπάθησε να μεταφράσει το κείμενο δεξιά και αριστερά της σαν να ήταν μέρος της.

Ακόμα πιο σημαντική παράμετρος που αφορά τη μη χρήση επίσημης γλώσσας στα κοινωνικά δίκτυα είναι η συχνή χρήση αργκό, ιδιωματικών φράσεων αλλά και αρκτικόλεξων όπως για παράδειγμα το *LOL (lough out loud)*, το *YOLO (you only live once)* και άλλα. Κύριο χαρακτηριστικό τους αποτελεί το γεγονός ότι συνήθως αφορούν την αγγλική γλώσσα και επειδή ως αρκτικόλεξα δεν έχουν κάποια κυριολεκτική σημασία συνήθως παραμένουν αμετάφραστα κι έτσι δεν δημιουργούν ιδιαίτερα προβλήματα. Ούτε στα κείμενα που εξετάστηκαν καταγράφηκε κάποιο πρόβλημα σχετικά με τη χρήση αρκτικόλεξων.

Σχετικά με τις ιδιωματικές φράσεις, το πιο συνηθισμένο σφάλμα που μπορεί να παρατηρηθεί είναι η κατά λέξη μετάφραση η οποία προφανώς και αποτελεί αστοχία στην πλειοψηφία των περιπτώσεων. Χαρακτηριστική είναι η ιταλική φράση *chiodo fisso* στο κείμενο *La Cina e il vino italino* που αποδίδεται κυριολεκτικά από το Facebook Translate ως *fixed nail* ενώ στην πραγματικότητα έχει το νόημα της έμμονης ιδέας. Ιδιωματικές φράσεις βεβαίως χρησιμοποιούνται και σε άλλα είδη κειμένου, όχι μόνο στα μέσα κοινωνικής δικτύωσης, ενδεχομένως με όχι τόσο μεγάλη συχνότητα.

Οι χρήστες των μέσων κοινωνικής δικτύωσης συνηθίζουν επίσης να χρησιμοποιούν πολλές συντομογραφίες στις δημοσιεύσεις τους, πολλές φορές επειδή υπάρχει περιορισμός στην έκταση των δημοσιεύσεων όπως για παράδειγμα στο *Twitter* όπου μια δημοσίευση δεν μπορεί να ξεπερνά τους 140 χαρακτήρες. Έτσι, πολλές φορές παρατηρούμε την απουσία χρήσης αποστροφών (γράφοντας πχ *havent* αντί για *haven't*, *Im* αντί για *I'm* κτλ) ή συντομογραφίες λέξεων όπως *u* αντί για *you*, *4* αντί για *for* (πρόθεση), *c* αντί για *see* και άλλες.

Σε γενικές γραμμές δεν παρατηρήθηκαν προβλήματα αυτού του είδους στα κείμενα που εξετάστηκαν. Αυτό οφείλεται κυρίως στο γεγονός ότι στο *FooTweets*, που αντλήθηκε απευθείας από το *Twitter* και είχε τις περισσότερες περιπτώσεις ιδιωματικής γραφής, είχε ήδη γίνει σχετική προεργασία κατά τη δημιουργία του παράλληλου σώματος κειμένων. Συντομογραφίες όπως οι παραπάνω αλλά και ιδιωματικές φράσεις τύπου *lil* αντί για *little* μεταφράστηκαν σε τυπική γλώσσα με προσέγγιση informal to formal (Sluyter-Gäthje et al. 2018: 2667) κατά την μετάφραση των πρωτότυπων αγγλόφωνων tweets στα γερμανικά. Στην παρούσα εργασία εξετάστηκε η μηχανική μετάφραση στα αγγλικά των ήδη μεταφρασμένων γερμανόφωνων tweets με τη χρήση των πρωτότυπων αγγλόφωνων tweets ως κείμενο αναφοράς. Το πρόβλημα λοιπόν δεν αποτυπώθηκε στο ίδιο το κείμενο της μετάφρασης, οι τι-

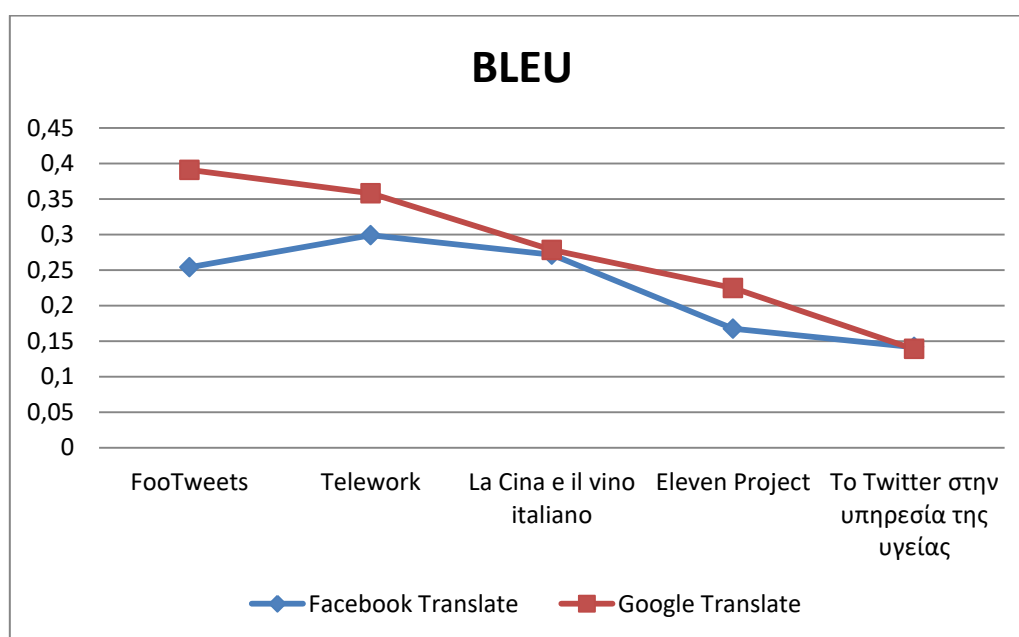
μές όμως των μετρικών κατά τη διαδικασία αυτόματης αξιολόγησης ήταν χαμηλότερες εξαιτίας αυτού ακριβώς του γεγονότος.

Υπήρξαν όμως και περιπτώσεις όπου αυτού του είδους η γραφή παρουσίασε ιδιαίτερο ενδιαφέρον. Χαρακτηριστική περίπτωση συνδυασμού αργκό και συντομογραφίας είναι η λέξη *Schiri* ,στο κείμενο *FooTweets*, η οποία αποτελεί συντόμευση της λέξης *Schiedsrichter* που σημαίνει *διαιτητής*. Το Google Translate την απόδωσε σε επίσημη γλώσσα ως *referee* ενώ η μεταφραστική μηχανή του Facebook διατήρησε τον ιδιωματικό της χαρακτήρα αποδίδοντας την ως *ref*. Άλλο ένα χαρακτηριστικό παράδειγμα είναι η παράλειψη της αποστρόφου στη φράση *I' uso* στο κείμενο του *Twitter στην υπηρεσία της υγείας*. Η φράση αναγράφηκε ως *Iuso*, μια λέξη χωρίς νόημα στα ιταλικά. Το Facebook Translate δεν αναγνώρισε το σφάλμα και την απόδωσε με λατινική γραφή ως *Iuse*, μια λέξη που επίσης δεν έχει κανένα νόημα ούτε στην ιταλική ούτε στην ελληνική γλώσσα. Το Google Translate αντίθετα το αναγνώρισε και την απόδωσε στα ελληνικά σωστά ως *η χρήση*.

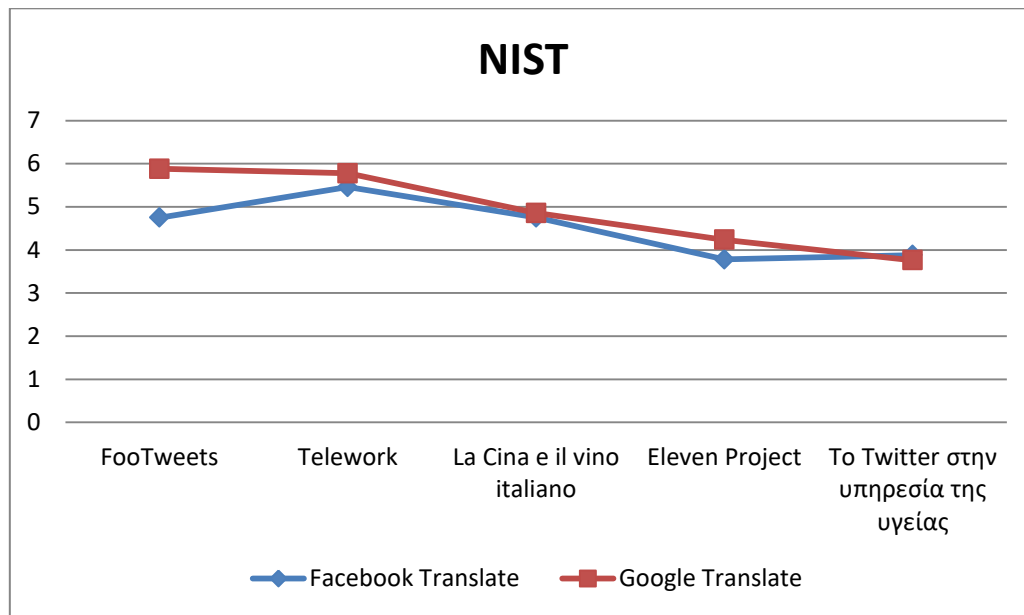
6.2 Συνολική Σύγκριση Facebook Translate και Google Translate

Κλείνοντας τη μελέτη των κειμένων παρουσιάζεται μία συνολική σύγκριση των δύο μεταφραστικών μηχανών με βάση την αυτόματη αλλά και την ανθρώπινη αξιολόγηση. Σκοπός της σύγκρισης είναι να δοθεί μια γενική εικόνα για τη συμπεριφορά των δύο μεταφραστικών εργαλείων καθώς χρησιμοποιήθηκαν για να μεταφράσουν κείμενα σε τρία διαφορετικά ζεύγη γλωσσών και για περισσότερες λεπτομέρειες πρέπει κάποιος να ανατρέξει στον σχολιασμό του κάθε κειμένου ξεχωριστά.

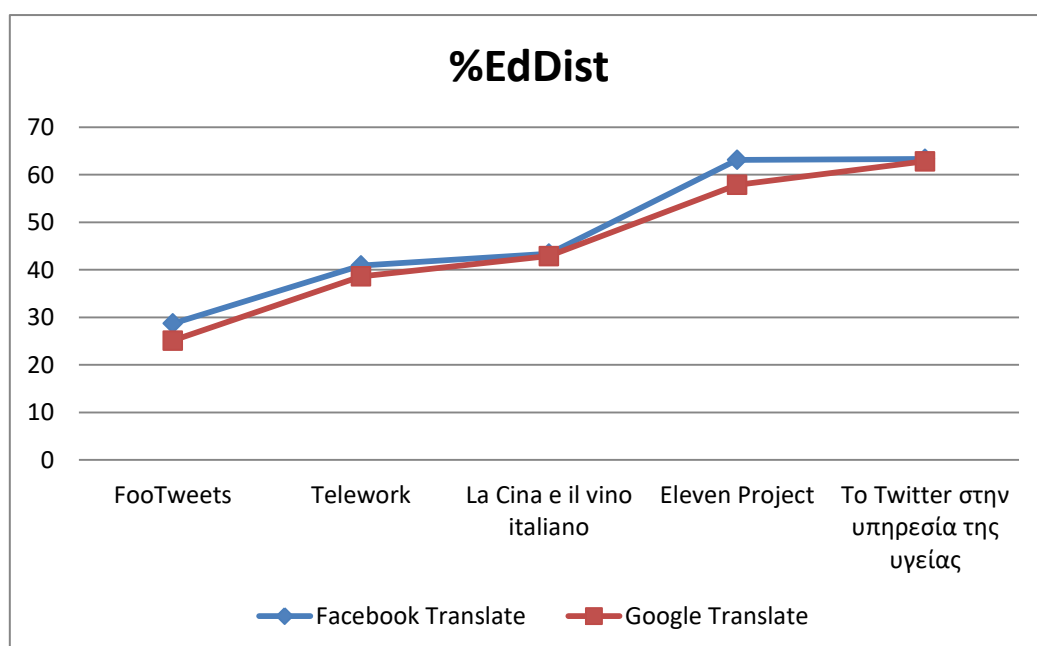
Στη συνολική σύγκριση παρουσιάζονται κάποια γραφήματα όπου απεικονίζεται η εικόνα της τιμής για κάθε μετρική αυτόματης αξιολόγησης ανάλογα με το κείμενο και τη μεταφραστική μηχανή που χρησιμοποιήθηκε. Όσον αφορά το κομμάτι της μη αυτόματης αξιολόγησης, παρουσιάζεται το σύνολο των τμημάτων κειμένων που αξιολογήθηκαν ως προς την ακρίβεια και ως προς την ευχέρεια.



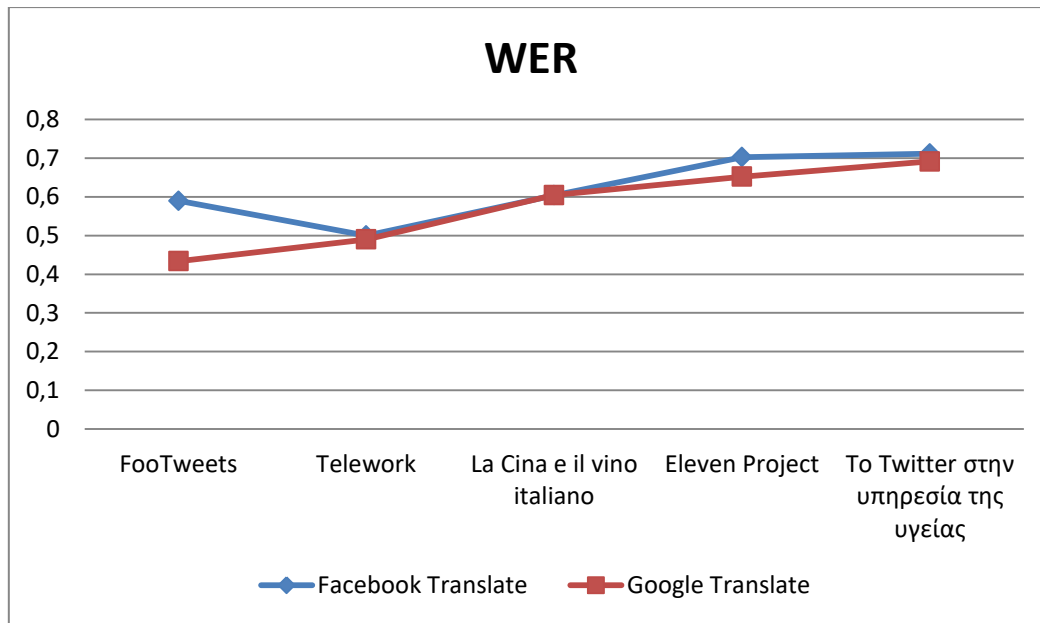
Όπως φαίνεται στο παραπάνω γράφημα, το Google Translate σημείωσε καλύτερες επιδόσεις σε σχέση με το Facebook Translate με βάση τη μετρική BLEU. Στα τρία από τα πέντε κείμενα έχει ξεκάθαρα υψηλότερες τιμές ενώ στα άλλα δύο οι τιμές είναι πολύ κοντά και σχεδόν ταυτίζονται. Ένα άλλο συμπέρασμα που βγαίνει από το εν λόγω γράφημα είναι ότι η τιμή της BLEU είναι χαμηλότερη στα κείμενα που αφορούν το ζεύγος ιταλικά-ελληνικά και για τις δύο μεταφραστικές μηχανές, γεγονός που κρίνεται φυσιολογικό μιας και οι δύο αυτές γλώσσες, και ειδικά τα ελληνικά, θεωρούνται λιγότερο «εμπορικές» από τα αγγλικά και τα γερμανικά.



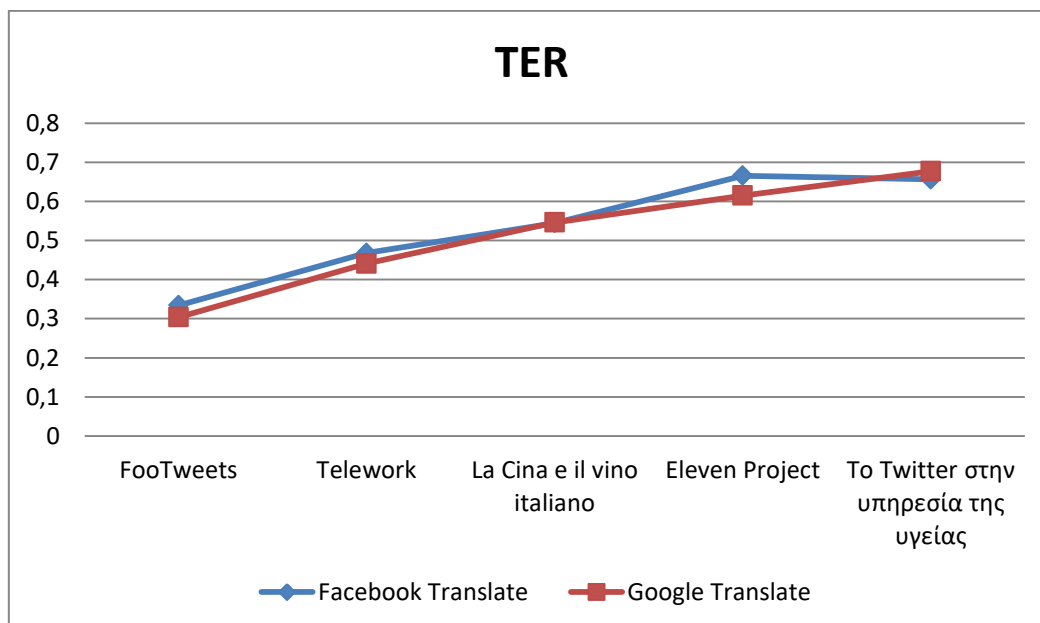
Παρόμοια εικόνα με τη BLEU παρουσιάζει και η NIST όπου βλέπουμε τις γραμμές με τη διακύμανση των τιμών της μετρικής να ακολουθούν περίπου την ίδια πορεία με τις αντίστοιχες της BLEU. Η συγκεκριμένη εικόνα είναι αναμενόμενη καθώς η NIST έχει παρόμοια φιλοσοφία με την BLEU.



Η γενική εικόνα των μετρικών που βασίζονται στην απόσταση διόρθωσης δείχνει και πάλι υπεροχή του Google Translate, αν και οι εικόνες στις τιμές μοιάζουν να έχουν μικρότερη απόκλιση. Στο παραπάνω γράφημα απεικονίζεται η διακύμανση της τιμής της %EdDist και φαίνεται καθαρά ότι η γραμμή του Facebook Translate βρίσκεται ψηλότερα από την αντίστοιχη του Google Translate, κάτι που συνεπάγεται καλύτερη απόδοση του τελευταίου. Αξιοσημείωτο είναι το γεγονός ότι σε δύο κείμενα οι τιμές σχεδόν ταυτίζονται. Πρόκειται μάλιστα για τα ίδια κείμενα στα οποία σχεδόν ταυτίζονται οι τιμές και των BLEU και NIST.

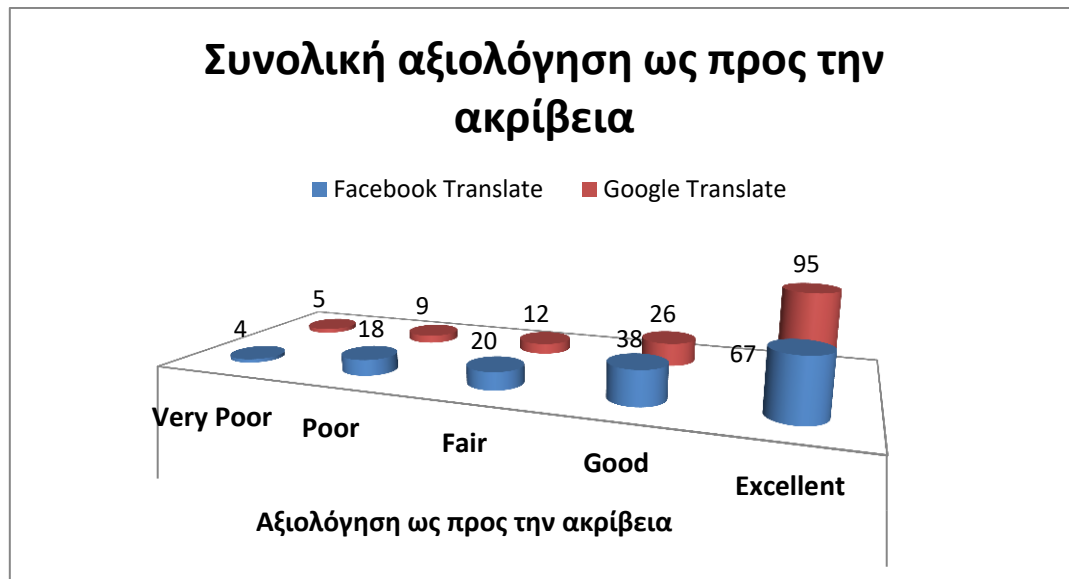


Παρόμοια εικόνα έχουμε και για τη μετρική WER με τη μεταφραστική μηχανή της Google να εμφανίζει ελαφριά υπεροχή. Πρέπει επίσης να σημειωθεί ότι παρατηρείται σχεδόν ταύτιση στις τιμές σε ένα ακόμα κείμενο, και πιο συγκεκριμένα στο *Telework*, και μάλιστα μεγαλύτερη σε σχέση με την ταύτιση που παρουσιάζεται στο *Twitter στην υπηρεσία της υγείας*.



Στο τελευταίο γράφημα παρουσιάζεται η συνολική εικόνα της TER. Είναι εμφανής και πάλι η ταύτιση των τιμών για το κείμενο *La Cina e il vino italiano* ενώ και πάλι η μεταφραστική μηχανή του Facebook υπολείπεται ελαφρώς. Η διαφοροποίηση σε σχέση με τις υπόλοιπες μετρικές καταγράφεται στο *Twitter στην υπηρεσία της υγείας* όπου το Facebook Translate παρουσιάζει σχετικά καλύτερη εικόνα σε σχέση με το εργαλείο της Google.

Η μελέτη των κειμένων ολοκληρώνεται με δύο γραφήματα στα οποία απεικονίζεται η συνολική εικόνα της ανθρώπινης αξιολόγησης ως προς την ακρίβεια και ως προς την ευχέρεια σε όλα τα κείμενα.



Όπως και στα γραφήματα όπου απεικονίστηκε η συνολική εικόνα των μετρικών αυτόματης αξιολόγησης, έτσι και σε αυτά της ανθρώπινης αξιολόγησης, το Google Translate παρουσιάζει καλύτερη εικόνα υπερτερώντας εμφανώς στα τμήματα κειμένου που έλαβαν άριστη αξιολόγηση τόσο ως προς την ακρίβεια όσο και ως προς την ευχέρεια.

Σε κάθε περίπτωση, υπάρχουν πάντοτε περιθώρια βελτίωσης και για τις δύο μεταφραστικές μηχανές αλλά και για όλες τις μεταφραστικές μηχανές γενικότερα καθώς η διαδικασία μηχανικής μετάφρασης παραμένει μια «ζωντανή» διαδικασία η οποία εξελίσσεται και βελτιώνεται συνεχώς. Το Facebook Translate μπορεί να υστερεί στην παρούσα μελέτη αλλά η εταιρεία που το αναπτύσσει σύστησε τον Οκτώβριο του 2020 για πρώτη φορά το M2M-100 (Fan 2020), ένα νέο σύστημα βασιζόμενο στο πολυγλωσσικό μοντέλο μηχανικής μετάφρασης Many-to-Many (Many-to-Many multilingual translation model). Το συγκεκριμένο μοντέλο βασίζεται στην απευθείας μετάφραση από μια γλώσσα σε μια άλλη χωρίς την παρεμβολή της αγγλικής γλώσσας και αναφέρεται μάλιστα χαρακτηριστικά ότι στις δοκιμές συγκέντρωσε βαθμολογία BLEU έως και 10 μονάδες υψηλότερη σε σχέση με τα αγγλοκεντρικά συστήματα. Την ίδια περίοδο, το Google μελετά τρόπους να βελτιώσει τη νευρωνική μηχανική μετάφραση χρησιμοποιώντας κείμενα αναφοράς παραφρασμένα από ανθρώπους (Human-Paraphrased References, Freitag et al. 2020). Το μέλλον της μηχανικής μετάφρασης διαγράφεται γεμάτο προκλήσεις και αναμένεται με μεγάλο ενδιαφέρον η πορεία των εξελίξεων.

7.

Εκτίμηση Ποιότητας

Όπως φάνηκε και από τη μελέτη των κειμένων που παρουσιάστηκε στα προηγούμενα κεφάλαια, οι μετρικές αυτόματης αξιολόγησης είναι σε θέση να δώσουν πολύ ικανοποιητικά αποτελέσματα σχετικά με την ποιότητα των μηχανικά μεταφρασμένων κειμένων. Το πρόβλημα βεβαίως είναι ότι για να αξιοποιηθούν είναι απαραίτητη και η ύπαρξη ανθρώπινων μεταφράσεων αναφοράς. Η έλλειψη κειμένων αναφοράς είναι άλλωστε ο λόγος για τον οποίο τα κείμενα που μελετήθηκαν, με εξαίρεση το απόσπασμα από το *FooTweets*, αντλήθηκαν από ιστοσελίδες στο διαδίκτυο και όχι από μέσα κοινωνικής δικτύωσης. Η εκτίμηση ποιότητας είναι μια μέθοδος που χρησιμοποιείται για να παρέχει αυτόματα ενδείξεις ποιότητας που αφορούν τη μηχανική μετάφραση χωρίς να βασίζεται σε ανθρώπινες μεταφράσεις αναφοράς. Με άλλα λόγια, είναι ένας τρόπος να εκτιμηθεί, χωρίς ανθρώπινη παρέμβαση, το πόσο καλές ή κακές είναι οι μεταφράσεις που παράγονται από ένα σύστημα μηχανικής μετάφρασης.

Αν και η εκτίμηση ποιότητας έχει ουσιαστικά τον ίδιο στόχο με την αξιολόγηση ενός μηχανικά μεταφρασμένου κειμένου, εντούτοις οι δύο διαδικασίες παρουσιάζουν χαρακτηριστικές διαφορές. Η ποιότητα της μηχανικής μετάφρασης μπορεί να αξιολογηθεί με δύο βασικούς τρόπους: την ανθρώπινη αξιολόγηση (όταν κάποιος άνθρωπος ελέγχει τη μετάφραση και παρέχει ανατροφοδότηση) και την αυτόματη αξιολόγηση με διάφορες μεθόδους που μπορούν δώσουν κάποια βαθμολογία στην ποιότητα της μετάφρασης χωρίς ανθρώπινη παρέμβαση.

Η αυτόματη αξιολόγηση του μηχανικά μεταφρασμένου κειμένου απαιτεί παραδοσιακά την ύπαρξη ενός τουλάχιστον κειμένου αναφοράς, μιας μετάφρασης δηλαδή του πρωτότυπου κειμένου από κάποιον άνθρωπο-μεταφραστή. Οι διαφορές και οι ομοιότητες μεταξύ του μηχανικά μεταφρασμένου κειμένου και της μετάφρασης αναφοράς επεξεργάζονται μέσω κατάλληλα σχεδιασμένων αλγορίθμων και μέσα από αυτή τη διαδικασία παράγεται κάποια βαθμολογία η οποία προσδιορίζει την ποιότητα του μηχανικού μεταφράσματος. Η εν λόγω προσέγγιση ακολουθείται από τις πιο χαρακτηριστικές μεθόδους αυτόματης αξιολόγησης με τη χρήση αυτόματων μετρικών όπως η BLEU και η NIST.

Ο παράγοντας που διαφοροποιεί την εκτίμηση ποιότητας από την αξιολόγηση ποιότητας ενός μηχανικά μεταφρασμένου κειμένου είναι το γεγονός ότι για την εκτίμηση ποιότητας δεν απαιτείται μετάφραση αναφοράς από άνθρωπο-μεταφραστή. Αναφέρεται χαρακτηριστικά ότι η εν λόγω διαδικασία «συνίσταται στην εκτίμηση της ποιότητας των αποτελεσμάτων εξόδου ενός συστήματος μηχανικής μετάφρασης για συγκεκριμένα δεδομένα εισόδου, χωρίς καμία πληροφορία σχετικά με τα αναμενόμενα αποτελέσματα εξόδου» (Specia et al., 2010).

Η εκτίμηση ποιότητας είναι ουσιαστικά μια πρόβλεψη ποιότητας βάσει ορισμένων χαρακτηριστικών όπως, για παράδειγμα, ο αριθμός ουσιαστικών ή προθετικών φράσεων στο κείμενο πηγή και στο μηχανικό μετάφρασμα (και η διαφορά τους), ο αριθμός των επώνυμων οντοτήτων (τοποθεσίες, άτομα, εταιρείες κτλ) και η παρουσία τους στα δύο κείμενα και πολλά άλλα. Με τη χρήση τεχνικών όπως η μηχανική μάθηση έχουν αναπτυχθεί διάφο-

ρα μοντέλα εκτίμησης ποιότητας τα οποία εξάγουν βαθμολογίες οι οποίες αντιπροσωπεύουν τις εκτιμήσεις για την ποιότητα ενός κειμένου που παράγεται από ένα σύστημα μηχανικής μετάφρασης.

7.1 Πλεονεκτήματα

Η εκτίμηση ποιότητας μπορεί να προσφέρει άμεσα σημαντικά πλεονεκτήματα και δυνατότητες όπως οι παρακάτω:

- Να αποφασίσει αν το μηχανικά μεταφρασμένο κείμενο είναι κατανοητό σε ικανοποιητικό βαθμό για τον αναγνώστη (Specia 2013)
- Να κάνει πιο αποτελεσματική τη διαδικασία της μετεπεξεργασίας του μηχανικά μεταφρασμένου κειμένου υποδεικνύοντας τμήματα του κειμένου που απαιτούν μεγαλύτερη προσπάθεια ή χρόνο να υποστούν μετεπεξεργασία σε σχέση με το χρόνο ή την προσπάθεια που θα απαιτούνταν αν μεταφράζονταν από άνθρωπο εξαρχής (Specia et al. 2009, Specia 2011)
- Να επιλέξει τα υψηλής ποιότητας τμήματα του μηχανικά μεταφρασμένου κειμένου (Soricut & Echihabi 2010)
- Να επιλέξει το πιο κατάλληλο από κείμενα που παράγονται από διαφορετικά συστήματα μηχανικής μετάφρασης (Shah & Specia 2014)
- Να επιλέξει αν είναι προτιμότερη η μετάφραση ενός κειμένου από ένα σύστημα μηχανικής μετάφρασης ή με βάση κάποια μεταφραστική μνήμη (He et al. 2010)
- Να αποφασίσει αν μια μετάφραση μπορεί να χρησιμοποιηθεί για την εκπαίδευση ενός συστήματος μηχανικής μετάφρασης (Specia 2013)
- Να υποδείξει λέξεις ή φράσεις στο μηχανικά μεταφρασμένο κείμενο που χρειάζονται επανεξέταση (Bach et al. 2011)

7.2 Επίπεδα ανάλυσης εκτίμησης ποιότητας

Η εκτίμηση της ποιότητας ενός κειμένου μπορεί να πραγματοποιηθεί σε διάφορα επίπεδα. Πιο συγκεκριμένα, μπορεί να πραγματοποιηθεί σε επίπεδο λέξεων, φράσεων, προτάσεων, ακόμα και σε επίπεδο ολόκληρων εγγράφων. Σε επίπεδο λέξεων, η εκτίμηση ποιότητας προσπαθεί να κάνει προβλέψεις ανάλογα με το αν οι λέξεις του κειμένου έχουν μεταφραστεί σωστά ή όχι απονέμοντας μια δυαδική ετικέτα σε κάθε λέξη. Λαμβάνονται επίσης υπόψη σφάλματα που προκύπτουν λόγω παράλειψης μετάφρασης κάποιων λέξεων ή παντελούς απουσίας τους από το μετάφρασμα. Η εκτίμηση ποιότητας σε επίπεδο φράσεων αποσκοπεί στην πρόβλεψη της ποιότητας των μεταφρασμένων φράσεων και προκύπτει από τα αποτελέσματα της εκτίμησης σε επίπεδο λέξεων. Για πιο λεπτομερή αποτελέσματα λαμβάνονται υπόψη η λανθασμένη σειρά λέξεων στο μεταφρασμένο κείμενο καθώς και λέξεις που ενδεχομένως να λείπουν.

Σε επίπεδο προτάσεων, η εκτίμηση ποιότητας στοχεύει στην απόδοση μιας βαθμολογίας σε μια μεταφρασμένη πρόταση με βάση τον αριθμό των λέξεων που πρέπει να αλλάξουν θέση ώστε η πρόταση να ταιριάζει με το κείμενο που παρέχεται στην πλησιέστερη μετάφραση αναφοράς. Στην πράξη, αφού δεν υπάρχει κείμενο αναφοράς, στόχος είναι η πρόβλεψη της βαθμολογίας TER (ή HTER) του εξεταζόμενου μεταφρασμένου τμήματος κειμένου η οποία

εκφράζει την απόσταση επεξεργασίας μεταξύ του εν λόγω τμήματος και μιας αντίστοιχης ιδανικής μετάφρασης αναφοράς.

Εκτίμηση ποιότητας μπορεί να γίνει ακόμα και σε επίπεδο εγγράφου. Σε αυτήν την περίπτωση η εκτίμηση ποιότητας επικεντρώνεται στον εντοπισμό σφαλμάτων σε ένα μηχανικά μεταφρασμένο κείμενο και στην εκχώρηση μιας βαθμολογίας ανάλογα με τον τύπο και τη σοβαρότητα των εν λόγω σφαλμάτων. Εξετάζονται η ακρίβεια, η ευχέρεια και το στυλ (για παράδειγμα το ύφος της γλώσσας) του μεταφρασμένου κειμένου καθώς και ο αντίκτυπος των σφαλμάτων που εντοπίστηκαν στο συνολικό νόημα του κειμένου.

7.3 Πλαίσια ανοιχτού κώδικα για εκτίμηση ποιότητας

Η ανάπτυξη πλαισίων για την εκτίμηση ποιότητας απαιτεί υψηλά επίπεδα τεχνογνωσίας και εμπειρίας πάνω στις μεθόδους μηχανικής μάθησης όμως τα τελευταία χρόνια έχουν αναπτυχθεί και δημοσιευτεί μερικά πλαίσια ανοιχτού κώδικα τα οποία είναι διαθέσιμα και ελεύθερα για χρήση, συνεισφορά και προτάσεις βελτίωσης. Τα εν λόγω συστήματα διευκολύνουν την εφαρμογή διαφόρων μεθόδων και επιπέδων εκτίμησης και επιτρέπουν ελέγχους και πειραματισμούς πάνω σε γλωσσικά δεδομένα για την εξαγωγή συμπερασμάτων.

Ένα από τα πρώτα πλαίσια ανοιχτού κώδικα για εκτίμηση ήταν το Quest (Specia et al. 2013), μετεξέλιξη του οποίου αποτέλεσε το Quest++ (Specia et al. 2016). Βασίζεται κατά κύριο λόγο στα γλωσσικά και στα κειμενικά χαρακτηριστικά και αποτελείται από δύο μέρη, τη μονάδα εξαγωγής χαρακτηριστικών και τη μονάδα μηχανικής μάθησης. Η μονάδα εξαγωγής χαρακτηριστικών έχει τη δυνατότητα να εξαγάγει μια πληθώρα χαρακτηριστικών από το κείμενο, σε τρία επίπεδα ανάλυσης. Σε επίπεδο λέξεων, προτάσεων και εγγράφου. Τα δεδομένα αυτά μεταβιβάζονται στη συνέχεια στο τμήμα μηχανικής μάθησης το οποίο τα χρησιμοποιεί για να δημιουργήσει ένα αξιόπιστο μοντέλο εκτίμησης ποιότητας.

Το πλαίσιο QUETCH (Kreutzer et al. 2015) αποτελείται από ένα τεχνητό νευρωνικό δίκτυο και έχει τη δυνατότητα να μαθαίνει με βάση χαρακτηριστικά εκτίμησης ποιότητας που εισάγονται από ακατέργαστο κείμενο. Περιορίζεται όμως στο επίπεδο λέξεων καθώς δεν μπορεί να υποστηρίξει υψηλότερα επίπεδα ανάλυσης.

Το deepQuest (Ive et al. 2018) είναι το τελευταίο και πιο σύγχρονο πλαίσιο που αναπτύχθηκε από την ομάδα ανάπτυξης των πλαισίων Quest που αναφέρθηκαν παραπάνω. Δημιουργήθηκε για εκτίμηση ποιότητας σε επίπεδο εγγράφου, μπορεί παρόλα αυτά να χρησιμοποιηθεί και να παράγει συμπεράσματα και στα υπόλοιπα επίπεδα. Παρέχει δύο διαφορετικές αρχιτεκτονικές νευρωνικής μηχανικής μάθησης και είναι σαφώς πιο γρήγορο στην εκπαίδευσή του σε σχέση με τα συγκριτικά μοντέλα.

Τέλος, ένα από τα πιο σύγχρονα πλαίσια ανοιχτού κώδικα είναι το OpenKiwi (Keppler et al. 2019), ίσως το πιο προσιτό και ολοκληρωμένο ως εφαρμογή. Υλοποιεί τέσσερις αρχιτεκτονικές μηχανικής μάθησης που έχουν τη δυνατότητα να κάνουν εκτίμηση ποιότητας στα διάφορα επίπεδα του κειμένου, έναν αριθμό προ-εκπαιδευμένων μοντέλων καθώς και διάφορους τρόπους τροποποίησης των δεδομένων και των πειραματικών ρυθμίσεων. Μπορεί να χρησιμοποιηθεί ως πακέτο Python αλλά και να εκτελεστεί από τη γραμμή εντολών ενός υπολογιστή.

Βιβλιογραφία

- Bach, N., Huang, F. & Al-Onaizan, Y. (2011). "Goodness: a method for measuring machine translation confidence" in: *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics*, pp 211–219. Portland, OR
- Bahdanau, D., [Kyunghyun, C.](#) & Bengio, Y. (2014). "Neural machine translation by jointly learning to align and translate" in: *International Conference on Learning Representations (ICLR)*. San Diego, CA
- 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 evaluation measures*, pp 65-72
- Brown, E. (2015). "Twitter brings Tweet translation to iOS, Android and Windows 8 phones" in: <https://www.zdnet.com/article/twitter-brings-tweet-translation-to-ios-android-and-windows-8-phones/> (τελευταία προσπέλαση: 26/03/2021)
- Brownlee, J. (2017). "A Gentle Introduction to Neural Machine Translation" in *Deep Learning for Natural Language Processing*, <https://machinelearningmastery.com/introduction-neural-machine-translation/> (τελευταία προσπέλαση: 26/03/2021)
- Castilho, S. (2017). "Human Evaluation: Why do we need it?", Talk at the 8th NLP Dublin meetup in: *Human & Linguistic Factors: The 8th NLP Dublin Meetup*, Dublin 2017, <https://www.slideshare.net/SebastianRuder/human-evaluation-why-do-we-need-it-dr-sheila-castilho> (τελευταία προσπέλαση: 26/03/2021)
- Castilho, S. (2019). "Machine Translation Evaluation: know your essentials, Tutorial at the RANLP '19" in: *Recent Advances in Natural Language Processing*. Varna, Bulgaria
- Dean, B. (2020). "Social Network Usage & Growth Statistics: How Many People Use Social Media in 2020?" in: <https://backlinko.com/social-media-users> (τελευταία προσπέλαση: 26/03/2021)
- Dorr, B., Snover, M. & Madnani, N. (2011). "Chapter One: Introduction" in: Dorr, B. (ed) "Part 5: Machine Translation Evaluation" in: Olive, J., McCary, J. and Christianson, C. (eds.) (2011). *Handbook of Natural Language Processing and Machine Translation: DARPA Global Autonomous Language Exploitation*, New York: Springer
- Espa-Gomis, M., Forcada, M., Ramirez-Sancez, D. & Hoang, H. (2019). "ParaCrawl: Web-scale parallel corpora for the languages of the EU" in: *Proceedings of MT Summit XVII, volume 2*, pp 118-119. Dublin, Ireland
- Fan, A. (2020). "Facebook: Introducing the First AI Model That Translates 100 Languages Without Relying on English" in: *Facebook Company Web Page*, <https://about.fb.com/news/2020/10/first-multilingual-machine-translation-model/> (τελευταία προσπέλαση: 26/03/2021)

- Freitag, M., Foster, G., Grangier, D., Cherry, C. Google Research (2020). “Human-Paraphrased References Improve Neural Machine Translation” in: *arXiv:2010.10245v1 [cs.CL]* 20 Oct 2020
- Goodfellow, I., Bengio, Y. & Courville, A. (2016). *Deep Learning (Adaptive Computation and Machine Learning series)*. Cambridge, MA: The MIT Press
- He, Y., Ma, Y., van Genabith, J. & Way, A. (2010). “Bridging SMT and TM with Translation Recommendation” in: *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, pp 622–630. Uppsala, Sweden
- Hern, A. (2017). “Chinese messaging app error sees n-word used in translation” in: *The Guardian, International Edition*, 13 October, <https://www.theguardian.com/technology/2017/oct/13/chinese-messaging-app-error-sees-n-word-used-in-translation> (τελευταία προσπέλαση: 26/03/2021)
- Hern, A. (2017). “Facebook translates 'good morning' into 'attack them', leading to arrest” in: *The Guardian, International Edition*, 24 October, <https://www.theguardian.com/technology/2017/oct/24/facebook-palestine-israel-translates-good-morning-attack-them-arrest> (τελευταία προσπέλαση: 26/03/2021)
- Hovy, E., King, M. & Popescu-Belis, A. (2002). “An Introduction to MT Evaluation” in: *Workshop of the LREC 2002 Workshop on Machine Translation Evaluation: Human Evaluators Meet Automated Metrics*, pp. 1-7. Las Palmas de Gran Canaria, Spain
- Ive, J., Blain, F. & Specia, L. (2018). “deepQuest: A Framework for Neural-based Quality Estimation” in: *Proceedings of the 27th International Conference on Computational Linguistics*, pp 3146–3157. Santa Fe, NM
- Kashif S. & Specia, L. (2014). “Quality estimation for translation selection” in: *“Proceedings of the 17th Annual Conference of the European Association for Machine Translation”*, pp 109-116. Dubrovnik, Croatia
- Kemp, S. (2020). “Digital 2020: Global Digital Overview” in: <https://datareportal.com/reports/digital-2020-global-digital-overview> (τελευταία προσπέλαση: 26/03/2021)
- Kepler, F., Trenous, J., Treviso, M., Vera, M. & Martins, A. F. T. (2019). “OpenKiwi: An Open Source Framework for Quality Estimation” in: *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, pp 117–122. Florence, Italy
- Koehn, P. (2020). *Neural Machine Translation*. Cambridge: Cambridge University Press
- Koehn, P. (2010). *Statistical Machine Translation*. Cambridge: Cambridge University Press
- Koehn, P., Heafield, K., Forcada, M. et al. (2018), *ParaCrawl Corpus version 1.0*, LINDAT/CLARIAH-CZ digital library at the Institute of Formal and Applied Linguistics (ÚFAL),

Faculty of Mathematics and Physics, Charles University, <http://hdl.handle.net/11372/LRT-2610> (τελευταία προσπέλαση: 26/03/2021)

Kreutzer, J., Schamoni, S. & Riezler, S. (2015). "QUality Estimation from ScraTCH (QUETCH): Deep Learning for Word-level Translation Quality Estimation" in: *Proceedings of the Tenth Workshop on Statistical Machine Translation*, pp 316–322. Lisbon, Portugal

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

Levenshtein, V. (1966). "Binary codes capable of correcting deletions, insertions, and reversals" in: *Soviet Physics - Doklady* 10 (8): pp. 707–710. Original in Russian: *Doklady Akademii Nauk SSSR*, V163 No4 845-848, 1965

Maucec, M. S. & Donaj, G. (2019). "Machine Translation and the Evaluation of its Quality" in: Sadollah, A. & Sinha, T. *Recent Trends in Computational Intelligence*, Open access peer-reviewed Edited Volume, <https://www.intechopen.com/books/recent-trends-in-computational-intelligence/machine-translation-and-the-evaluation-of-its-quality> (τελευταία προσπέλαση: 26/03/2021)

Moorkens, J., Castilho, S., Gaspari, F. & Doherty, S. (eds.) (2018). *Translation Quality Assessment – From Principles to Practice*. New York: Springer

Nießen, S., Och, F., Leusch, G. & Ney, H. (2000). "An evaluation tool for machine translation: Fast evaluation for mt research" in: *Proceedings of the 2nd International Conference on Language Resources and Evaluation*. Athens, Greece

Och, F. J. (2003). "Minimum error rate training in statistical machine translation" in *Proceedings of the 41st Annual Meeting of the ACL*, pp 160–167

Olive, J., McCary, J. and Christianson, C. (eds.) (2011). *Handbook of Natural Language Processing and Machine Translation: DARPA Global Autonomous Language Exploitation*. New York: Springer

Oliver, A. (2020). "MTUOC: easy and free integration of NMT systems in professional translation environments" in: *Proceedings of the 22nd Annual Conference of the European Association for Machine Translation*. Lisbon, Portugal

Oliver, A. (2020). "MTUOC: Tools for training and translating with NMT and SMT systems", <https://sourceforge.net/projects/mtuoc/> (τελευταία προσπέλαση: 26/03/2021)

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*, pp. 311–318. Philadelphia, PA

Prokopidis, P., Papavassiliou, V. & Piperidis, S. (2016). «Parallel Global Voices: a Collection of Multilingual Corpora with Citizen Media Stories» in: *Proceedings of the Tenth International Conference on Language Resources and Evaluation*, pp 900-905. Portorož, Slovenia

Russell, S. & Norvig, P. (2009). *Artificial Intelligence: A Modern Approach*, 3rd Edition. London: Pearson

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

Sluyter-Gäthje, H., Lohar, P., Afli, H., & Way, A. (2018). «FooTweets: A bilingual parallel corpus of worldcup tweets» in: *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, pp 2666-2670. Miyazaki, Japan

Snover, M., Dorr, B., Schwartz, R., Micciulla, L. & 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*. Cambridge, MA

Soricut, R. & Echiabi, A. (2010). “Trustrank: Inducing trust in automatic translations via ranking” in: *ACL 2010 - 48th Annual Meeting of the Association for Computational Linguistics, Proceedings of the Conference*, pp 612–621. Uppsala, Sweden

Specia, L. (2011). “Exploiting objective annotations for measuring translation post-editing effort” in *Proceedings of the 15th Conference of the European Association for Machine Translation*, pp 73–80. Leuven, Belgium

Specia, L., Paetzold, G. H. & Scarton C. (2015). “Multi-level Translation Quality Prediction with QUEST++” in: *Proceedings of ACL-IJCNLP 2015 System Demonstrations*, pp 115–120. Beijing, China

Specia, L., Raj, D. & Turchi, M. (2010). «Machine translation evaluation versus quality estimation» in: *Machine Translation 24*, pp. 39–50. New York: Springer

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

Specia, L., Turchi, M., Cancedda, N., Dymetman, M. & Cristianini, N. (2009). “Estimating the Sentence-Level Quality of Machine Translation Systems” in: *Proceedings of the 13th Annual Conference of the European Association for Machine Translation*, pp 28-37. Barcelona, Spain

Tortermvasana, K. (2020). “FB told to apologise for 'mistranslation'” in: *Bangkok Post*, 31 July, <https://www.bangkokpost.com/thailand/general/1960155/fb-told-to-apologise-for-mistranslation> (τελευταία προσπέλαση: 26/03/2021)

Vashee, K. (2019). “Understanding MT Quality: BLEU Scores” in: *SDL Trados Blog*, <https://blog.sdl.com/blog/understanding-mt-quality-bleu-scores.html> (τελευταία προσπέλαση: 26/03/2021)

Williams, P., Senrnrich, R., Post, M. & Koehn, P. (2016). *Syntax-based Statistical Machine Translation (Synthesis Lectures on Human Language Technologies)*. San Rafael, CA: Morgan & Claypool Publisher

Σύνδεσμοι

Λεξικά – Γλωσσικές Βάσεις Δεδομένων

<https://glosbe.com/>

<https://iate.europa.eu/home>

<https://pons.com>

<https://www.dict.cc/>

<https://www.duden.de/>

<https://www.linguee.com/>

<https://www.treccani.it/vocabolario/>

<https://www.wordreference.com/engr/>

Μεταφραστικές Μηχανές

<https://translate.google.gr>

<https://www.facebook.com/>

Εργαλεία Αυτόματης Αξιολόγησης και Εκτίμησης Ποιότητας

<https://github.com/aoliverg/MTUOC>

<https://github.com/juliakreutzer/quetch>

<https://github.com/Unbabel/OpenKiw>

<https://www.quest.dcs.shef.ac.uk/>

Οργανισμοί και πρωτοβουλίες που σχετίζονται με τη μηχανική μετάφραση

American Machine Translation Association - <https://amtaweb.org/>

Association For Computational Linguistics - <https://www.aclweb.org>

European Association for Machine Translation - <http://eamt.org/>

International Conference on Language Resources and Evaluation - <http://www.lrec-conf.org/>

Statistical Machine Translation, a website dedicated to research in statistical machine translation - <http://www.statmt.org/>

Multidimensional Quality Metrics (MQM) Definition

<http://www.qt21.eu/mqm-definition/definition-2015-12-30.html>

Στατιστικά Στοιχεία

<https://backlinko.com/social-media-users>

<https://datareportal.com/reports/digital-2020-global-digital-overview>

<https://www.marketingprofs.com/charts/2014/25346/the-most-popular-topics-on-facebook-and-twitter>

Παράρτημα

Στο παράρτημα, για κάθε ένα από τα κείμενα που μελετήθηκαν παρατίθεται το κείμενο στη γλώσσα αφετηρίας, το κείμενο αναφοράς, τα μεταφράσματα και η αξιολόγηση ως προς την ακρίβεια και την ευχέρεια των μηχανικών μεταφρασμάτων και για τις δύο μεταφραστικές μηχανές, καθώς και η αντίστοιχη κατηγοριοποίηση λαθών. Για λόγους οικονομίας σελίδων, τα μεταφράσματα για κάθε μεταφραστική μηχανή παρατίθενται παράλληλα με την αξιολόγησή τους ενώ στην κατηγοριοποίηση λαθών παρουσιάζονται μόνο οι κατηγορίες όπου εντοπίστηκαν σφάλματα. Τα πρωτότυπα αρχεία excel επισυνάπτονται ως ξεχωριστά αρχεία στην ηλεκτρονική έκδοση της εργασίας. Στο τέλος του παραρτήματος παρατίθεται επίσης ο κώδικας για τον Greek Tokenizer.

FooTweets

Κείμενο Αφετηρίας

es ist total deprimierend , zu sehen , dass \$ mein # Mexico den # Worldcup verlassen hat , aber niemand kann leugnen , sie haben gekämpft wie die Löwen ! ! ! ! ! # VamosMexico

verdammte Scheiße # Brazil , ich weiß , ihr habt vergessen , einige Stadien fertig zubauen , aber habt ihr auch vergessen \$ euer Team zu trainieren ? # Worldcup # BrazilvsMexico

@ themoonexpress Lass mich raten ... du bist kein Fan des südamerikanischen Tanzes ? # sarcasm # Worldcup

ich brauche eine Wiederholung mit Torlinientechnik für dieses # robben Tor , Ich bin nicht sicher , ob er drin war # Worldcup

kein Abseits von Higuain ! 🤔 der verdammte Schiedsrichter hat es verdorben ! # Worldcup # ARG

ich denke \$ es ist zu früh für schlechten Ballismus ... # Worldcup

der einzige Grund , warum ich mich nach diesem langweiligen Spiel nicht umgebracht habe , ist das # USA in einer Stunde spielt . # Worldcup

bringt einen Gedanken für die Königin Maxima der Niederlande auf , da ich argentinisch bin , sage ich , es kann sein , dass sie nach dem Halbfinale keine Königin ist ! : wooooo Orange ! Weltmeisterschaft

nach einer 5-1 Niederlage gegen die Holländer , muss sich Spanien so fühlen [http : / / t.co / urc1aV7uJ9](http://t.co/urc1aV7uJ9) # Worldcup

und die Geschichte wiederholt sich mit den Schweizern , sie entscheiden sich , sich nicht zu wehren .. Zing # Worldcup # SUIvsFRA

es ist so seltsam zu sehen , dass # MEX rot und schwarz als \$ ihre alternativen Farben tragen . # Worldcup

Frankreich verschießt einen Elfmeter , trotzdem noch 2-0 für Frankreich # Worldcup

wir glauben , \$ unser # Belgium Chef ist ein wenig verwirrt darüber , bei wem er mitfiebern soll in dem # BELvsUSA Spiel . # Worldcup [http : / / t.co / zm0sVFGJXb](http://t.co/zm0sVFGJXb)

Siri besitzt einige Frechheit . 🤖 . # pummeled # brazil # germany # Worldcup [http : / / t.co / EGjMM6fF9u](http://t.co/EGjMM6fF9u)

erste Wasserpause in der Geschichte der Weltmeisterschaft ... Nicht sicher , wie ich darüber denke . # Worldcup # USAvsPOR

Worldcup | Ghana leidet unter ihrer 2 - 1 Niederlage gegen die USA : Ghanas schwarze Stars verloren \$ ihr erstes Spiel der Weltmeisterschaft 2014 : ... [http : / / t.co / GLVO5G8wBE](http://t.co/GLVO5G8wBE)

tut mir leid Brasilien veranstalte nächstes Mal keine Weltmeisterschaft # Worldcup Host # SSFootball

ich hätte nie gedacht , dass Spanien jemals in einem Spiel so unüberzeugend \$ aussehen würden # Worldcup # ESP

Spanien vs Niederlande 1-5 # WTF # omg # Worldcup lol

kann ich dich bitte mit einem Klumpen Cheddar-Käse bekanntmachen ! los England ! # Worldcup

OK eine kleine Unaufmerksamkeit : - mal sehen , wie tor-hungrig das Team sein wird : - es geht nur darum , wie sehr man es will # Worldcup

Herzinfarkt ist die einzige Option # Worldcup

unnütz ! England nutzlos ! # eng # Worldcup

unglückliches Brasilien , es war Klose

warum hat der Schiedsrichter aufgehört dort zuspitzen ? es ist keine Kopfverletzung . es lässt sich darüber streiten , ob das überhaupt eine Verletzung war ! # Worldcup

spanisches wurde das Tor nicht gegeben ; Abseits # Worldcup # sucks # ESPNED

wenn man in diesem # Worldcup nur zwei Tore macht , sollte man wahrscheinlich damit rechnen bald nach Hause zu fahren . # ITAURU

warte was ? nicht , wenn er in den Ruhestand gehen muss # Gerrard # Worldcup

verdammt # Worldcup hier und # Worldcup da. hab es jetzt schon satt

zum ersten Mal hat mich das zur Toilette gehen gekriegt und ich habe das Tor verpasst ! # Worldcup # England

USA Man of the Match ~ NO ONE ! # scmotm # WorldCupProblems # Worldcup # ESPN

Messis Eilmeldung an den verletzten Neymar [http : / / t.co / QILAKzDb5e](http://t.co/QILAKzDb5e)

lol , ich habe dieses Spiel nur geguckt , um mich über das Team , das verliert , lustig zu machen.Ich war darüber noch nie so begeistert , TBH . ▪ # Worldcup

WK2014 : van Persie : die Niederländer müssen eine harte Zeit erwarten [http : / / t.co / gREqhKADkY](http://t.co/gREqhKADkY)

süßes Italien hat gerade einen kleinen Zittern # avicii # Worldcup

Howard Webb ist ein schrecklicher Schiri # Worldcup

darf ich der erste sein der sagt , Totalschaden für Spanien ! # Spainvschile # Worldcup

die Leute sagen , es ist sinnlos das @ england Spiel morgen zu gucken . ja , sie sind raus aus dem # Worldcup aber ich warte immer noch darauf

@ MiseryBear Armer kleine Bär . vielleicht könnten du und @ officialkylie einander helfen , das Elend zu bedauern # Worldcup

sind Brasilien und England verkleidet ? # Worldcup

Κείμενο Αναφοράς

complete depression to see \$ my # Mexico leave the # WorldCup , but no one can deny they fought like lions ! ! ! ! ! # VamosMexico

fucking hell # Brazil , I know you forgot to finish some stadiums , but did you also forget to train \$ your team ? # WorldCup # BrazilvsMexico

@ themoonexpress Im going to take a stab in the dark and say ... youre not a fan of the South American dance ? # sarcasm # WorldCup

I need a goal line technology replay of that # Robben goal , Im not sure if it went in # World-Cup

not an offside for Higuain ! ☹ damn referee spoilt it ! # worldcup # ARG

I think \$ its to early 4 bad belleism ... # Worldcup

only reason I havent killed myself after that boring game is cos # USA will be playing in an hour . # WorldCup

spare a thought for Queen Maxima of the Nerherlands , being Argentine she may not be Queen after the semi final ! : Wooooo Orange ! Worldcup

after a 5-1 loss to the Dutch , this must be how Spain feels [http : / / t.co / urc1aV7uJ9](http://t.co/urc1aV7uJ9) # worldcup

and history repeats itself with the Swiss , they decide to not put up a fight ... zing # World-Cup # SUIvsFRA

its so strange to see # MEX wear red and black as \$ their alternate colors . # WorldCup

France miss a pen still 2-0 to France # WorldCup

we think \$ our # Belgium boss is a little confused on who to root for in the # BELvsUSA game . # WorldCup [http : / / t.co / zm0sVFGJXb](http://t.co/zm0sVFGJXb)

Siri has some sass . ☹ . # pummeled # brazil # germany # worldcup [http : / / t.co / EGjMM6fF9u](http://t.co/EGjMM6fF9u)

first water break ever in the history of the World Cup ... Not sure how I feel about it . # WorldCup # USAvsPOR

WorldCup | Ghana suffers 2 – 1 Loss to USA : Ghana ' s Black Stars lost \$ their first match of the 2014 World Cup : ... [http : / / t.co / GLVO5G8wBE](http://t.co/GLVO5G8wBE)

sorry Brazil dont host a worldcup nxt time # WorldCup # SSFootball

I never would have thought that Spain would look so unconvincing in \$ their play ever # WorldCup # ESP

Spain vs netherlands 1-5 # wtf # omg # WorldCup lol

can I introduced you please to a lump of Cheddar Cheese ! cmon England ! # WorldCup

OK bit of a lapse in concentration : - lets see how hungry the team come out : - its all about how bad u want it # WorldCup

heart attack is the only option # WorldCup

useless ! England useless ! # eng # worldcup

unlucky Brazil , it was Klose

why has the ref stopped play there ? its not a head injury . debatable whether its an injury at all ! # WorldCup

Spanish goal denied ; offside # worldcup # sucks # ESPNED

if you score only two goals at this # WorldCup you probably should expect to be heading home . # ITAURU

wait What ? not if , he must retire # Gerrard # WorldCup

damn # worldcup there and # worldcup there. tired already

for the first time I have fallen for the going toilet and missing the goal ! # worldcup # England

USA Man of the Match ~ NO ONE ! # scmotm # WorldCupProblems # worldcup # espn

Messis special message to injured Neymar [http : / / t.co / QlAKzDb5e](http://t.co/QlAKzDb5e)

lol , I just watched this match , , just to troll the team that will lose .I never was so much excited about it , tbh . ▪ # WorldCup

WK2014 : van Persie : Dutch must expect a hard time [http : / / t.co / gREqhKADkY](http://t.co/gREqhKADkY)

little Italys havin a lil shaker rn # avicii # WorldCup

Howard Webb is a terrible ref # WorldCup

may I be the first to say massive burns to Spain ! # Spainvschile # WorldCup

people saying its pointless watching @ england tomorrow . yes there out of the # WorldCup but Ill still be waiting it

@ MiseryBear Poor little bear . perhaps you and @ officialkylie could help each other commiserate # WorldCup misery

are Brazil England in disguise ? # WorldCup

Facebook Translate – Αξιολόγηση ακρίβειας και ευχέρειας

ST	MT Output	Adequacy	Fluency
es ist total deprimierend , zu sehen , dass \$ mein # Mexico den # Worldcup verlassen hat , aber niemand kann leugnen , sie haben gekämpft wie die Löwen ! ! ! ! ! # VamosMexico	it's totally depressing to see that \$ my #Mexico has left the #Worldcup, but nobody can deny, they fought like the lions! ! #VamosMexico	Good	Excellent
verdammte Scheiße # Brazil , ich weiß , ihr habt vergessen , einige Stadien fertig zubauen , aber habt ihr auch vergessen \$ euer Team zu trainieren ? # Worldcup # BrazilvsMexico	Damn shit #Brazil I know y'all forgot to finish some stadiums but did y'all forget to train \$ your team too? #Worldcup #BrazilvsMexico	Good	Excellent
@ themoonexpress Lass mich raten ... du bist kein Fan des südamerikanischen Tanzes ? # sarcasm # Worldcup	@ themoon express Let me guess... you're not a fan of South American dance? #sarcasm #worldcup	Excellent	Fair
ich brauche eine Wiederholung mit Torlinientechnik für dieses # robben Tor , Ich bin nicht sicher , ob er drin war # Worldcup	i need a repeat with goal line tech for this #robben goal, not sure if he was in it #Worldcup	Poor	Excellent
kein Abseits von Higuain ! ☹ der verdammte Schiedsrichter hat es verdorben ! # Worldcup # ARG	no offside Higuain! ☹ the damn ref spoiled it! #Worldcup #ARG	Excellent	Excellent
ich denke \$ es ist zu früh für schlechten Ballismus ... # Worldcup	i think \$ it's too early for bad ballism... #Worldcup	Excellent	Excellent
der einzige Grund , warum ich mich nach diesem langweiligen Spiel nicht umgebracht habe , ist das # USA in einer Stunde spielt . # Worldcup	the only reason I didn't kill myself after this boring game is the #USA plays in an hour. #Worldcup	Good	Fair
bringt einen Gedanken für die Königin Maxima der Niederlande auf , da ich argentinisch bin , sage ich , es kann sein , dass sie nach dem Halbfinale keine Königin ist ! : wooooo Orange ! Weltmeisterschaft	brings up a thought for Queen Maxima of Netherlands, since I'm Argentinian, I say she may not be a queen after the semi finals! : wooooo orange! World Cup	Poor	Good
nach einer 5-1 Niederlage gegen die Holländer , muss sich Spanien so fühlen http : // t.co / urc1aV7uJ9 # Worldcup	After a 5-1 defeat to the Dutch, Spain must feel this way http: // t. co / urc1aV 7uJ 9 #Worldcup	Excellent	Excellent

ST	MT Output	Adequacy	Fluency
und die Geschichte wiederholt sich mit den Schweizern , sie entscheiden sich , sich nicht zu wehren .. Zing # Worldcup # SUIvsFRA	and history repeats itself with the Swiss, they choose not to fight back.. Zing #Worldcup #SUIvsFRA	Excellent	Excellent
es ist so seltsam zu sehen , dass # MEX rot und schwarz als \$ ihre alternativen Farben tragen . # Worldcup	it's so weird to see #MEX wearing red and black as \$ their alternative colors. #Worldcup	Excellent	Excellent
Frankreich verschießt einen Elfmeter , trotzdem noch 2-0 für Frankreich # Worldcup	France shoots a penalty, but still 2-0 for France #Worldcup	Very poor	Excellent
wir glauben , \$ unser # Belgium Chef ist ein wenig verwirrt darüber , bei wem er mitfeiern soll in dem # BELvsUSA Spiel . # Worldcup http : / / t.co / zm0sVFGJXb	we think \$ our #Belgium boss is a little confused about who to cheer on in the #BelvsUSA game. #Worldcup http: / / t. co / zm0sVFGJXb	Excellent	Excellent
# Siri besitzt einige Frechheit . ☹ . # pummeled # brazil # germany # Worldcup http : / / t.co / EGjMM6fF9u	#Siri has some sassiness. ☹. ☹ #chummed #brazil #germany #Worldcup http: / / t. co / EGjM 6fF 9u	Excellent	Excellent
erste Wasserpause in der Geschichte der Weltmeisterschaft ... Nicht sicher , wie ich darüber denke . # Worldcup # USAvsPOR	first water break in the history of the World Cup... Not sure how I feel about it. #Worldcup #USAvsPOR	Good	Excellent
# Worldcup | Ghana leidet unter ihrer 2 - 1 Niederlage gegen die USA : Ghanas schwarze Stars verloren \$ ihr erstes Spiel der Weltmeisterschaft 2014 : ... http : / / t.co / GLVO5G8wBE	#Worldcup &# 124; Ghana suffers from her 2-1 defeat to the USA: Ghana's black stars lost \$ her first game of the 2014 World Cup :... http: / / t. co / GLVO5G 8wBE	Fair	Fair
tut mir leid Brasilien veranstalte nächstes Mal keine Weltmeisterschaft # Worldcup Host # SSFootball	sorry Brazil.... not hosting a world cup next time #Worldcuphost #SSFootball	Poor	Good
ich hätte nie gedacht , dass Spanien jemals in einem Spiel so unüberzeugend \$ aussehen würden # Worldcup # ESP	i never thought Spain would ever look so unsurpassing \$ in a game #Worldcup #ESP	Very poor	Excellent
Spanien vs Niederlande 1-5 # WTF # omg # Worldcup lol	Spain vs Netherlands 1-5 #WTF #omg #Worldcup lol	Excellent	Excellent
kann ich dich bitte mit einem Klumpen Cheddar-Käse bekanntmachen ! los England ! # Worldcup	can i please get you acquainted with a lump of cheddar cheese! Let's go England! #Worldcup	Good	Excellent

ST	MT Output	Adequacy	Fluency
OK eine kleine Unaufmerksamkeit : - mal sehen , wie tor-hungrig das Team sein wird : - es geht nur darum , wie sehr man es will # Worldcup	OK a little inattention: - let's see how goal-hungry the team will be: - it's all about how bad you want it #Worldcup	Excellent	Excellent
Herzinfarkt ist die einzige Option # Worldcup	Heart attack is the only option #Worldcup	Excellent	Excellent
unnütz ! England nutzlos ! # eng # Worldcup	useless! England useless! #tight #worldcup	Excellent	Excellent
unglückliches Brasilien , es war Klose	unlucky Brazil it was Klose	Excellent	Good
warum hat der Schiedsrichter aufgehört dort zuspiesen ? es ist keine Kopfverletzung . es lässt sich darüber streiten , ob das überhaupt eine Verletzung war ! # Worldcup	why did the ref stop playing there? it's not a head injury. it can be argued about whether this was even an injury! #Worldcup	Good	Excellent
spanisches wurde das Tor nicht gegeben ; Abseits # Worldcup # sucks # ESPNED	The goal wasn't given in Spanish; offside #Worldcup #sucks #ESPNED	Fair	Good
wenn man in diesem # Worldcup nur zwei Tore macht , sollte man wahrscheinlich damit rechnen bald nach Hause zu fahren . # ITAURU	If you only score two goals in this #Worldcup, you should probably expect to drive home soon. #ITAURU	Good	Excellent
warte was ? nicht , wenn er in den Ruhestand gehen muss # Gerrard # Worldcup	Wait for what? not when he needs to retire #Gerrard #Worldcup	Poor	Excellent
verdammt # Worldcup hier und # Worldcup da. hab es jetzt schon satt	damn #worldcup here and #worldcup there. tired of it already	Excellent	Excellent
zum ersten Mal hat mich das zur Toilette gehen gekriegt und ich habe das Tor verpasst ! # Worldcup # England	for the first time this got me going to the bathroom and I missed the gate! #Worldcup #England	Very poor	Excellent
USA Man of the Match ~ NO ONE ! # scmotm # WorldCupProblems # Worldcup # ESPN	USA Man of the Match ~ NO ONE ! # scmotm # WorldCupProblems # Worldcup # ESPN	Excellent	Excellent
Messis Eilmeldung an den verletzten Neymar http : // t.co / QlAKzDb5e	Messi's breaking message to injured Neymar http: // t. co / QlAKzDb5e	Poor	Excellent

ST	MT Output	Adequacy	Fluency
lol , ich habe dieses Spiel nur geguckt , um mich über das Team , das verliert , lustig zu machen.Ich war darüber noch nie so begeistert , TBH . ☹️ # Worldcup	lol I just watched this game to make fun of the team losing. Never been so excited about this TBH. ☹️ #Worldcup	Poor	Good
# WK2014 : van Persie : die Niederländer müssen eine harte Zeit erwarten http : // t.co / gREqhKADKY	#WK2014: van Persie: Dutch must expect a hard time http: // t. co / gREqhKADKY	Excellent	Good
süßes Italien hat gerade einen kleines Zittern # avicii # Worldcup	cute Italy just got a little tremor #avicii #worldcup	Excellent	Excellent
Howard Webb ist ein schrecklicher Schiri # Worldcup	Howard Webb is a terrible ref #Worldcup	Excellent	Excellent
darf ich der erste sein der sagt , Totalschaden für Spanien ! # Spainvschile # Worldcup	May I be the first to say total damage to Spain! #Spainvschile #Worldcup	Fair	Good
die Leute sagen , es ist sinnlos das @ england Spiel morgen zu gucken . ja , sie sind raus aus dem # Worldcup aber ich warte immer noch darauf	people say it's pointless to watch the @ england game tomorrow. yes they are out of the #worldcup but I'm still waiting for it	Excellent	Excellent
@ MiseryBear Armer kleine Bär . vielleicht könnten du und @ officialkylie einander helfen , das Elend zu bedauern # Worldcup	@ MiseryBear poor little bear. maybe you and @ official kylie could help each other regret the misery #Worldcup	Fair	Excellent
sind Brasilien und England verkleidet ? # Worldcup	are brazil and england disguised? #Worldcup	Excellent	Excellent

Google Translate – Αξιολόγηση ακρίβειας και ευχέρειας

ST	MT Output	Adequacy	Fluency
es ist total deprimierend , zu sehen , dass \$ mein # Mexico den # Worldcup verlassen hat , aber niemand kann leugnen , sie haben gekämpft wie die Löwen ! ! ! ! ! # VamosMexico	it's totally depressing to see that \$ my # Mexico has left the # World Cup, but no one can deny that they fought like the lions! ! ! ! ! # VamosMexico	Good	Excellent
verdammte Scheiße # Brazil , ich weiß , ihr habt vergessen , einige Stadien fertig zubauen , aber habt ihr auch vergessen \$ euer Team zu trainieren ? # Worldcup # BrazilvsMexico	Damn shit # Brazil, I know you forgot to finish building some stadiums, but did you forget \$ to train your team too? # Worldcup # BrazilvsMexico	Excellent	Excellent
@ themoonexpress Lass mich raten ... du bist kein Fan des südamerikanischen Tanzes ? # sarcasm # Worldcup	@ themoonexpress Let me guess ... you are not a fan of the South American dance? # sarcasm # Worldcup	Excellent	Fair
ich brauche eine Wiederholung mit Torlinientechnik für dieses # robben Tor , Ich bin nicht sicher , ob er drin war # Worldcup	I need a goal-line repetition for this # robben goal, I'm not sure if it was # Worldcup	Poor	Excellent
kein Abseits von Higuain ! ☹ der verdammte Schiedsrichter hat es verdorben ! # Worldcup # ARG	no apart from Higuain! ☹ the damn referee spoiled it! # Worldcup # ARG	Very poor	Good
ich denke \$ es ist zu früh für schlechten Ballismus ... # Worldcup	I think \$ it's too early for bad ballism ... # Worldcup	Excellent	Excellent
der einzige Grund , warum ich mich nach diesem langweiligen Spiel nicht umgebracht habe , ist das # USA in einer Stunde spielt . # Worldcup	the only reason I didn't kill myself after this boring game is because #USA is playing in an hour. # Worldcup	Excellent	Excellent
bringt einen Gedanken für die Königin Maxima der Niederlande auf , da ich argentinisch bin , sage ich , es kann sein , dass sie nach dem Halbfinale keine Königin ist ! : wooooo Orange ! Weltmeisterschaft	brings up a thought for Queen Maxima of the Netherlands, since I'm Argentine, I say, it may be that she won't be a queen after the semifinals! : wooooo orange! World Championship	Poor	Good
nach einer 5-1 Niederlage gegen die Holländer , muss sich Spanien so fühlen http : // t.co / urc1aV7uJ9 # Worldcup	after a 5-1 loss to the Dutch, Spain must feel this way http: // t.co / urc1aV7uJ9 # Worldcup	Excellent	Excellent

ST	MT Output	Adequacy	Fluency
und die Geschichte wiederholt sich mit den Schweizern , sie entscheiden sich , sich nicht zu wehren .. Zing # Worldcup # SUIvsFRA	and history repeats itself with the Swiss, they decide not to fight back .. Zing # Worldcup # SUIvsFRA	Excellent	Excellent
es ist so seltsam zu sehen , dass # MEX rot und schwarz als \$ ihre alternativen Farben tragen . # Worldcup	it's so strange to see # MEX wearing red and black as \$ their alternate colors. # Worldcup	Excellent	Excellent
Frankreich verschießt einen Elfmeter , trotzdem noch 2-0 für Frankreich # Worldcup	France misses a penalty, but still 2-0 for France #Worldcup	Excellent	Excellent
wir glauben , \$ unser # Belgium Chef ist ein wenig verwirrt darüber , bei wem er mitfiebern soll in dem # BELvsUSA Spiel . # Worldcup http : / / t.co / zm0sVFGJXb	we think \$ our # Belgium boss is a little confused about who to cheer for in the # BELvsUSA game. # Worldcup http: / / t.co / zm0sVFGJXb	Good	Excellent
# Siri besitzt einige Frechheit . ☹ . # pummeled # brazil # germany # Worldcup http : / / t.co / EGjMM6fF9u	# Siri has some cheek. ☹. # pummeled # brazil # germany # Worldcup http: / / t.co / EGjMM6fF9u	Excellent	Excellent
erste Wasserpause in der Geschichte der Weltmeisterschaft ... Nicht sicher , wie ich darüber denke . # Worldcup # USAvsPOR	first water break in World Cup history ... not sure how I think about that. # Worldcup # USAvsPOR	Excellent	Excellent
# Worldcup | Ghana leidet unter ihrer 2 - 1 Niederlage gegen die USA : Ghanas schwarze Stars verloren \$ ihr erstes Spiel der Weltmeisterschaft 2014 : ... http : / / t.co / GLVO5G8wBE	# Worldcup & # 124; Ghana suffers from their 2-1 loss to the USA: Ghana's black stars lost \$ their first game of the 2014 World Cup: ... http: / / t.co / GLVO5G8wBE	Fair	Excellent
tut mir leid Brasilien veranstalte nächstes Mal keine Weltmeisterschaft # Worldcup Host # SSFootball	I'm sorry Brazil don't host a World Cup # Worldcup Host # SSFootball next time	Excellent	Excellent
ich hätte nie gedacht , dass Spanien jemals in einem Spiel so unüberzeugend \$ aussehen würden # Worldcup # ESP	I never thought that Spain would ever look so unconvincing in a game \$ # Worldcup # ESP	Excellent	Excellent
Spanien vs Niederlande 1-5 # WTF # omg # Worldcup lol	Spain vs Netherlands 1-5 # WTF # omg # Worldcup lol	Excellent	Excellent
kann ich dich bitte mit einem Klumpen Cheddar-Käse bekanntmachen ! los England ! # Worldcup	can I introduce you to a lump of cheddar cheese, please! go England! # Worldcup	Excellent	Excellent

ST	MT Output	Adequacy	Fluency
OK eine kleine Unaufmerksamkeit : - mal sehen , wie tor-hungrig das Team sein wird : - es geht nur darum , wie sehr man es will # Worldcup	OK a little inattention: - let's see how goal-hungry the team will be: - it's all about how much you want it # Worldcup	Excellent	Excellent
Herzinfarkt ist die einzige Option # Worldcup	Heart attack is the only option # Worldcup	Excellent	Excellent
unnütz ! England nutzlos ! # eng # Worldcup	useless! England useless! # eng # Worldcup	Excellent	Excellent
unglückliches Brasilien , es war Klose	unhappy Brazil, it was Klose	Excellent	Excellent
warum hat der Schiedsrichter aufgehört dort zuspiesen ? es ist keine Kopfverletzung . es lässt sich darüber streiten , ob das überhaupt eine Verletzung war ! # Worldcup	why did the referee stop playing there? it's not a head injury. it is debatable whether that was an injury at all! # Worldcup	Excellent	Excellent
spanisches wurde das Tor nicht gegeben ; Abseits # Worldcup # sucks # ESPNED	Spanish was not given the gate; Offside # Worldcup # sucks # ESPNED	Very poor	Good
wenn man in diesem # Worldcup nur zwei Tore macht , sollte man wahrscheinlich damit rechnen bald nach Hause zu fahren . # ITAURU	if you only score two goals in this #Worldcup, you should probably expect to go home soon. # ITAURU	Excellent	Excellent
warte was ? nicht , wenn er in den Ruhestand gehen muss # Gerrard # Worldcup	wait what not if he has to retire # Gerrard # Worldcup	Poor	Poor
verdammt # Worldcup hier und # Worldcup da. hab es jetzt schon satt	damn #worldcup here and #worldcup there. already tired of it	Excellent	Excellent
zum ersten Mal hat mich das zur Toilette gehen gekriegt und ich habe das Tor verpasst ! # Worldcup # England	for the first time this got me to the bathroom and I missed the gate! # Worldcup # England	Very poor	Excellent
USA Man of the Match ~ NO ONE ! # scmotm # WorldCupProblems # Worldcup # ESPN	USA Man of the Match ~ NO ONE! # scmotm # WorldCupProblems # Worldcup # ESPN	Excellent	Excellent
Messis Eilnachricht an den verletzten Neymar http : // t.co / QlAKzDb5e	Messi's urgent message to the injured Neymar http: // t.co / QlAKzDb5e	Poor	Excellent

ST	MT Output	Adequacy	Fluency
lol , ich habe dieses Spiel nur geguckt , um mich über das Team , das verliert , lustig zu machen.Ich war darüber noch nie so begeistert , TBH . ☹️ # Worldcup	lol, I just watched this game to make fun of the team that was losing, I've never been so excited about it, TBH. ☹️ # Worldcup	Poor	Excellent
# WK2014 : van Persie : die Niederländer müssen eine harte Zeit erwarten http : // t.co / gREqhKADkY	# WK2014: van Persie: the Dutch have to expect a hard time http: / / t.co / gREqhKADkY	Excellent	Excellent
süßes Italien hat gerade einen kleines Zittern # avicii # Worldcup	sweet italy just got a little tremor # avicii # worldcup	Excellent	Excellent
Howard Webb ist ein schrecklicher Schiri # Worldcup	Howard Webb is a terrible referee #Worldcup	Good	Excellent
darf ich der erste sein der sagt , Totalschaden für Spanien ! # Spainvschile # Worldcup	may I be the first to say, total loss for Spain! # Spainvschile # Worldcup	Excellent	Excellent
die Leute sagen , es ist sinnlos das @ england Spiel morgen zu gucken . ja , sie sind raus aus dem # Worldcup aber ich warte immer noch darauf	people say there is no point in watching @ england tomorrow. yes, they are out of the #worldcup but i'm still waiting for it	Good	Excellent
@ MiseryBear Armer kleine Bär . vielleicht könnten du und @ officialkylie einander helfen , das Elend zu bedauern # Worldcup	@ MiseryBear Poor little bear. maybe you and @officialkylie could help each other regret the misery #Worldcup	Fair	Excellent
sind Brasilien und England verkleidet ? # Worldcup	are Brazil and England in disguise? # Worldcup	Excellent	Excellent

Facebook Translate – Κατηγοριοποίηση Λαθών

ST	MT output	No error	Accuracy	Addition	Omission	Mistranslation	Terminology	Fluency	Punctuation	Grammar	Style	Awkward
es ist total dep- rimierend , zu seh- en , dass \$ mein # Mexico den # Worldcup verlas- sen hat , aber nie- mand kann leugnen , sie haben gekämpft wie die Löwen ! ! ! ! ! # VamosMexico	it's totally depress- ing to see that \$ my #Mexico has left the #Worldcup, but nobody can deny, they fought like the lions! ! #VamosMexico		1	1								
verdammte Scheiße # Brazil , ich weiß , ihr habt vergessen , einige Stadien fertig zub- auen , aber habt ihr auch vergessen \$ euer Team zu trainieren ? # Worldcup # BrazilvsMexico	Damn shit #Brazil I know y'all forgot to finish some stadi- ums but did y'all forget to train \$ your team too? #Worldcup #BrazilvsMexico										2	2
@ themoonexpress Lass mich raten ... du bist kein Fan des südamerikan- ischen Tanzes ? # sarcasm # World- cup	@ themoon ex- press Let me guess... you're not a fan of South American dance? #sarcasm #worldcup							1		1		
ich brauche eine Wiederholung mit Torlinientechnik für dieses # robben	i need a repeat with goal line tech for this #robben goal, not sure if he		3	1		1	1					

ST	MT output	No error	Accuracy	Addition	Omission	Mistranslation	Terminology	Fluency	Punctuation	Grammar	Style	Awkward
Tor , Ich bin nicht sicher , ob er drin war # Worldcup	was in it #Worldcup											
kein Abseits von Higuain ! ☹ der verdammte Schiedsrichter hat es verdorben ! # Worldcup # ARG	no offside Higuain! ☹ the damn ref spoiled it! #Worldcup #ARG	1										
ich denke \$ es ist zu früh für schlechten Ballismus ... # Worldcup	i think \$ it's too early for bad ballism... #Worldcup	1										
der einzige Grund , warum ich mich nach diesem langweiligen Spiel nicht umgebracht habe , ist das # USA in einer Stunde spielt . # Worldcup	the only reason I didn't kill myself after this boring game is the #USA plays in an hour. #Worldcup							1		1		
bringt einen Gedanken für die Königin Maxima der Niederlande auf , da ich argentinisch bin , sage ich , es kann sein , dass sie nach dem Halbfinale keine Königin ist ! : wooooo Orange ! Weltmeisterschaft	brings up a thought for Queen Maxima of Netherlands, since I'm Argentinian, I say she may not be a queen after the semi finals! : wooooo orange! World Cup							1		1		

ST	MT output	No error	Accuracy	Addition	Omission	Mistranslation	Terminology	Fluency	Punctuation	Grammar	Style	Awkward
nach einer 5-1 Niederlage gegen die Holländer , muss sich Spanien so fühlen http : // t.co / urc1aV7uJ9 # Worldcup	After a 5-1 defeat to the Dutch, Spain must feel this way http: // t. co / urc1aV 7uJ 9 #Worldcup	1										
und die Geschichte wiederholt sich mit den Schweizern , sie entscheiden sich , sich nicht zu wehren .. Zing # Worldcup # SUIvsFRA	and history repeats itself with the Swiss, they choose not to fight back.. Zing #Worldcup #SUIvsFRA	1										
es ist so seltsam zu sehen , dass # MEX rot und schwarz als \$ ihre alternativen Farben tragen . # Worldcup	it's so weird to see #MEX wearing red and black as \$ their alternative colors. #Worldcup	1										
Frankreich verschießt einen Elfmeter , trotzdem noch 2-0 für Frankreich # Worldcup	France shoots a penalty, but still 2-0 for France #Worldcup		1			1						
wir glauben , \$ unser # Belgium Chef ist ein wenig verwirrt darüber , bei wem er mitfiebern soll in dem # BELvsUSA Spiel . # Worldcup http : /	we think \$ our #Belgium boss is a little confused about who to cheer on in the #BelvsUSA game. #Worldcup http: // t. co / zm0svFGJXb	1										

ST	MT output	No error	Accuracy	Addition	Omission	Mistranslation	Terminology	Fluency	Punctuation	Grammar	Style	Awkward
/ t.co / zm0sVFGJXb												
# Siri besitzt einige Frechheit . ☹ . # pummeled # brazil # germany # Worldcup http : // t.co / EGjMM6fF9u	#Siri has some sassiness. ☹. ☹ #chummed #brazil #germany #World- cup http: // t. co / EGjM 6fF 9u	1										
erste Wasserpause in der Geschichte der Weltmeister- schaft ... Nicht sicher , wie ich darüber denke . # Worldcup # USAvsPOR	first water break in the history of the World Cup... Not sure how I feel about it. #Worldcup #USAvsPOR		1			1						
# Worldcup | Ghana leidet unter ihrer 2 - 1 Nieder- lage gegen die USA : Ghanas schwarze Stars verloren \$ ihr erstes Spiel der Weltmeisterschaft 2014 : ... http : // t.co / GLVO5G8wBE	#Worldcup &# 124; Ghana suffers from her 2-1 defeat to the USA: Ghana's black stars lost \$ her first game of the 2014 World Cup :... http: // t. co / GLVO5G 8wBE							2		2		
tut mir leid Brasili- en veranstalte nächstes Mal keine Weltmeisterschaft # Worldcup Host #	sorry Brazil.... not hosting a world cup next time #Worldcuphost #SSFootball		1			1						

ST	MT output	No error	Accuracy	Addition	Omission	Mistranslation	Terminology	Fluency	Punctuation	Grammar	Style	Awkward
SSFootball												
ich hätte nie gedacht , dass Spanien jemals in einem Spiel so unüberzeugend \$ aussehen würden # Worldcup # ESP	i never thought Spain would ever look so unsurpassing \$ in a game #Worldcup #ESP		1			1						
Spanien vs Niederlande 1-5 # WTF # omg # Worldcup lol	Spain vs Netherlands 1-5 #WTF #omg #Worldcup lol	1										
kann ich dich bitte mit einem Klumpen Cheddar-Käse bekanntmachen ! los England ! # Worldcup	can i please get you acquainted with a lump of cheddar cheese! Let's go England! #Worldcup										1	1
OK eine kleine Unaufmerksamkeit : - mal sehen , wie tor-hungrig das Team sein wird : - es geht nur darum , wie sehr man es will # Worldcup	OK a little inattention: - let's see how goal-hungry the team will be: - it's all about how bad you want it #Worldcup	1										
Herzinfarkt ist die einzige Option # Worldcup	Heart attack is the only option #Worldcup	1										
unnütz ! England nutzlos ! # eng # Worldcup	useless! England useless! #tight #worldcup	1										

ST	MT output	No error	Accuracy	Addition	Omission	Mistranslation	Terminology	Fluency	Punctuation	Grammar	Style	Awkward
unglückliches Brasilien , es war Klose	unlucky Brazil it was Klose							1	1			
warum hat der Schiedsrichter aufgehört dort zuspiesen ? es ist keine Kopfverletzung . es lässt sich darüber streiten , ob das überhaupt eine Verletzung war ! # Worldcup	why did the ref stop playing there? it's not a head injury. it can be argued about whether this was even an injury! #Worldcup										1	1
spanisches wurde das Tor nicht gegeben ; Abseits # Worldcup # sucks # ESPNED	The goal wasn't given in Spanish; offside #Worldcup #sucks #ESPNED							1		1		
wenn man in diesem # Worldcup nur zwei Tore macht , sollte man wahrscheinlich damit rechnen bald nach Hause zu fahren . # ITAURU	If you only score two goals in this #Worldcup, you should probably expect to drive home soon. #ITAURU										1	1
warte was ? nicht , wenn er in den Ruhestand gehen muss # Gerrard # Worldcup	Wait for what? not when he needs to retire #Gerrard #Worldcup	1										
verdammt # Worldcup hier und # Worldcup da.	damn #worldcup here and #worldcup there. tired of	1										

ST	MT output	No error	Accuracy	Addition	Omission	Mistranslation	Terminology	Fluency	Punctuation	Grammar	Style	Awkward
hab es jetzt schon satt	it already											
zum ersten Mal hat mich das zur Toilette gehen gekriegt und ich habe das Tor verpasst ! # Worldcup # England	for the first time this got me going to the bathroom and I missed the gate ! #Worldcup #England		2			1	1					
USA Man of the Match ~ NO ONE ! # scmotm # WorldCupProblems # Worldcup # ESPN	USA Man of the Match ~ NO ONE ! # scmotm # WorldCupProblems # Worldcup # ESPN	1										
Messis Eilmeldung an den verletzten Neymar http : // t.co / QILAKzDb5e	Messi's breaking message to injured Neymar http: // t.co / QILAKzDb5e		1			1						
lol , ich habe dieses Spiel nur geguckt , um mich über das Team , das verliert , lustig zu machen.Ich war darüber noch nie so begeistert , TBH . ☹ # Worldcup	lol I just watched this game to make fun of the team losing . Never been so excited about this TBH. ☹ #Worldcup		1		1			1		1		
# WK2014 : van Persie : die Niederländer müssen eine harte Zeit erwarten http : // t.co / gREqhKADkY	#WK2014: van Persie: Dutch must expect a hard time http: // t.co / gREqhKADkY		1		1							

ST	MT output	No error	Accuracy	Addition	Omission	Mistranslation	Terminology	Fluency	Punctuation	Grammar	Style	Awkward
süßes Italien hat gerade einen kleinen Zittern #avicii # Worldcup	cute Italy just got a little tremor #avicii #worldcup	1										
Howard Webb ist ein schrecklicher Schiri # Worldcup	Howard Webb is a terrible ref #Worldcup	1										
darf ich der erste sein der sagt , Totalschaden für Spanien ! #Spainvschile # Worldcup	May I be the first to say total damage to Spain! #Spainvschile #Worldcup										1	1
die Leute sagen , es ist sinnlos das @england Spiel morgen zu gucken . ja , sie sind raus aus dem # Worldcup aber ich warte immer noch darauf	people say it's pointless to watch the @england game tomorrow. yes they are out of the #worldcup but I'm still waiting for it	1										
@ MiseryBear Armer kleine Bär . vielleicht könnten du und @ official-kylie einander helfen , das Elend zu bedauern # Worldcup	@ MiseryBear poor little bear. maybe you and @ official kylie could help each other regret the misery #Worldcup		1			1						
sind Brasilien und England verkleidet ? # Worldcup	are brazil and england disguised? #Worldcup	1										

Google Translate – Κατηγοριοποίηση Λαθών

ST	MT output	No error	Accuracy	Addition	Omission	Mistranslation	Terminology	Fluency	Punctuation	Grammar	Style	Inconsistent Style	Verity	Culture-specific element
es ist total deprimierend , zu sehen , dass \$ mein # Mexico den # Worldcup verlassen hat , aber niemand kann leugnen , sie haben gekämpft wie die Löwen ! ! ! ! ! # VamosMexico	it's totally depressing to see that \$ my # Mexico has left the # World Cup, but no one can deny that they fought like the lions! ! ! ! ! # VamosMexico		1	1										
verdammte Scheiße # Brazil , ich weiß , ihr habt vergessen , einige Stadien fertig zubauen , aber habt ihr auch vergessen \$ euer Team zu trainieren ? # Worldcup # BrazilvsMexico	Damn shit # Brazil, I know you forgot to finish building some stadiums, but did you forget \$ to train your team too? # Worldcup # BrazilvsMexico												1	1
@ themoonexpress Lass mich raten ... du bist kein Fan des südamerikanischen Tanzes ? # sarcasm # Worldcup	@ themoonexpress Let me guess ... you are not a fan of the South American dance? # sarcasm # Worldcup							1		1				
ich brauche eine Wiederholung mit Torlinientechnik für dieses # robben Tor , Ich bin nicht sicher , ob er drin war # Worldcup	I need a goal-line repetition for this # robben goal, I'm not sure if it was # Worldcup		2		1		1							

ST	MT output	No error	Accuracy	Addition	Omission	Mistranslation	Terminology	Fluency	Punctuation	Grammar	Style	Inconsistent Style	Verity	Culture-specific element
kein Abseits von Higuain ! ☹ der verdammte Schiedsrichter hat es verdorben ! # Worldcup # ARG	no apart from Higuain! ☹ the damn referee spoiled it! # Worldcup # ARG		1				1							
ich denke \$ es ist zu früh für schlechten Ballismus ... # Worldcup	I think \$ it's too early for bad ballism ... # Worldcup	1												
der einzige Grund , warum ich mich nach diesem langweiligen Spiel nicht umgebracht habe , ist das # USA in einer Stunde spielt . # Worldcup	the only reason I didn't kill myself after this boring game is because #USA is playing in an hour. # Worldcup	1												
bringt einen Gedanken für die Königin Maxima der Niederlande auf , da ich argentinisch bin , sage ich , es kann sein , dass sie nach dem Halbfinale keine Königin ist ! : wooooo Orange ! Weltmeisterschaft	brings up a thought for Queen Maxima of the Netherlands, since I'm Argentine, I say, it may be that she won't be a queen after the semifinals! : wooooo orange! World Championship							1		1				
nach einer 5-1 Niederlage gegen die Holländer , muss sich Spanien so fühlen http : // t.co / urc1aV7uJ9 # World-	after a 5-1 loss to the Dutch, Spain must feel this way http: // t.co / urc1aV7uJ9 # Worldcup	1												

ST	MT output	No error	Accuracy	Addition	Omission	Mistranslation	Terminology	Fluency	Punctuation	Grammar	Style	Inconsistent Style	Verity	Culture-specific element
cup														
und die Geschichte wiederholt sich mit den Schweizern , sie entscheiden sich , sich nicht zu wehren .. Zing # Worldcup # SUIvsFRA	and history repeats itself with the Swiss, they decide not to fight back .. Zing # Worldcup # SUIvsFRA	1												
es ist so seltsam zu sehen , dass # MEX rot und schwarz als \$ ihre alternativen Farben tragen . # Worldcup	it's so strange to see # MEX wearing red and black as \$ their alternate colors. # Worldcup	1												
Frankreich verschießt einen Elfmeter , trotzdem noch 2-0 für Frankreich # Worldcup	France misses a penalty, but still 2-0 for France #Worldcup	1												
wir glauben , \$ unser # Belgium Chef ist ein wenig verwirrt darüber , bei wem er mitfiebern soll in dem # BELvsUSA Spiel . # Worldcup http : // t.co / zm0sVFGJXb	we think \$ our # Belgium boss is a little confused about who to cheer for in the # BELvsUSA game. # Worldcup http: // t.co / zm0sVFGJXb		1			1								
# Siri besitzt einige Frechheit . ☹ . # pummeled # brazil # germany # Worldcup	# Siri has some cheek. ☹. # pummeled # brazil # germany # World-	1												

ST	MT output	No error	Accuracy	Addition	Omission	Mistranslation	Terminology	Fluency	Punctuation	Grammar	Style	Inconsistent Style	Verity	Culture-specific element
http : // t.co / EGjMM6fF9u	cup http: // t.co / EGjMM6fF9u													
erste Wasserpause in der Geschichte der Weltmeisterschaft ... Nicht sicher , wie ich darüber denke . # Worldcup # USAvsPOR	first water break in World Cup history ... not sure how I think about that. # Worldcup # USAvsPOR	1												
# Worldcup & #124; Ghana leidet unter ihrer 2 - 1 Niederlage gegen die USA : Ghanas schwarze Stars verloren \$ ihr erstes Spiel der Weltmeisterschaft 2014 : ... http : // t.co / GLVO5G8wBE	# Worldcup & # 124; Ghana suffers from their 2-1 loss to the USA: Ghana's black stars lost \$ their first game of the 2014 World Cup: ... http: / / t.co / GLVO5G8wBE							1		1				
tut mir leid Brasilien veranstalte nächstes Mal keine Weltmeisterschaft # Worldcup Host # SSFootball	I'm sorry Brazil don't host a World Cup # Worldcup Host # SSFootball next time	1												
ich hätte nie gedacht , dass Spanien jemals in einem Spiel so unüberzeugend \$ aussehen würden # Worldcup # ESP	I never thought that Spain would ever look so unconvincing in a game \$ # Worldcup # ESP	1												

ST	MT output	No error	Accuracy	Addition	Omission	Mistranslation	Terminology	Fluency	Punctuation	Grammar	Style	Inconsistent Style	Verity	Culture-specific element
Spanien vs Niederlande 1-5 # WTF # omg # Worldcup lol	Spain vs Netherlands 1-5 # WTF # omg # Worldcup lol	1												
kann ich dich bitte mit einem Klumpen Cheddar-Käse bekanntmachen ! los England ! # Worldcup	can I introduce you to a lump of cheddar cheese, please! go England! # Worldcup	1												
OK eine kleine Unaufmerksamkeit : - mal sehen , wie torhungrig das Team sein wird : - es geht nur darum , wie sehr man es will # Worldcup	OK a little inattention: - let's see how goal-hungry the team will be: - it's all about how much you want it # Worldcup	1												
Herzinfarkt ist die einzige Option # Worldcup	Heart attack is the only option # Worldcup	1												
unnütz ! England nutzlos ! # eng # Worldcup	useless! England useless! # eng # Worldcup	1												
unglückliches Brasilien , es war Klose	unhappy Brazil, it was Klose	1												
warum hat der Schiedsrichter aufgehört dort zupspielen ? es ist keine Kopfverletzung . es lässt sich darüber streiten , ob das überhaupt eine Verletzung war ! # Worldcup	why did the referee stop playing there? it's not a head injury. it is debatable whether that was an injury at all! # Worldcup	1												

ST	MT output	No error	Accuracy	Addition	Omission	Mistranslation	Terminology	Fluency	Punctuation	Grammar	Style	Inconsistent Style	Verity	Culture-specific element
spanisches wurde das Tor nicht gegeben ; Abseits # Worldcup # sucks # ESPNED	Spanish was not given the gate ; Off-side # Worldcup # sucks # ESPNED		1				1							
wenn man in diesem # Worldcup nur zwei Tore macht , sollte man wahrscheinlich damit rechnen bald nach Hause zu fahren . # ITAURU	if you only score two goals in this #Worldcup, you should probably expect to go home soon. # ITAURU	1												
warte was ? nicht , wenn er in den Ruhestand gehen muss # Gerrard # Worldcup	wait what not if he has to retire # Gerrard # Worldcup							1	1	1				
verdammt # Worldcup hier und # Worldcup da. hab es jetzt schon satt	damn #worldcup here and #worldcup there. already tired of it	1												
zum ersten Mal hat mich das zur Toilette gehen gekriegt und ich habe das Tor verpasst ! # Worldcup # England	for the first time this got me to the bathroom and I missed the gate ! # Worldcup # England		2			1	1							
USA Man of the Match ~ NO ONE ! # scmotm # WorldCup-Problems # Worldcup # ESPN	USA Man of the Match ~ NO ONE! # scmotm # WorldCupProblems # Worldcup # ESPN	1												
Messis Eilmnachricht an den verletzten Neymar http : // t.co /	Messi's urgent message to the injured Neymar http: // t.co		1			1								

ST	MT output	No error	Accuracy	Addition	Omission	Mistranslation	Terminology	Fluency	Punctuation	Grammar	Style	Inconsistent Style	Verity	Culture-specific element
QLAKzDb5e	/ QLAKzDb5e													
lol , ich habe dieses Spiel nur geguckt , um mich über das Team , das verliert , lustig zu machen.Ich war darüber noch nie so begeistert , TBH . ☹ # Worldcup	lol, I just watched this game to make fun of the team that was losing, I've never been so excited about it, TBH. ☹ # Worldcup		1			1								
# WK2014 : van Persie : die Niederländer müssen eine harte Zeit erwarten http : / / t.co / gREqhKADkY	# WK2014: van Persie: the Dutch have to expect a hard time http: // t.co / gREqhKADkY	1												
süßes Italien hat gerade einen kleines Zittern # avicii # Worldcup	sweet italy just got a little tremor # avicii # worldcup	1												
Howard Webb ist ein schrecklicher Schiri # Worldcup	Howard Webb is a terrible referee #Worldcup										1	1		
darf ich der erste sein der sagt , Totalschaden für Spanien ! # Spainvschile # Worldcup	may I be the first to say, total loss for Spain! # Spainvschile # Worldcup	1												
die Leute sagen , es ist sinnlos das @ england Spiel morgen zu gucken . ja , sie sind raus aus dem # Worldcup aber ich	people say there is no point in watching @ england tomorrow. yes, they are out of the #worldcup but i'm still		1		1									

ST	MT output	No error	Accuracy	Addition	Omission	Mistranslation	Terminology	Fluency	Punctuation	Grammar	Style	Inconsistent Style	Verity	Culture-specific element
warte immer noch darauf	waiting for it													
@ MiseryBear Armer kleine Bär . vielleicht könntest du und @ officialkylie einander helfen , das Elend zu bedauern # Worldcup	@ MiseryBear Poor little bear. maybe you and @officialkylie could help each other regret the misery #Worldcup		1			1								
sind Brasilien und England verkleidet ? # Worldcup	are Brazil and England in disguise? # Worldcup	1												

Telework

Κείμενο Αφιερώσεως

Un Sistema di Formazione, Orientamento e Informazione sul Telelavoro

Il progetto

Il Progetto “TES – TELEWORK EDUCATION SYSTEM”, realizzato nel triennio 2000-2003, ha visto la partecipazione di un partenariato di cinque paesi: Italia, Grecia, Irlanda, Romania e Spagna e composto da Pubbliche Amministrazioni, Agenzie per il Lavoro, Istituti Scolastici Superiori, Università, Centri di Ricerca, Organizzazioni di Base degli operatori della formazione.

Il Progetto ha realizzato un innovativo modello di Sistema per la Formazione, l'Orientamento e l'Informazione al Telelavoro che sia in grado di garantire, attraverso una significativa azione formativa dei lavoratori, una corretta introduzione del telelavoro stesso nelle organizzazioni.

Ma perché questa formazione sia efficace ed efficiente, TES ha posto alla base della sua ricerca l'assunto che l'idea di “formazione al telelavoro” non è sufficiente e che sia necessario parlare di “educazione al telelavoro”, intendendo il momento educativo come l'avvio di un processo di acquisizione di quelle competenze di base necessarie sia ai lavoratori che a coloro che si accingono ad entrare nel mondo del lavoro, ma anche a tutti coloro che hanno la funzione di decisori e gestori nelle organizzazioni.

In tal senso il Progetto si pone obiettivi che possono avere ricadute molto forti sia sul mondo del lavoro che sul sistema formativo ed educativo.

Il Telelavoro (o “e-Work”) è definito, nella sua moderna accezione, come “una metodologia di lavoro che presuppone la fornitura a distanza e, se possibile, in modo asincrono, di una prestazione lavorativa attraverso l'uso, in tutto o in parte, delle tecnologie dell'Informazione e della Comunicazione”.

Il telelavoratore, in questo senso, può rappresentare un potente strumento di evoluzione delle organizzazioni nel processo di produzione e gestione del lavoro.

Una qualsiasi organizzazione (piccola o grande, pubblica o privata) ma anche singoli soggetti che operano nel contesto dell'auto-imprenditorialità, possono ottenere dal telelavoro significative innovazioni di prodotto e/o di processo.

Ma perché queste siano effettivamente tali, uno degli aspetti che le stesse organizzazioni devono curare è proprio quello della formazione e dell'aggiornamento dei lavoratori.

Saper telelavorare non è una competenza legata semplicemente a conoscenze di tipo tecnologico (saper usare il computer, Internet, la posta elettronica, ecc.), ma è una competenza organizzativa e culturale.

Il progetto vuole rispondere in modo innovativo a questa esigenza di formazione.

Per il raggiungimento dei suoi obiettivi TES ha realizzato un'articolata infrastruttura tecnologica di tipo distribuito che eroga i servizi formativi ed informativi.

Questa infrastruttura è gestita da un centro servizi - denominato TTC - TES Teleservices Centre (collocato presso l'ITI "F'GIORDANI") – ed utilizza uno specifico portale (TES Portal), accessibile all'indirizzo www.leonardotes.org , per erogare i propri servizi:

servizi formativi e di supporto per operatori della formazione , quali:

- corsi di formazione a distanza sul telelavoro,
- supporto all'ideazione, progettazione, realizzazione di nuovi moduli formativi e corsi sul telelavoro,
- uso di ambienti di sperimentazione del telelavoro (basati sulla simulazione dei processi)

servizi formativi per utenti finali, quali:

- corsi di formazione a distanza di introduzione al telelavoro e alla sua applicazione in vari contesti
- erogazione di materiali di supporto e di approfondimento

servizi informativi e di orientamento sul telelavoro rivolti a vari target e riguardanti, in particolare:

- gli aspetti normativi relativi al telelavoro
- le buone prassi per il telelavoro realizzate in Europa
- le iniziative di formazione realizzate
- un orientamento generale nel mondo del telelavoro in Europa.

Κείμενο Αναφοράς

A Telework Training, Guidance and Information System

THE PROJECT

The Project "TES – TELEWORK EDUCATION SYSTEM" witnesses participation by a partnership of five countries: Italy, Greece, Ireland, Romania, and Spain, composed of Public Administrations, Labour Agencies, Higher Scholastic Institutions, Universities, Research Centres, and Organizations dedicated to trainers.

The Project, realized in three years (2001 – 2003), has created an innovative model for a Training, Orientation, and Information System for Telework that can guarantee a proper introduction to telework in organisations through meaningful training programs for employees.

But in order for this training to be effective and efficient, TES has based its research on the founding assumption that the idea of "telework training" alone is not sufficient, and it is also necessary to talk about "telework education", which poses the educational event as the beginning of a process of acquiring the basic skills necessary both for workers and people preparing to enter the workplace, but also to people in the roles of decision makers and managers in organisations.

In this sense the Project sets goals that can entail powerful consequences both in the workplace and on the training and education system.

Telework (or "e-Work") is defined in its modern sense as "an employment methodology that presupposes distance supply of services, asynchronously if possible, using wholly or in part Information and Communications Technologies".

In this sense, the teleworker can represent a powerful tool in the evolution of organisations in the production process and labour management.

Any organisation (small or large, public or private), and even individuals working in the framework of self-employment, can obtain significant innovations in products and/or process from telework.

But in order for this to take place effectively, one of the aspects that these same organisations must handle is precisely that of training and updating of employees.

Knowing how to telework is not a skill simply tied to technical competence (knowing how to use a computer, the Internet, e-mail, etc.), but rather is an organizational and cultural skill.

The project aims to offer an innovative response to this need for training, starting from an experience already conducted in Naples, of which in some ways, it represents a natural continuation.

In order to reach its objectives, TES created an articulated technological infrastructure, distributed in nature, that provides training and information services.

This infrastructure is managed by a services centre—called TTC – TES Teleservices Center (located at the "F.GIORNDANI" Inst.—and uses one specific portal (TES Portal), which can be accessed at www.leonardotes.org, to provide the following services:

training services and support for trainers, such as:

- e-learning training courses on telework
- assistance in conceiving, planning, and creating new training modules and courses on telework
- use of experimental telework environments (based on the simulation of the processes)

training services for end-users, such as:

- introductory e-learning courses on telework and its application in various contexts
- supply of support materials and further information

information and orientation services on telework aimed at the various targets, and in particular regarding:

- the legal aspects involved with telework
- good practices for telework created in Europe
- training initiatives created
- a general orientation on the world of telework in Europe

Facebook Translate – Αξιολόγηση ακρίβειας και ευχέρειας

ST	MT Output	Adequacy	Fluency
Un Sistema di Formazione, Orientamento e Informazione sul Telelavoro	A Telework Training, Guidance and Information System	Excellent	Excellent
Il progetto	The project	Excellent	Excellent
Il Progetto “TES – TELEWORK EDUCATION SYSTEM”, realizzato nel triennio 2000-2003, ha visto la partecipazione di un partenariato di cinque paesi: Italia, Grecia, Irlanda, Romania e Spagna e composto da Pubbliche Amministrazioni, Agenzie per il Lavoro, Istituti Scolastici Superiori, Università, Centri di Ricerca, Organizzazioni di Base degli operatori della formazione.	The Project “ TES - TELEWORK EDUCATION SYSTEM ”, carried out in the three-year period 2000-2003, participated in a partnership of five countries: Italy, Greece, Ireland, Romania and Spain and composed of Public Administrations, Job Agencies, Institutes Superior Schools, University, Research Centers, Basic Organizations of Training Operators.	Poor	Good
Il Progetto ha realizzato un innovativo modello di Sistema per la Formazione, l’Orientamento e l’Informazione al Telelavoro che sia in grado di garantire, attraverso una significativa azione formativa dei lavoratori, una corretta introduzione del telelavoro stesso nelle organizzazioni.	The Project has created an innovative model of System for Training, Guidance and Telework Information that can ensure, through significant training action by workers, the correct introduction of teleworking itself into organizations.	Good	Excellent
Ma perché questa formazione sia efficace ed efficiente, TES ha posto alla base della sua ricerca l’assunto che l’idea di “formazione al telelavoro” non è sufficiente e che sia necessario parlare di “educazione al telelavoro”, intendendo il momento educativo come l’avvio di un processo di acquisizione di quelle competenze di base necessarie sia ai lavoratori che a coloro che si accingono ad entrare nel mondo del lavoro, ma anche a tutti coloro che hanno la funzione di decisori e gestori nelle organizzazioni.	But for this training to be efficient and efficient, TES has put the assumption that the idea of “ telework training ” is not enough and that it is necessary to talk about “ teleworking education ”, meaning the educational moment like the launching a process of acquiring the basic skills necessary for both workers and those who are about to enter the world of work, but also for all those who function as decision-makers and managers in organizations.	Poor	Excellent

ST	MT Output	Adequacy	Fluency
In tal senso il Progetto si pone obiettivi che possono avere ricadute molto forti sia sul mondo del lavoro che sul sistema formativo ed educativo.	In this respect, the Project sets goals that can have very strong consequences both on the world of work and on the educational and educational system.	Good	Excellent
Il Telelavoro (o "e-Work") è definito, nella sua moderna accezione, come "una metodologia di lavoro che presuppone la fornitura a distanza e, se possibile, in modo asincrono, di una prestazione lavorativa attraverso l'uso, in tutto o in parte, delle tecnologie dell'Informazione e della Comunicazione".	Telework (or " e-Work ") is defined, in its modern sense, as " a working methodology that presupposes the distance supply and, if possible, asynchronously, of a work performance through use, in full or partly of Information and Communication Technologies ".	Good	Fair
Il telelavoratore, in questo senso, può rappresentare un potente strumento di evoluzione delle organizzazioni nel processo di produzione e gestione del lavoro.	The teleworker, in this sense, can be a powerful tool for the organization s' evolution in the process of production and management of work.	Excellent	Fair
Una qualsiasi organizzazione (piccola o grande, pubblica o privata) ma anche singoli soggetti che operano nel contesto dell'auto-imprenditorialità, possono ottenere dal telelavoro significative innovazioni di prodotto e/o di processo.	Any organization (small or large, public or private) but also individual entities working in the context of self-entrepreneurship, can obtain significant product and / or process innovations from teleworking.	Excellent	Excellent
Ma perché queste siano effettivamente tali, uno degli aspetti che le stesse organizzazioni devono curare è proprio quello della formazione e dell'aggiornamento dei lavoratori.	But for these to actually be such, one of the aspects that the organizations themselves need to take care of is precisely the training and updating of workers.	Good	Excellent
Saper telelavorare non è una competenza legata semplicemente a conoscenze di tipo tecnologico (saper usare il computer, Internet, la posta elettronica, ecc.), ma è una competenza organizzativa e culturale.	Knowing how to teleworking is not a competence simply linked to technological knowledge (knowing how to use computer, Internet, inbox, etc.), but it's an organizational and cultural competence.	Excellent	Good
Il progetto vuole rispondere in modo innovativo a questa esigenza di formazione.	The project wants to innovatively respond to this training need.	Excellent	Excellent

ST	MT Output	Adequacy	Fluency
Per il raggiungimento dei suoi obiettivi TES ha realizzato un'articolata infrastruttura tecnologica di tipo distribuito che eroga i servizi formativi ed informativi.	In order to achieve its objectives, TES has created an articulated technology infrastructure of a distributed type that provides training and information services.	Good	Excellent
Questa infrastruttura è gestita da un centro servizi - denominato TTC - TES Teleservices Centre (collocato presso l'ITI "F'GIORDANI") – ed utilizza uno specifico portale (TES Portal), accessibile all'indirizzo www.leonardotes.org , per erogare i propri servizi:	This infrastructure is operated by a service center - TTC - TES Teleservices Centre (located at ITI " F ' GiORDANI ") - and uses a specific portal (TES Portal), accessible at www.leonardotes.org , to provide your services:	Good	Excellent
servizi formativi e di supporto per operatori della formazione , quali:	Training and support services for training providers, such as:	Excellent	Excellent
- corsi di formazione a distanza sul telelavoro,	- distance training courses on telework,	Excellent	Excellent
- supporto all'ideazione, progettazione, realizzazione di nuovi moduli formativi e corsi sul telelavoro,	- support for the idea, planning, the creation of new training modules and teleworking courses,	Fair	Excellent
- uso di ambienti di sperimentazione del telelavoro (basati sulla simulazione dei processi)	- use of telework experimentation environments (based on process simulation)	Excellent	Excellent
servizi formativi per utenti finali, quali:	Training services for end users, such as:	Excellent	Excellent
- corsi di formazione a distanza di introduzione al telelavoro e alla sua applicazione in vari contesti	- training courses at a distance of introduction to telework and its application in various contexts	Excellent	Good
- erogazione di materiali di supporto e di approfondimento	- supply of support and deepening materials	Fair	Excellent
servizi informativi e di orientamento sul telelavoro rivolti a vari target e riguardanti, in particolare:	Telework information and orientation services aimed at various targets, in particular:	Good	Excellent
- gli aspetti normativi relativi al telelavoro	- teleworking regulations	Excellent	Excellent
- le buone prassi per il telelavoro realizzate in Europa	- good practice for teleworking in Europe	Excellent	Good
- le iniziative di formazione realizzate	- training initiatives carried out	Excellent	Excellent

ST	MT Output	Adequacy	Fluency
- un orientamento generale nel mondo del telelavoro in Europa.	- a general orientation in the world of teleworking in Europe.	Excellent	Excellent

Google Translate – Αξιολόγηση ακρίβειας και ευχέρειας

ST	MT Output	Adequacy	Fluency
Un Sistema di Formazione, Orientamento e Informazione sul Telelavoro	A Training, Orientation and Information System on Telework	Excellent	Excellent
Il progetto	The project	Excellent	Excellent
Il Progetto “TES – TELEWORK EDUCATION SYSTEM”, realizzato nel triennio 2000-2003, ha visto la partecipazione di un partenariato di cinque paesi: Italia, Grecia, Irlanda, Romania e Spagna e composto da Pubbliche Amministrazioni, Agenzie per il Lavoro, Istituti Scolastici Superiori, Università, Centri di Ricerca, Organizzazioni di Base degli operatori della formazione.	The "TES - TELEWORK EDUCATION SYSTEM" Project, carried out in the three-year period 2000-2003, saw the participation of a partnership of five countries: Italy, Greece, Ireland, Romania and Spain and made up of Public Administrations, Employment Agencies, Schools Superiors, Universities, Research Centers, Basic Organizations of training operators.	Fair	Excellent
Il Progetto ha realizzato un innovativo modello di Sistema per la Formazione, l'Orientamento e l'Informazione al Telelavoro che sia in grado di garantire, attraverso una significativa azione formativa dei lavoratori, una corretta introduzione del telelavoro stesso nelle organizzazioni.	The Project has created an innovative system model for Telework Training, Orientation and Information that is able to ensure, through a significant training action for workers, a correct introduction of telework itself in organizations.	Excellent	Excellent
Ma perché questa formazione sia efficace ed efficiente, TES ha posto alla base della sua ricerca l'assunto che l'idea di “formazione al telelavoro” non è sufficiente e che sia necessario parlare di “educazione al telelavoro”, intendendo il momento educativo come l'avvio di un processo di acquisizione di quelle competenze di base necessarie sia ai lavoratori che a coloro che si accingono ad entrare nel mondo del lavoro, ma anche a tutti coloro che hanno la funzione di decisori e gestori nelle organizzazioni.	But for this training to be effective and efficient, TES has placed at the basis of its research the assumption that the idea of "telework training" is not enough and that it is necessary to speak of "telework education", meaning the educational moment as the launch of a process of acquiring those basic skills necessary for both workers and those who are about to enter the world of work, but also for all those who have the function of decision makers and managers in organizations.	Fair	Excellent

ST	MT Output	Adequacy	Fluency
In tal senso il Progetto si pone obiettivi che possono avere ricadute molto forti sia sul mondo del lavoro che sul sistema formativo ed educativo.	In this sense, the Project sets itself objectives that can have very strong repercussions both on the world of work and on the training and educational system.	Excellent	Excellent
Il Telelavoro (o "e-Work") è definito, nella sua moderna accezione, come "una metodologia di lavoro che presuppone la fornitura a distanza e, se possibile, in modo asincrono, di una prestazione lavorativa attraverso l'uso, in tutto o in parte, delle tecnologie dell'Informazione e della Comunicazione".	Teleworking (or "e-Work") is defined, in its modern meaning, as "a working methodology that presupposes the remote supply and, if possible, asynchronously, of a work performance through the use, in all or in part, of information and communication technologies".	Good	Good
Il telelavoratore, in questo senso, può rappresentare un potente strumento di evoluzione delle organizzazioni nel processo di produzione e gestione del lavoro.	In this sense, the teleworker can represent a powerful tool for the evolution of organizations in the process of production and management of work.	Excellent	Fair
Una qualsiasi organizzazione (piccola o grande, pubblica o privata) ma anche singoli soggetti che operano nel contesto dell'auto-imprenditorialità, possono ottenere dal telelavoro significative innovazioni di prodotto e/o di processo.	Any organization (small or large, public or private) but also individual subjects operating in the context of self-entrepreneurship, can obtain significant product and / or process innovations from teleworking.	Excellent	Excellent
Ma perché queste siano effettivamente tali, uno degli aspetti che le stesse organizzazioni devono curare è proprio quello della formazione e dell'aggiornamento dei lavoratori.	But for these to be effectively such, one of the aspects that the same organizations must take care of is precisely that of training and updating workers.	Fair	Excellent
Saper telelavorare non è una competenza legata semplicemente a conoscenze di tipo tecnologico (saper usare il computer, Internet, la posta elettronica, ecc.), ma è una competenza organizzativa e culturale.	Knowing how to telework is not a skill linked simply to technological knowledge (knowing how to use computers, Internet, e-mail, etc.), but it is an organizational and cultural competence.	Excellent	Excellent
Il progetto vuole rispondere in modo innovativo a questa esigenza di formazione.	The project aims to respond in an innovative way to this training need.	Excellent	Excellent

ST	MT Output	Adequacy	Fluency
Per il raggiungimento dei suoi obiettivi TES ha realizzato un'articolata infrastruttura tecnologica di tipo distribuito che eroga i servizi formativi ed informativi.	In order to achieve its objectives, TES has created an articulated distributed type technological infrastructure that provides training and information services.	Excellent	Excellent
Questa infrastruttura è gestita da un centro servizi - denominato TTC - TES Teleservices Centre (collocato presso l'ITI "F'GIORDANI") – ed utilizza uno specifico portale (TES Portal), accessibile all'indirizzo www.leonardotes.org , per erogare i propri servizi:	This infrastructure is managed by a service center - called TTC - TES Teleservices Center (located at the ITI "F'GIORDANI") - and uses a specific portal (TES Portal), accessible at www.leonardotes.org , to provide its services:	Excellent	Excellent
servizi formativi e di supporto per operatori della formazione , quali:	training and support services for training operators, such as:	Excellent	Excellent
- corsi di formazione a distanza sul telelavoro,	- distance training courses on teleworking,	Excellent	Excellent
- supporto all'ideazione, progettazione, realizzazione di nuovi moduli formativi e corsi sul telelavoro,	- support for the conception, design, implementation of new training modules and courses on teleworking,	Excellent	Excellent
- uso di ambienti di sperimentazione del telelavoro (basati sulla simulazione dei processi)	- use of telework testing environments (based on process simulation)	Excellent	Excellent
servizi formativi per utenti finali, quali:	training services for end users, such as:	Excellent	Excellent
- corsi di formazione a distanza di introduzione al telelavoro e alla sua applicazione in vari contesti	- distance training courses for introduction to teleworking and its application in various contexts	Excellent	Excellent
- erogazione di materiali di supporto e di approfondimento	- provision of support and in-depth materials	Excellent	Excellent
servizi informativi e di orientamento sul telelavoro rivolti a vari target e riguardanti, in particolare:	information and guidance services on telework aimed at various targets and concerning, in particular:	Excellent	Excellent
- gli aspetti normativi relativi al telelavoro	- the regulatory aspects relating to teleworking	Excellent	Excellent
- le buone prassi per il telelavoro realizzate in Europa	- good practices for teleworking implemented in Europe	Excellent	Excellent
- le iniziative di formazione realizzate	- the training initiatives carried out	Excellent	Excellent

ST	MT Output	Adequacy	Fluency
- un orientamento generale nel mondo del telelavoro in Europa.	- a general orientation in the world of teleworking in Europe.	Excellent	Excellent

Facebook Translate - Κατηγοριοποίηση λαθών

ST	MT output	No error	Accuracy	Addition	Omission	Mistranslation	Fluency	Punctuation	Grammar	Style	Awkward
Un Sistema di Formazione, Orientamento e Informazione sul Telelavoro	A Telework Training, Guidance and Information System	1									
Il progetto	The project	1									
Il Progetto "TES – TELEWORK EDUCATION SYSTEM", realizzato nel triennio 2000-2003, ha visto la partecipazione di un partenariato di cinque paesi: Italia, Grecia, Irlanda, Romania e Spagna e composto da Pubbliche Amministrazioni, Agenzie per il Lavoro, Istituti Scolastici Superiori, Università, Centri di Ricerca, Organizzazioni di Base degli operatori della formazione.	The Project " TES - TELEWORK EDUCATION SYSTEM ", carried out in the three-year period 2000-2003, participated in a partnership of five countries: Italy, Greece, Ireland, Romania and Spain and composed of Public Administrations, Job Agencies, Institutes Superior Schools, University, Research Centers, Basic Organizations of Training Operators.		4		1	3	1		1	1	1
Il Progetto ha realizzato un innovativo modello di Sistema per la Formazione, l'Orientamento e l'Informazione al Telelavoro che sia in grado di garantire, attraverso una significativa azione formativa dei lavoratori, una corretta introduzione del telelavoro stesso nelle	The Project has created an innovative model of System for Training, Guidance and Telework Information that can ensure, through significant training action by workers, the correct introduction of tele-		1			1					

ST	MT output	No error	Accuracy	Addition	Omission	Mistranslation	Fluency	Punctuation	Grammar	Style	Awkward
organizzazioni.	working itself into organizations.										
Ma perché questa formazione sia efficace ed efficiente, TES ha posto alla base della sua ricerca l'assunto che l'idea di "formazione al telelavoro" non è sufficiente e che sia necessario parlare di "educazione al telelavoro", intendendo il momento educativo come l'avvio di un processo di acquisizione di quelle competenze di base necessarie sia ai lavoratori che a coloro che si accingono ad entrare nel mondo del lavoro, ma anche a tutti coloro che hanno la funzione di decisori e gestori nelle organizzazioni.	But for this training to be efficient and efficient, TES has put the assumption that the idea of " telework training " is not enough and that it is necessary to talk about " teleworking education ", meaning the educational moment like the launching a process of acquiring the basic skills necessary for both workers and those who are about to enter the world of work, but also for all those who function as decision-makers and managers in organizations.		3		2	1				3	3

ST	MT output	No error	Accuracy	Addition	Omission	Mistranslation	Fluency	Punctuation	Grammar	Style	Awkward
In tal senso il Progetto si pone obiettivi che possono avere ricadute molto forti sia sul mondo del lavoro che sul sistema formativo ed educativo.	In this respect, the Project sets goals that can have very strong consequences both on the world of work and on the educational and educational system.		1			1					
Il Telelavoro (o "e-Work") è definito, nella sua moderna accezione, come "una metodologia di lavoro che presuppone la fornitura a distanza e, se possibile, in modo asincrono, di una prestazione lavorativa attraverso l'uso, in tutto o in parte, delle tecnologie dell'Informazione e della Comunicazione".	Telework (or "e-Work") is defined, in its modern sense, as "a working methodology that presupposes the distance supply and, if possible, asynchronously, of a work performance through use, in full or partly of Information and Communication Technologies".		1			1	3	1	2		
Il telelavoratore, in questo senso, può rappresentare un potente strumento di evoluzione delle organizzazioni nel processo di produzione e gestione del lavoro.	The teleworker, in this sense, can be a powerful tool for the organization s' evolution in the process of production and management of work.						2	1	1		

ST	MT output	No error	Accuracy	Addition	Omission	Mistranslation	Fluency	Punctuation	Grammar	Style	Awkward
Una qualsiasi organizzazione (piccola o grande, pubblica o privata) ma anche singoli soggetti che operano nel contesto dell'auto-imprenditorialità, possono ottenere dal telelavoro significative innovazioni di prodotto e/o di processo.	Any organization (small or large, public or private) but also individual entities working in the context of self-entrepreneurship, can obtain significant product and / or process innovations from teleworking.	1									
Ma perché queste siano effettivamente tali, uno degli aspetti che le stesse organizzazioni devono curare è proprio quello della formazione e dell'aggiornamento del lavoratori.	But for these to actually be such, one of the aspects that the organizations themselves need to take care of is precisely the training and updating of workers.		1		1						
Saper telelavorare non è una competenza legata semplicemente a conoscenze di tipo tecnologico (saper usare il computer, Internet, la posta elettronica, ecc.), ma è una competenza organizzativa e culturale.	Knowing how to teleworking is not a competence simply linked to technological knowledge (knowing how to use computer , Internet, inbox, etc.), but it's an organizational and cultural competence.						1		1		
Il progetto vuole rispondere in modo innovativo a questa esigenza di formazione.	The project wants to innovatively respond to this training need.	1									

ST	MT output	No error	Accuracy	Addition	Omission	Mistranslation	Fluency	Punctuation	Grammar	Style	Awkward
Per il raggiungimento dei suoi obiettivi TES ha realizzato un'articolata infrastruttura tecnologica di tipo distribuito che eroga i servizi formativi ed informativi.	In order to achieve its objectives, TES has created an articulated technology infrastructure of a distributed type that provides training and information services.		1	1							
Questa infrastruttura è gestita da un centro servizi - denominato TTC - TES Teleservices Centre (collocato presso l'ITI "F'GIORDANI") – ed utilizza uno specifico portale (TES Portal), accessibile all'indirizzo www.leonardotes.org , per erogare i propri servizi:	This infrastructure is operated by a service center - TTC - TES Teleservices Centre (located at ITI " F ' GiORDANI ") - and uses a specific portal (TES Portal), accessible at www.leonardotes.org , to provide your services:		2		1	1					
servizi formativi e di supporto per operatori della formazione , quali:	Training and support services for training providers, such as:	1									
- corsi di formazione a distanza sul telelavoro,	- distance training courses on telework,	1									
- supporto all'ideazione, progettazione, realizzazione di nuovi moduli formativi e corsi sul telelavoro,	- support for the idea , planning, the creation of new training modules and teleworking courses,		2			2					
- uso di ambienti di sperimentazione del telelavoro (basati sulla simulazione dei processi)	- use of telework experimentation environments (based on process simulation)	1									

ST	MT output	No error	Accuracy	Addition	Omission	Mistranslation	Fluency	Punctuation	Grammar	Style	Awkward
servizi formativi per utenti finali, quali:	Training services for end users, such as:	1									
- corsi di formazione a distanza di introduzione al telelavoro e alla sua applicazione in vari contesti	- training courses at a distance of introduction to telework and its application in various contexts						1		1		
- erogazione di materiali di supporto e di approfondimento	- supply of support and deepening materials		1			1					
servizi informativi e di orientamento sul telelavoro rivolti a vari target e riguardanti, in particolare:	Telework information and orientation services aimed at various targets, in particular:		1		1						
- gli aspetti normativi relativi al telelavoro	- teleworking regulations	1									
- le buone prassi per il telelavoro realizzate in Europa	- good practice for teleworking in Europe						1		1		
- le iniziative di formazione realizzate	- training initiatives carried out	1									
- un orientamento generale nel mondo del telelavoro in Europa.	- a general orientation in the world of teleworking in Europe.	1									

Google Translate - Κατηγοριοποίηση λαθών

ST	MT output	No error	Accuracy	Omission	Mistranslation	Fluency	Grammar	Style	Awkward
Un Sistema di Formazione, Orientamento e Informazione sul Telelavoro	A Training, Orientation and Information System on Telework	1							
Il progetto	The project	1							
Il Progetto "TES – TELEWORK EDUCATION SYSTEM", realizzato nel triennio 2000-2003, ha visto la partecipazione di un partenariato di cinque paesi: Italia, Grecia, Irlanda, Romania e Spagna e composto da Pubbliche Amministrazioni, Agenzie per il Lavoro, Istituti Scolastici Superiori, Università, Centri di Ricerca, Organizzazioni di Base degli operatori della formazione.	The "TES - TELEWORK EDUCATION SYSTEM" Project, carried out in the three-year period 2000-2003, saw the participation of a partnership of five countries: Italy, Greece, Ireland, Romania and Spain and made up of Public Administrations, Employment Agencies, Schools Superiors, Universities, Research Centers, Basic Organizations of training operators.		3	1	2				
Il Progetto ha realizzato un innovativo modello di Sistema per la Formazione, l'Orientamento e l'Informazione al Telelavoro che sia in grado di garantire, attraverso una significativa azione formativa dei lavoratori, una corretta introduzione del telelavoro stesso nelle organizzazioni.	The Project has created an innovative system model for Telework Training, Orientation and Information that is able to ensure, through a significant training action for workers, a correct introduction of telework itself in organizations.	1							

ST	MT output	No error	Accuracy	Omission	Mistranslation	Fluency	Grammar	Style	Awkward
Ma perché questa formazione sia efficace ed efficiente, TES ha posto alla base della sua ricerca l'assunto che l'idea di "formazione al telelavoro" non è sufficiente e che sia necessario parlare di "educazione al telelavoro", intendendo il momento educativo come l'avvio di un processo di acquisizione di quelle competenze di base necessarie sia ai lavoratori che a coloro che si accingono ad entrare nel mondo del lavoro, ma anche a tutti coloro che hanno la funzione di decisori e gestori nelle organizzazioni.	But for this training to be effective and efficient, TES has placed at the basis of its research the assumption that the idea of "telework training" is not enough and that it is necessary to speak of "telework education", meaning the educational moment as the launch of a process of acquiring those basic skills necessary for both workers and those who are about to enter the world of work, but also for all those who have the function of decision makers and managers in organizations.		2	1	1			2	2
In tal senso il Progetto si pone obiettivi che possono avere ricadute molto forti sia sul mondo del lavoro che sul sistema formativo ed educativo.	In this sense, the Project sets itself objectives that can have very strong repercussions both on the world of work and on the training and educational system.	1							
Il Telelavoro (o "e-Work") è definito, nella sua moderna accezione, come "una metodologia di lavoro che presuppone la fornitura a distanza e, se possibile, in modo asincrono, di una prestazione lavorativa attraverso l'uso, in tutto o in parte, delle tecnologie dell'Informazione e della Comunicazione".	Teleworking (or "e-Work") is defined, in its modern meaning, as "a working methodology that presupposes the remote supply and, if possible, asynchronously, of a work performance through the use, in all or in part, of information and communication technologies".		1		1	1	1		

ST	MT output	No error	Accuracy	Omission	Mistranslation	Fluency	Grammar	Style	Awkward
Il telelavoratore, in questo senso, può rappresentare un potente strumento di evoluzione delle organizzazioni nel processo di produzione e gestione del lavoro.	In this sense, the teleworker can represent a powerful tool for the evolution of organizations in the process of production and management of work.					1	1		
Una qualsiasi organizzazione (piccola o grande, pubblica o privata) ma anche singoli soggetti che operano nel contesto dell'auto-imprenditorialità, possono ottenere dal telelavoro significative innovazioni di prodotto e/o di processo.	Any organization (small or large, public or private) but also individual subjects operating in the context of self-entrepreneurship, can obtain significant product and / or process innovations from teleworking.	1							
Ma perché queste siano effettivamente tali, uno degli aspetti che le stesse organizzazioni devono curare è proprio quello della formazione e dell'aggiornamento dei lavoratori.	But for these to be effectively such, one of the aspects that the same organizations must take care of is precisely that of training and updating workers.		2	1	1				
Saper telelavorare non è una competenza legata semplicemente a conoscenze di tipo tecnologico (saper usare il computer, Internet, la posta elettronica, ecc.), ma è una competenza organizzativa e culturale.	Knowing how to telework is not a skill linked simply to technological knowledge (knowing how to use computers, Internet, e-mail, etc.), but it is an organizational and cultural competence.	1							
Il progetto vuole rispondere in modo innovativo a questa esigenza di formazione.	The project aims to respond in an innovative way to this training need.	1							
Per il raggiungimento dei suoi obiettivi TES ha realizzato un'articolata infrastruttura tecnologica di tipo distribuito che eroga i servizi formativi ed informativi.	In order to achieve its objectives, TES has created an articulated distributed type technological infrastructure that provides training and information services.	1							

ST	MT output	No error	Accuracy	Omission	Mistranslation	Fluency	Grammar	Style	Awkward
Questa infrastruttura è gestita da un centro servizi - denominato TTC - TES Teleservices Centre (collocato presso l'ITI "F'GIORDANI") – ed utilizza uno specifico portale (TES Portal), accessibile all'indirizzo www.leonardotes.org , per erogare i propri servizi:	This infrastructure is managed by a service center - called TTC - TES Teleservices Center (located at the ITI "F'GIORDANI") - and uses a specific portal (TES Portal), accessible at www.leonardotes.org , to provide its services:	1							
servizi formativi e di supporto per operatori della formazione , quali:	training and support services for training operators, such as:	1							
- corsi di formazione a distanza sul telelavoro,	- distance training courses on teleworking,	1							
- supporto all'ideazione, progettazione, realizzazione di nuovi moduli formativi e corsi sul telelavoro,	- support for the conception, design, implementation of new training modules and courses on teleworking,	1							
- uso di ambienti di sperimentazione del telelavoro (basati sulla simulazione dei processi)	- use of telework testing environments (based on process simulation)		1		1				
servizi formativi per utenti finali, quali:	training services for end users, such as:	1							
- corsi di formazione a distanza di introduzione al telelavoro e alla sua applicazione in vari contesti	- distance training courses for introduction to teleworking and its application in various contexts	1							
- erogazione di materiali di supporto e di approfondimento	- provision of support and in-depth materials	1							
servizi informativi e di orientamento sul telelavoro rivolti a vari target e riguardanti, in particolare:	information and guidance services on telework aimed at various targets and concerning, in particular:	1							

ST	MT output	No error	Accuracy	Omission	Mistranslation	Fluency	Grammar	Style	Awkward
- gli aspetti normativi relativi al telelavoro	- the regulatory aspects relating to teleworking	1							
- le buone prassi per il telelavoro realizzate in Europa	- good practices for teleworking implemented in Europe	1							
- le iniziative di formazione realizzate	- the training initiatives carried out	1							
- un orientamento generale nel mondo del telelavoro in Europa.	- a general orientation in the world of teleworking in Europe.	1							

La Cina e il vino italiano

Κείμενο Αφιετηρίας

La Cina e il vino italiano...

Che cosa ci fa una laureata in lingua cinese, specializzata in economia e storia dei paesi dell'Estremo Oriente, in una prestigiosa cantina di Langa?

Roberta Picco, 30 anni da Saluzzo, ha deciso di collegare le sue due grandi passioni, il vino e la cultura orientale, e per capire e mettere in pratica questo ha percorso molti chilometri e trascorso anni in Cina, per poi approdare al Master in Cultura del Vino Italiano a Pollenzo.

“Ho studiato Lingue orientali a Venezia, quindi ho proseguito con una laurea magistrale in economia e storia della Cina.

La Cina è sempre stata il mio sogno e terminati gli studi sono partita e vi sono rimasta circa due anni, curando i rapporti per aziende italiane in loco.

Non mi sono fermata, però, e rientrata in Italia ho lavorato a Genova nel settore dell'import-export” ci racconta Roberta, nella pausa pranzo del suo tirocinio presso Gaja a Barbaresco.

Il vino, però rimane un chiodo fisso per Roberta, e quindi la sua esperienza nel campo enologico continua:

lei vuole saperne di più, inizia a studiare, a praticare il mondo del vino, sia per lavoro che per curiosità. In particolare utilizza le sue competenze linguistiche collaborando con alcune aziende vinicole.

Ho quindi lavorato principalmente come interprete presso cantine in Langa, cercando di approfondire la mia conoscenza dei vini.

Durante i miei soggiorni in Cina ho potuto capire che, ancora per molti, il vino rappresenta uno status symbol, qualcosa per ricchi.

Con il passare del tempo però, il consumo di vino si sta allargando non solo nelle città principali, ma anche in quelle cosiddette di seconda fascia, che sono comunque sempre delle metropoli enormi:

si tratta quindi di un settore emergente che ha un grande futuro – continua Roberta.

“Dunque ormai nuovamente stabilitami in Italia mi sono attivata per trovare una formazione adeguata:

da anni seguivo l'attività dell'Università di Pollenzo:

due anni fa ho partecipato ad una Giornata Aperta, ma il Master non era ancora partito.

Sono rimasta in contatto con l'UNISG e quando ho scoperto che il corso era stato attivato, mi sono iscritta”.

Roberta modestamente ammette di non aver ancora imparato tutto sul vino:

“Il mondo del vino è molto variegato, c’è sempre da imparare e in un anno non si può apprendere tutto.

Però il Master è una buona base dalla quale partire per ampliare le proprie conoscenze.

“Mi occupo delle visite in cantina, che avvengono su prenotazione.

La visita presso Gaja è un momento di grande formazione, per me è una sfida perché mi sto impegnando al massimo per essere all’altezza.

Inoltre essere qui è una straordinaria opportunità di crescita per me:

sto imparando moltissimo ogni giorno”.

Ci viene spontaneo chiederle se ha già fatto visite in cinese.

“Non ancora – risponde Roberta – ad oggi effettuo le visite in inglese.

La Cina rimane però sempre nei miei obiettivi.

Oggi purtroppo è ancora la Francia che gioca un ruolo principale sul mercato cinese:

se pensano al vino, i cinesi pensano a quello francese.

Vorrei provare a cambiare questa abitudine:

mi piacerebbe che in Cina il binomio vino e Italia fosse vincente”.

Κείμενο Αναφοράς

Italian wine meets China...

What is an Asian language graduate with a specialization in Chinese economics and history doing in a prestigious Langhe winery?

Roberta Picco, 30, originally from Saluzzo, has made it her life's goal to connect her two great passions, wine and Asian culture. This has led her to travel many miles and spend several years in China before bringing her to the Master in Italian Wine Culture in Pollenzo.

"I studied Asian languages in Venice, then continued with a graduate degree in Chinese economics and history.

China has always been my dream, and once I finished my studies I spent about two years there, managing relations for Italian companies in the country.

I didn't stay, though, and once back in Italy I worked in Genoa in the import-export sector," Roberta tells us during her lunch break (she's currently interning at Gaja in Barbaresco—more about that later).

Wine has always been a fixation for Roberta, and now she is expanding her experience in the enological field.

She wants to know more, to study and be immersed in the world of wine, for both professional and personal reasons.

"So I've worked as an interpreter in wineries in the Langhe, trying to deepen my knowledge of wine.

During my time in China I realized how wine still represents a status symbol for many people, something for the rich.

But over time wine consumption is starting to grow, and not just in the main cities, but also in the so-called 'second-tier' cities, which are still enormous metropolises.

So this is an emerging sector with a great future," continues Roberta.

"Once I'd settled back in Italy, I started looking for a suitable course.

For years I'd been following the activities of the University of Gastronomic Sciences in Pollenzo.

Two years ago I came to an Open Day, but the Master hadn't started yet.

I stayed in touch with UNISG and when I found out the program had been launched, I enrolled."

Roberta makes it clear she still has a lot to learn about wine.

“The world of wine is very varied, there’s always more to find out and in one year you can’t learn everything.

But the Master is a good base from which to broaden your knowledge.

"I take care of winery visits, which are booked in advance.

A visit to Gaja is very educational experience, and for me it’s also a challenge because I’m working as hard as possible to make sure I’m up to the task.

Plus being here is an extraordinary opportunity for growth for me.

I’m learning so much every day.”

We wonder if she’s already given tours in Chinese.

“Not yet,” she answers. “So far I’ve only done them in English.

But China is still in my sights.

At the moment unfortunately it’s France that plays the biggest role in the Chinese market.

If the Chinese think about wine, they think about French wine.

I’d like to try to change that.

I’d like to see the pairing of wine and Italy triumphing in China.”

Facebook Translate – Αξιολόγηση ακρίβειας και ευχέρειας

ST	MT Output	Adequacy	Fluency
La Cina e il vino italiano...	China and Italian wine...	Excellent	Excellent
Che cosa ci fa una laureata in lingua cinese, specializzata in economia e storia dei paesi dell'Estremo Oriente, in una prestigiosa cantina di Langa?	What is a Chinese language graduate doing, specializing in economics and history of the Far East countries, in a prestigious basement of Langa?	Fair	Good
Roberta Picco, 30 anni da Saluzzo, ha deciso di collegare le sue due grandi passioni, il vino e la cultura orientale, e per capire e mettere in pratica questo ha percorso molti chilometri e trascorso anni in Cina, per poi approdare al Master in Cultura del Vino Italiano a Pollenzo.	Roberta Picco, 30 years old from Saluzzo, decided to connect his two great passions, wine and eastern culture, and to understand and practice this he travelled many kilometers and spent years in China, and then landed to the Master's degree Italian wine culture in Pollenzo.	Fair	Excellent
"Ho studiato Lingue orientali a Venezia, quindi ho proseguito con una laurea magistrale in economia e storia della Cina.	" I studied oriental languages in Venice, so I proceeded with a master's degree in economics and history of China.	Excellent	Excellent
La Cina è sempre stata il mio sogno e terminati gli studi sono partita e vi sono rimasta circa due anni, curando i rapporti per aziende italiane in loco.	China has always been my dream and I have finished my studies and I have left for about two years, treating reports for Italian companies on site.	Fair	Excellent
Non mi sono fermata, però, e rientrata in Italia ho lavorato a Genova nel settore dell'import-export" ci racconta Roberta, nella pausa pranzo del suo tirocinio presso Gaja a Barbaresco.	I didn't stop, however, and I went back to Italy I worked in Genoa in the import-export sector " Roberta tells us, during the lunch break of her internship at Gaja in Barbaresco.	Fair	Good
Il vino, però rimane un chiodo fisso per Roberta, e quindi la sua esperienza nel campo enologico continua:	Wine, however, remains a fixed nail for Roberta, and therefore her experience in the oenological field continues:	Poor	Excellent
lei vuole saperne di più, inizia a studiare, a praticare il mondo del vino, sia per lavoro che per curiosità.	she wants to know more, she starts studying, practicing the wine world, both for work and for curiosity.	Good	Excellent

ST	MT Output	Adequacy	Fluency
"Ho quindi lavorato principalmente come interprete presso cantine in Langa, cercando di approfondire la mia conoscenza dei vini.	"I therefore worked mainly as an interpreter at wineries in Langa, trying to deepen my knowledge of wines.	Good	Excellent
Durante i miei soggiorni in Cina ho potuto capire che, ancora per molti, il vino rappresenta uno status symbol, qualcosa per ricchi.	During my stays in China I could understand that, still for many, wine represents a status symbol, something for rich people.	Good	Excellent
Con il passare del tempo però, il consumo di vino si sta allargando non solo nelle città principali, ma anche in quelle cosiddette di seconda fascia, che sono comunque sempre delle metropoli enormi:	As time passes, however, wine consumption is expanding not only in the main cities, but also in the so-called second-end cities, which are always huge metropolis,	Fair	Good
si tratta quindi di un settore emergente che ha un grande futuro – continua Roberta.	so it is an emerging sector with a great future - Roberta continues.	Good	Excellent
"Dunque ormai nuovamente stabilitami in Italia mi sono attivata per trovare una formazione adeguata:	" So now I'm established again in Italy, I've activated myself to find an adequate training:	Fair	Excellent
da anni seguivo l'attività dell'Università di Pollenzo:	for years I've been following the activity of the University of Pollenzo:	Excellent	Excellent
due anni fa ho partecipato ad una Giornata Aperta, ma il Master non era ancora partito.	two years ago I participated in an open day, but the Master hadn't started yet.	Excellent	Excellent
Sono rimasta in contatto con l'UNISG e quando ho scoperto che il corso era stato attivato, mi sono iscritta".	I stayed in touch with UNISG and when I found out the course was activated, I signed up ".	Good	Excellent
Roberta modestamente ammette di non aver ancora imparato tutto sul vino:	Roberta modestly admits that she hasn't learned everything about wine yet:	Excellent	Excellent
"Il mondo del vino è molto variegato, c'è sempre da imparare e in un anno non si può apprendere tutto.	" The world of wine is very varied, there's always to learn and in a year you can't learn everything.	Good	Excellent

ST	MT Output	Adequacy	Fluency
Però il Master è una buona base dalla quale partire per ampliare le proprie conoscenze.	But the Master is a good foundation from which to start to expand your knowledge.	Good	Good
“Mi occupo delle visite in cantina, che avvengono su prenotazione.	” I take care of cellar visits, which happen on reservation.	Good	Excellent
La visita presso Gaja è un momento di grande formazione, per me è una sfida perché mi sto impegnando al massimo per essere all'altezza.	Visiting Gaja is a moment of great training, for me it's a challenge because I'm trying hard to be up to it.	Fair	Excellent
Inoltre essere qui è una straordinaria opportunità di crescita per me:	Also being here is an amazing opportunity to grow for me:	Excellent	Good
sto imparando moltissimo ogni giorno”.	I'm learning a lot everyday ”.	Excellent	Excellent
Ci viene spontaneo chiederle se ha già fatto visite in cinese.	It's spontaneous to ask her if she's already made visits to Chinese.	Poor	Excellent
“Non ancora – risponde Roberta – ad oggi effettuo le visite in inglese.	” Not yet - Roberta responds - I'm visiting English today.	Poor	Excellent
La Cina rimane però sempre nei miei obiettivi.	China always stays in my goals though.	Excellent	Excellent
Oggi purtroppo è ancora la Francia che gioca un ruolo principale sul mercato cinese:	Unfortunately, France is still playing a leading role in the Chinese market today:	Excellent	Excellent
se pensano al vino, i cinesi pensano a quello francese.	if they think of wine, the Chinese think of the French one.	Excellent	Excellent
Vorrei provare a cambiare questa abitudine:	I would like to try to change this habit:	Excellent	Good
mi piacerebbe che in Cina il binomio vino e Italia fosse vincente”.	I would like to see the combination of wine and Italy win in China ”.	Good	Excellent

Google Translate – Αξιολόγηση ακρίβειας και ευχέρειας

ST	MT Output	Adequacy	Fluency
La Cina e il vino italiano...	China and Italian wine...	Excellent	Excellent
Che cosa ci fa una laureata in lingua cinese, specializzata in economia e storia dei paesi dell'Estremo Oriente, in una prestigiosa cantina di Langa?	What is a Chinese-language graduate doing, specializing in economics and history of Far Eastern countries, in a prestigious Langa winery?	Good	Good
Roberta Picco, 30 anni da Saluzzo, ha deciso di collegare le sue due grandi passioni, il vino e la cultura orientale, e per capire e mettere in pratica questo ha percorso molti chilometri e trascorso anni in Cina, per poi approdare al Master in Cultura del Vino Italiano a Pollenzo.	Roberta Picco, 30 years old from Saluzzo, has decided to connect her two great passions, wine and oriental culture, and to understand and put this into practice she has traveled many kilometers and spent years in China, before arriving at the Master in Culture of Italian Wine in Pollenzo.	Good	Excellent
"Ho studiato Lingue orientali a Venezia, quindi ho proseguito con una laurea magistrale in economia e storia della Cina.	"I studied Oriental Languages in Venice, then I continued with a master's degree in economics and history of China.	Excellent	Excellent
La Cina è sempre stata il mio sogno e terminati gli studi sono partita e vi sono rimasta circa due anni, curando i rapporti per aziende italiane in loco.	China has always been my dream and after completing my studies I left and stayed there for about two years, taking care of the relationships for Italian companies on site.	Good	Excellent
Non mi sono fermata, però, e rientrata in Italia ho lavorato a Genova nel settore dell'import-export" ci racconta Roberta, nella pausa pranzo del suo tirocinio presso Gaja a Barbaresco.	I didn't stop, however, and when I returned to Italy I worked in Genoa in the import-export sector "Roberta tells us, during the lunch break of her internship at Gaja in Barbaresco.	Good	Excellent
Il vino, però rimane un chiodo fisso per Roberta, e quindi la sua esperienza nel campo enologico continua:	Wine, however, remains an obsession for Roberta, and therefore her experience in the oenological field continues:	Excellent	Excellent

ST	MT Output	Adequacy	Fluency
lei vuole saperne di più, inizia a studiare, a praticare il mondo del vino, sia per lavoro che per curiosità.	she wants to know more, she begins to study, to practice the world of wine, both for work and for curiosity.	Good	Excellent
"Ho quindi lavorato principalmente come interprete presso cantine in Langa, cercando di approfondire la mia conoscenza dei vini.	"I therefore worked mainly as an interpreter at wineries in the Langa, trying to deepen my knowledge of wines.	Good	Excellent
Durante i miei soggiorni in Cina ho potuto capire che, ancora per molti, il vino rappresenta uno status symbol, qualcosa per ricchi.	During my stays in China I was able to understand that, still for many, wine represents a status symbol, something for the rich.	Excellent	Excellent
Con il passare del tempo però, il consumo di vino si sta allargando non solo nelle città principali, ma anche in quelle cosiddette di seconda fascia, che sono comunque sempre delle metropoli enormi:	With the passage of time, however, the consumption of wine is expanding not only in the main cities, but also in the so-called second-tier ones, which are always huge metropolises:	Good	Excellent
si tratta quindi di un settore emergente che ha un grande futuro – continua Roberta.	it is therefore an emerging sector that has a great future – Roberta continues.	Excellent	Excellent
“Dunque ormai nuovamente stabilitami in Italia mi sono attivata per trovare una formazione adeguata:	So by now, having settled back in Italy, I took steps to find adequate training:	Excellent	Excellent
da anni seguivo l’attività dell’Università di Pollenzo:	for years I had been following the activities of the University of Pollenzo:	Excellent	Excellent
due anni fa ho partecipato ad una Giornata Aperta, ma il Master non era ancora partito.	two years ago I participated in an Open Day, but the Master had not yet started.	Excellent	Excellent
Sono rimasta in contatto con l’UNISG e quando ho scoperto che il corso era stato attivato, mi sono iscritta”.	I kept in touch with UNISG and when I found out that the course had been activated, I signed up .	Excellent	Excellent
Roberta modestamente ammette di non aver ancora imparato tutto sul vino:	Roberta modestly admits that she has not yet learned everything about wine:	Excellent	Excellent

ST	MT Output	Adequacy	Fluency
“Il mondo del vino è molto variegato, c’è sempre da imparare e in un anno non si può apprendere tutto.	"The world of wine is very varied, there is always something to learn and you can't learn everything in a year.	Excellent	Excellent
Però il Master è una buona base dalla quale partire per ampliare le proprie conoscenze.	However, the Master is a good basis from which to start to expand one's knowledge.	Good	Good
“Mi occupo delle visite in cantina, che avvengono su prenotazione.	“I take care of the visits to the cellar, which take place by appointment.	Excellent	Excellent
La visita presso Gaja è un momento di grande formazione, per me è una sfida perché mi sto impegnando al massimo per essere all’altezza.	The visit to Gaja is a time of great training, for me it is a challenge because I am doing my utmost to be up to it.	Fair	Excellent
Inoltre essere qui è una straordinaria opportunità di crescita per me:	Also being here is an extraordinary growth opportunity for me:	Excellent	Good
sto imparando moltissimo ogni giorno”.	I am learning a lot every day ”.	Excellent	Excellent
Ci viene spontaneo chiederle se ha già fatto visite in cinese.	We naturally ask you if you have already made visits in Chinese.	Very Poor	Excellent
“Non ancora – risponde Roberta – ad oggi effettuo le visite in inglese.	“Not yet - Roberta replies - to date I carry out the visits in English.	Excellent	Excellent
La Cina rimane però sempre nei miei obiettivi.	However, China always remains in my goals.	Excellent	Excellent
Oggi purtroppo è ancora la Francia che gioca un ruolo principale sul mercato cinese:	Today, unfortunately, it is still France that plays a main role on the Chinese market:	Excellent	Excellent
se pensano al vino, i cinesi pensano a quello francese.	if they think of wine, the Chinese think of the French one.	Excellent	Excellent
Vorrei provare a cambiare questa abitudine:	I would like to try to change this habit:	Excellent	Good
mi piacerebbe che in Cina il binomio vino e Italia fosse vincente”.	I would like the combination of wine and Italy to be successful in China ”.	Excellent	Excellent

Facebook Translate - Κατηγοριοποίηση λαθών

ST	MT output	No error	Accuracy	Addition	Omission	Mistranslation	Untranslated	Fluency	Punctuation	Grammar	Style	Awkward
La Cina e il vino italiano...	China and Italian wine...	1										
Che cosa ci fa una laureata in lingua cinese, specializzata in economia e storia dei paesi dell'Estremo Oriente, in una prestigiosa cantina di Langa?	What is a Chinese language graduate doing , specializing in economics and history of the Far East countries, in a prestigious base-ment of Langa?		2			1	1	1		1		
Roberta Picco, 30 anni da Saluzzo, ha deciso di collegare le sue due grandi passioni, il vino e la cultura orientale, e per capire e mettere in pratica questo ha percorso molti chilometri e trascorso anni in Cina, per poi approdare al Master in Cultura del Vino Italiano a Pollenzo.	Roberta Picco, 30 years old from Saluzzo, decided to connect his two great passions, wine and eastern culture, and to understand and practice this he travelled many kilometers and spent years in China, and then landed to the Master's degree Italian wine culture in Pollenzo.		5		2	3					2	2
"Ho studiato Lingue orientali a Venezia, quindi ho proseguito con una laurea magistrale in eco-	"I studied oriental languages in Venice, so I proceeded with a master's degree in econom-	1										

ST	MT output	No error	Accuracy	Addition	Omission	Mistranslation	Untranslated	Fluency	Punctuation	Grammar	Style	Awkward
nomia e storia della Cina.	ics and history of China.											
La Cina è sempre stata il mio sogno e terminati gli studi sono partita e vi sono rimasta circa due anni, curando i rapporti per aziende italiane in loco.	China has always been my dream and I have finished my studies and I have left for about two years, treating reports for Italian companies on site.		3		1	2						
Non mi sono fermata, però, e rientrata in Italia ho lavorato a Genova nel settore dell'import-export" ci racconta Roberta, nella pausa pranzo del suo tirocinio presso Gaja a Barbaresco.	I didn't stop, however, and I went back to Italy I worked in Genoa in the import-export sector " Roberta tells us, during the lunch break of her internship at Gaja in Barbaresco.		1			1		1	1		1	1
Il vino, però rimane un chiodo fisso per Roberta, e quindi la sua esperienza nel campo enologico continua:	Wine, however, remains a fixed nail for Roberta, and therefore her experience in the oenological field continues:		1			1						
lei vuole saperne di più, inizia a studiare, a praticare il mondo del vino, sia	she wants to know more, she starts studying, practicing the wine world ,		1		1							

ST	MT output	No error	Accuracy	Addition	Omission	Mistranslation	Untranslated	Fluency	Punctuation	Grammar	Style	Awkward
per lavoro che per curiosità.	both for work and for curiosity.											
"Ho quindi lavorato principalmente come interprete presso cantine in Langa, cercando di approfondire la mia conoscenza dei vini.	"I therefore worked mainly as an interpreter at wineries in Langa, trying to deepen my knowledge of wines.		1				1					
Durante i miei soggiorni in Cina ho potuto capire che, ancora per molti, il vino rappresenta uno status symbol, qualcosa per ricchi.	During my stays in China I could understand that, still for many, wine represents a status symbol, something for rich people.		1			1						
Con il passare del tempo però, il consumo di vino si sta allargando non solo nelle città principali, ma anche in quelle cosiddette di seconda fascia, che sono comunque sempre delle metropoli enormi:	As time passes, however, wine consumption is expanding not only in the main cities, but also in the so-called second-end cities, which are always huge metropolis,		1			1		1		1		
si tratta quindi di un settore emergente che ha un grande futuro – continua Roberta.	so it is an emerging sector with a great future - Roberta continues.										1	1

ST	MT output	No error	Accuracy	Addition	Omission	Mistranslation	Untranslated	Fluency	Punctuation	Grammar	Style	Awkward
“Dunque ormai nuovamente stabilitami in Italia mi sono attivata per trovare una formazione adeguata:	“ So now I'm established again in Italy, I've activated myself to find an adequate training:		2	1		1						
da anni seguivo l'attività dell'Università di Pollenzo:	for years I've been following the activity of the University of Pollenzo:	1										
due anni fa ho partecipato ad una Giornata Aperta, ma il Master non era ancora partito.	two years ago I participated in an open day, but the Master hadn't started yet.	1										
Sono rimasta in contatto con l'UNISG e quando ho scoperto che il corso era stato attivato, mi sono iscritta”.	I stayed in touch with UNISG and when I found out the course was activated, I signed up ”.		1		1							
Roberta modestamente ammette di non aver ancora imparato tutto sul vino:	Roberta modestly admits that she hasn't learned everything about wine yet:	1										
“Il mondo del vino è molto variegato, c'è sempre da imparare e in un anno non si può apprendere tutto.	“ The world of wine is very varied, there's always to learn and in a year you can't learn everything.		1		1							

ST	MT output	No error	Accuracy	Addition	Omission	Mistranslation	Untranslated	Fluency	Punctuation	Grammar	Style	Awkward
Però il Master è una buona base dalla quale partire per ampliare le proprie conoscenze.	But the Master is a good foundation from which to start to expand your knowledge.		1	1				1		1	1	1
“Mi occupo delle visite in cantina, che avvengono su prenotazione.	” I take care of cellar visits, which happen on reservation.											
La visita presso Gaja è un momento di grande formazione, per me è una sfida perché mi sto impegnando al massimo per essere all'altezza.	Visiting Gaja is a moment of great training, for me it's a challenge because I'm trying hard to be up to it.		1			1						
Inoltre essere qui è una straordinaria opportunità di crescita per me:	Also being here is an amazing opportunity to grow for me:							1		1	1	1
sto imparando moltissimo ogni giorno”.	I'm learning a lot everyday ”.	1										
Ci viene spontaneo chiederle se ha già fatto visite in cinese.	It's spontaneous to ask her if she's already made visits to Chinese.		1			1						
“Non ancora – risponde Roberta – ad oggi effettuo le visite in inglese.	” Not yet - Roberta responds - I'm visiting English today.		2			2						
La Cina rimane però sempre nei miei obiettivi.	China always stays in my goals though.	1										

ST	MT output	No error	Accuracy	Addition	Omission	Mistranslation	Untranslated	Fluency	Punctuation	Grammar	Style	Awkward
Oggi purtroppo è ancora la Francia che gioca un ruolo principale sul mercato cinese:	Unfortunately, France is still playing a leading role in the Chinese market today:	1										
se pensano al vino, i cinesi pensano a quello francese.	if they think of wine, the Chinese think of the French one.	1										
Vorrei provare a cambiare questa abitudine:	I would like to try to change this habit:										1	1
mi piacerebbe che in Cina il binomio vino e Italia fosse vincente".	I would like to see the combination of wine and Italy win in China ".										1	1

Google Translate - Κατηγοριοποίηση λαθών

ST	MT output	No error	Accuracy	Addition	Omission	Mistranslation	Untranslated	Fluency	Grammar	Style	Awkward
La Cina e il vino italiano...	China and Italian wine...	1									
Che cosa ci fa una laureata in lingua cinese, specializzata in economia e storia dei paesi dell'Estremo Oriente, in una prestigiosa cantina di Langa?	What is a Chinese-language graduate doing , specializing in economics and history of Far Eastern countries, in a prestigious Langa winery?		1				1	1	1		
Roberta Picco, 30 anni da Saluzzo, ha deciso di collegare le sue due grandi passioni, il vino e la cultura orientale, e per capire e mettere in pratica questo ha percorso molti chilometri e trascorso anni in Cina, per poi approdare al Master in Cultura del Vino Italiano a Pollenzo.	Roberta Picco, 30 years old from Saluzzo, has decided to connect her two great passions, wine and oriental culture, and to understand and put this into practice she has traveled many kilometers and spent years in China, before arriving at the Master in Culture of Italian Wine in Pollenzo.		1		1					1	1
"Ho studiato Lingue orientali a Venezia, quindi ho proseguito con una laurea magistrale in economia e storia della Cina.	"I studied Oriental Languages in Venice, then I continued with a master's degree in economics and history of China.	1									
La Cina è sempre stata il mio sogno e terminati gli studi sono partita e vi sono rimasta circa due anni, curando i rapporti per aziende italiane in loco.	China has always been my dream and after completing my studies I left and stayed there for about two years, taking care of the relationships for Italian companies on site.		1	1							

ST	MT output	No error	Accuracy	Addition	Omission	Mistranslation	Untranslated	Fluency	Grammar	Style	Awkward
Non mi sono fermata, però, e rientrata in Italia ho lavorato a Genova nel settore dell'import-export" ci racconta Roberta, nella pausa pranzo del suo tirocinio presso Gaja a Barbaresco.	I didn't stop, however, and when I returned to Italy I worked in Genoa in the import-export sector "Roberta tells us, during the lunch break of her internship at Gaja in Barbaresco.		1			1				1	1
Il vino, però rimane un chiodo fisso per Roberta, e quindi la sua esperienza nel campo enologico continua:	Wine, however, remains an obsession for Roberta, and therefore her experience in the oenological field continues:	1									
lei vuole saperne di più, inizia a studiare, a praticare il mondo del vino, sia per lavoro che per curiosità.	she wants to know more, she begins to study, to practice the world of wine, both for work and for curiosity.		1		1						
"Ho quindi lavorato principalmente come interprete presso cantine in Langa, cercando di approfondire la mia conoscenza dei vini.	"I therefore worked mainly as an interpreter at wineries in the Langa , trying to deepen my knowledge of wines.		1				1				
Durante i miei soggiorni in Cina ho potuto capire che, ancora per molti, il vino rappresenta uno status symbol, qualcosa per ricchi.	During my stays in China I was able to understand that, still for many, wine represents a status symbol, something for the rich.	1									
Con il passare del tempo però, il consumo di vino si sta allargando non solo nelle città principali, ma anche in quelle cosiddette di seconda fascia, che sono comunque sempre delle	With the passage of time , however, the consumption of wine is expanding not only in the main cities, but also in the so-called second-tier ones, which are always huge metropolises:										

ST	MT output	No error	Accuracy	Addition	Omission	Mistranslation	Untranslated	Fluency	Grammar	Style	Awkward
metropoli enormi:											
si tratta quindi di un settore emergente che ha un grande futuro – continua Roberta.	it is therefore an emerging sector that has a great future - Roberta continues.	1									
“Dunque ormai nuovamente stabilitami in Italia mi sono attivata per trovare una formazione adeguata:	So by now, having settled back in Italy, I took steps to find adequate training:	1									
da anni seguivo l’attività dell’Università di Pollenzo:	for years I had been following the activities of the University of Pollenzo:	1									
due anni fa ho partecipato ad una Giornata Aperta, ma il Master non era ancora partito.	two years ago I participated in an Open Day, but the Master had not yet started.	1									
Sono rimasta in contatto con l’UNISG e quando ho scoperto che il corso era stato attivato, mi sono iscritta”.	I kept in touch with UNISG and when I found out that the course had been activated, I signed up .	1									
Roberta modestamente ammette di non aver ancora imparato tutto sul vino:	Roberta modestly admits that she has not yet learned everything about wine:	1									
“Il mondo del vino è molto variegato, c’è sempre da imparare e in un anno non si può apprendere tutto.	"The world of wine is very varied, there is always something to learn and you can't learn everything in a year.	1									

ST	MT output	No error	Accuracy	Addition	Omission	Mistranslation	Untranslated	Fluency	Grammar	Style	Awkward
Però il Master è una buona base dalla quale partire per ampliare le proprie conoscenze.	However, the Master is a good basis from which to start to expand one's knowledge.		1	1				1	1		
“Mi occupo delle visite in cantina, che avvengono su prenotazione.	“I take care of the visits to the cellar, which take place by appointment.	1									
La visita presso Gaja è un momento di grande formazione, per me è una sfida perché mi sto impegnando al massimo per essere all'altezza.	The visit to Gaja is a time of great training , for me it is a challenge because I am doing my utmost to be up to it.		1			1					
Inoltre essere qui è una straordinaria opportunità di crescita per me:	Also being here is an extraordinary growth opportunity for me:							1	1		
sto imparando moltissimo ogni giorno”.	I am learning a lot every day”.	1									
Ci viene spontaneo chiederle se ha già fatto visite in cinese.	We naturally ask you if you have already made visits in Chinese.		3			3					
“Non ancora – risponde Roberta – ad oggi effettuo le visite in inglese.	“Not yet - Roberta replies - to date I carry out the visits in English.	1									
La Cina rimane però sempre nei miei obiettivi.	However, China always remains in my goals.	1									
Oggi purtroppo è ancora la Francia che gioca un ruolo principale sul mercato cinese:	Today, unfortunately, it is still France that plays a main role on the Chinese market:	1									
se pensano al vino, i cinesi pensano a quello francese.	if they think of wine, the Chinese think of the French one.	1									
Vorrei provare a cambiare questa abitudine:	I would like to try to change this habit:									1	1

ST	MT output	No error	Accuracy	Addition	Omission	Mistranslation	Untranslated	Fluency	Grammar	Style	Awkward
mi piacerebbe che in Cina il binomio vino e Italia fosse vincente”.	I would like the combination of wine and Italy to be successful in China ”.	1									

Eleven Project

Κείμενο Αφιετηρίας

11 novembre 2011: unisciti e condividi la tua vita con il mondo

Cosa starai facendo l'11 novembre del 2011?

Condividilo con il resto del mondo tramite l'11Eleven Project [come tutti gli altri link, eccetto ove diversamente indicato] (Progetto 11Undici), aiutando così le Nazioni Unite a conseguire gli Obiettivi di Sviluppo del Millennio.

Raccontaci qualcosa di te, della tua vita e del mondo che ti circonda.

La tua storia per noi è importante, perciò ti chiediamo di partecipare al progetto.

Dopo l' 11.11.2011, ti verrà chiesto di presentare una foto, un video o un file audio che entreranno a far parte di tre progetti speciali: un libro fotografico, una compilation musicale e un film documentario.

I fondi raccolti grazie ai 3 progetti andranno a sostenere le associazioni benefiche che si impegnano ad aiutare l'umanità contribuendo al raggiungimento degli 8 Obiettivi di Sviluppo del Millennio delle Nazioni Unite.

11-11-11 Menti Creative Unite - 3 mesi alla partenza!

Video di 11Eleven Project su Youtube

A partire dall'11 agosto ci si può iscrivere e registrare online per partecipare l'11 novembre documentando un pezzetto della tua vita.

Basta avere una video, foto o della musica, si può diventare 'Ambasciatore' diffondendo il messaggio nella propria città, dando così voce alla propria comunità.

Puoi seguire 11Eleven Project su Facebook e Twitter per rimanere aggiornato sulle sue attività.

Poi, l'11 novembre 2011 (11/11/11) accendi il tuo computer, cellulare, macchina fotografica o registratore e documenta, in qualsiasi lingua, una parte significativa della tua giornata.

Dopo l'11 novembre, inoltraci video, fotografie o musica e il team di 11Eleven Project provvederà a montarne il contenuto per trasformarlo in un film.

Il 21 Settembre 2012, il video documentario verrà proiettato nel corso della Giornata Internazionale dell'ONU per la Pace.

Global Voices sta collaborando con 11Eleven Project per far crescere la consapevolezza riguardo ai Millenium Development Goals (MDGs) [Obiettivi di Sviluppo del Millennio dell'ONU].

A novembre, esattamente l'11/11/11, invitiamo i cittadini del mondo ad aiutarci a creare un ritratto condiviso dell'umanità con foto, video, audio e post che raccontino questa giornata.

Non serve far riferimento agli MDG, l'importante è che TI rappresenti .

Per documentarti meglio sulle 8 sfide per lo sviluppo e le misure adottate per conseguirle entro il 2015, consulta la nostra Sezione Speciale.

Incoraggiamo infine tutti a scrivere un post sui propri blog sul tema “Il mondo che vorremmo avere tra 100 anni”, enfatizzando l'importanza degli MDG.

Linka al tuo post la nostra pagina Millenium Development Goals e noi lo inseriremo in una lista di post dello Speciale.

Unisciti a noi, e l'11/11/11 celebriamo assieme l'umanità!

Κείμενο Αναφοράς

Κόσμος: Μοιράσου τη ζωή σου με τον κόσμο στις 11-11-11

Τι θα κάνετε στις 11 Νοεμβρίου 2011;

Μοιραστείτε το με τον υπόλοιπο κόσμο μέσω του πρότζεκτ 11Eleven Project και βοηθήστε στην επίτευξη των Αναπτυξιακών Στόχων της Χιλιετίας των Ηνωμένων Εθνών!

Πείτε μας κάτι για τη ζωή σας και το κοινωνικό σας περιβάλλον.

Η ιστορία σας είναι σημαντική για εμάς, εγγραφείτε και λάβετε μέρος στο πρότζεκτ 11.11.2011.

Μετά τις 11.11.2011 - Θα σας ζητηθεί να υποβάλετε είτε φωτογραφίες, ταινία ή ηχογραφημένο κείμενο online, όπου εμείς θα το μετατρέψουμε σε τριών ειδών πρότζεκτ -σε ένα φωτογραφικό βιβλίο, σε μια συλλογή μουσικής και σε ένα ντοκιμαντέρ.

Τα χρήματα που θα συγκεντρωθούν από τα έσοδα αυτών των τριών έργων θα διατεθούν για τη στήριξη Φιλανθρωπικών Ιδρυμάτων, που στηρίζουν την ανθρωπότητα βοηθώντας στην υλοποίηση των 8 Αναπτυξιακών Στόχων των Ηνωμένων Εθνών για τη Χιλιετία.

11-11-11 Δημιουργικά μυαλά ενωμένα - 3 μήνες έμειναν!

Βίντεο του 11Eleven Project στο Youtube

Από τις 11 Αυγούστου μπορείτε να εγγραφείτε και να συμμετέχετε. Έτσι σας δίνετε η δυνατότητα να καταγράψετε ένα κομμάτι της ζωής σας στις 11 Νοεμβρίου, είτε μέσω ταινίας, φωτογραφίας ή μουσικής.

Μπορείτε επίσης να εγγραφείτε ως «Πρεσβευτής», προσπαθώντας να γνωστοποιήσετε το πρόγραμμα στην πόλη σας και να διασφαλίσετε ότι οι «φωνές» της κοινότητάς σας θα εισακουστούν και θα γίνουν γνωστές.

Ακολουθήστε το πρότζεκτ «11 Eleven» στο Facebook και το Twitter για να είστε ενήμεροι για τις δραστηριότητές του προγράμματος.

Εν συνεχεία, στις 11 Νοεμβρίου 2011, (11/11/11), χρησιμοποιήστε είτε υπολογιστή, κινητό τηλέφωνο, φωτογραφική μηχανή ή συσκευή εγγραφής ήχου για να καταγράψετε ένα σημαντικό γεγονός αυτής της ημέρας σε οποιαδήποτε γλώσσα.

Μετά τις 11 Νοεμβρίου υποβάλετε το βίντεο, τις φωτογραφίες ή τη μουσική σας και η ομάδα του πρότζεκτ «11Eleven» θα επεξεργαστεί το περιεχόμενο σε μια ταινία,

και στις 21 Σεπτεμβρίου 2012, αυτό το ντοκιμαντέρ θα προβληθεί στη παγκόσμια ημέρα ειρήνης των Ηνωμένων Εθνών.

Η ιστοσελίδα του Global Voices συνεργάζεται με το πρότζεκτ «11Eleven» των Ηνωμένων Εθνών για να ευαισθητοποιήσει σχετικά με τους στόχους της Χιλιετίας για την Ανάπτυξη (ΑΣΧ) « Millenium Development Goals (MDGs)».

Τον Νοέμβριο, στις 11/11/11 καλούμε όλο τον κόσμο να συμβάλει στη δημιουργία ενός πορτρέτου της ανθρωπότητας σε συνεργασία με τους πολίτες: όπου θα εμπεριέχονται φωτογραφίες, ταινίες, ήχο ή δημοσιεύσεις σε blog για τη συγκεκριμένη ημέρα.

Δεν επιβάλλεται να σχετίζονται με τους ΑΣΧ Millenium Development Goals (MDGs), το σημαντικό είναι να εκπροσωπεί ΕΣΑΣ.

Μπορείτε να διαβάσετε περισσότερα για τις 8 αναπτυξιακές πολιτικές και τα μέτρα που λαμβάνονται για την επίλυσή τους ως το 2015 στην ιστοσελίδα με το ειδικό αφιέρωμα..

Επίσης ενθαρρύνουμε όσους ενδιαφέρονται να γράψουν σε blog για το “Πώς θα ήθελαν να είναι ο κόσμος σε 100 χρόνια”, σε σχέση με τους στόχους της Χιλιετίας.

Συνδεθείτε με την ιστοσελίδα μας «Millenium Development Goals» , όπου θα βρείτε εκεί αναρτημένη τη δημοσίευσή σας.

Λάβετε μέρος για να γιορτάσουμε τη μέρα που είναι αφιερωμένη στην ανθρωπότητα, 11/11/11, όλοι μαζί.

Facebook Translate – Αξιολόγηση ακρίβειας και ευχέρειας

ST	MT Output	Adequacy	Fluency
es ist total deprimierend , zu sehen , dass \$ mein # Mexico den # Worldcup verlassen hat , aber niemand kann leugnen , sie haben gekämpft wie die Löwen ! ! ! ! ! # VamosMexico	it's totally depressing to see that \$ my #Mexico has left the #Worldcup, but nobody can deny, they fought like the lions! ! #VamosMexico	Good	Excellent
verdammte Scheiße # Brazil , ich weiß , ihr habt vergessen , einige Stadien fertig zubauen , aber habt ihr auch vergessen \$ euer Team zu trainieren ? # Worldcup # BrazilvsMexico	Damn shit #Brazil I know y'all forgot to finish some stadiums but did y'all forget to train \$ your team too? #Worldcup #BrazilvsMexico	Good	Excellent
@ themoonexpress Lass mich raten ... du bist kein Fan des südamerikanischen Tanzes ? # sarcasm # Worldcup	@ themoon express Let me guess... you're not a fan of South American dance? #sarcasm #worldcup	Excellent	Fair
ich brauche eine Wiederholung mit Torlinientechnik für dieses # robben Tor , Ich bin nicht sicher , ob er drin war # Worldcup	i need a repeat with goal line tech for this #robben goal, not sure if he was in it #Worldcup	Poor	Excellent
kein Abseits von Higuain ! ☹ der verdammte Schiedsrichter hat es verdorben ! # Worldcup # ARG	no offside Higuain! ☹ the damn ref spoiled it! #Worldcup #ARG	Excellent	Excellent
ich denke \$ es ist zu früh für schlechten Ballismus ... # Worldcup	i think \$ it's too early for bad ballism... #Worldcup	Excellent	Excellent
der einzige Grund , warum ich mich nach diesem langweiligen Spiel nicht umgebracht habe , ist das # USA in einer Stunde spielt . # Worldcup	the only reason I didn't kill myself after this boring game is the #USA plays in an hour. #Worldcup	Good	Fair
bringt einen Gedanken für die Königin Maxima der Niederlande auf , da ich argentinisch bin , sage ich , es kann sein , dass sie nach dem Halbfinale keine Königin ist ! : wooooo Orange ! Weltmeisterschaft	brings up a thought for Queen Maxima of Netherlands, since I'm Argentinian, I say she may not be a queen after the semi finals! : wooooo orange! World Cup	Poor	Good
nach einer 5-1 Niederlage gegen die Holländer , muss sich Spanien so fühlen http : // t.co / urc1aV7uJ9 # Worldcup	After a 5-1 defeat to the Dutch, Spain must feel this way http: // t. co / urc1aV 7uJ 9 #Worldcup	Excellent	Excellent

ST	MT Output	Adequacy	Fluency
und die Geschichte wiederholt sich mit den Schweizern , sie entscheiden sich , sich nicht zu wehren .. Zing # Worldcup # SUIvsFRA	and history repeats itself with the Swiss, they choose not to fight back.. Zing #Worldcup #SUIvsFRA	Excellent	Excellent
es ist so seltsam zu sehen , dass # MEX rot und schwarz als \$ ihre alternativen Farben tragen . # Worldcup	it's so weird to see #MEX wearing red and black as \$ their alternative colors. #Worldcup	Excellent	Excellent
Frankreich verschießt einen Elfmeter , trotzdem noch 2-0 für Frankreich # Worldcup	France shoots a penalty, but still 2-0 for France #Worldcup	Very poor	Excellent
wir glauben , \$ unser # Belgium Chef ist ein wenig verwirrt darüber , bei wem er mitfiebern soll in dem # BELvsUSA Spiel . # Worldcup http : / / t.co / zm0sVFGJXb	we think \$ our #Belgium boss is a little confused about who to cheer on in the #BelvsUSA game. #Worldcup http: / / t. co / zm0sVFGJXb	Excellent	Excellent
# Siri besitzt einige Frechheit . ☹ . # pummeled # brazil # germany # Worldcup http : / / t.co / EGjMM6f9u	#Siri has some sassiness. ☹. ☹ #chummed #brazil #germany #Worldcup http: / / t. co / EGjM 6fF 9u	Excellent	Excellent
erste Wasserpause in der Geschichte der Weltmeisterschaft ... Nicht sicher , wie ich darüber denke . # Worldcup # USAvsPOR	first water break in the history of the World Cup... Not sure how I feel about it. #Worldcup #USAvsPOR	Good	Excellent
# Worldcup | Ghana leidet unter ihrer 2 - 1 Niederlage gegen die USA : Ghanas schwarze Stars verloren \$ ihr erstes Spiel der Weltmeisterschaft 2014 : ... http : / / t.co / GLVO5G8wBE	#Worldcup &# 124; Ghana suffers from her 2-1 defeat to the USA: Ghana's black stars lost \$ her first game of the 2014 World Cup :... http: / / t. co / GLVO5G 8wBE	Fair	Fair
tut mir leid Brasilien veranstalte nächstes Mal keine Weltmeisterschaft # Worldcup Host # SSFootball	sorry Brazil.... not hosting a world cup next time #Worldcuphost #SSFootball	Poor	Good
ich hätte nie gedacht , dass Spanien jemals in einem Spiel so unüberzeugend \$ aussehen würden # Worldcup # ESP	i never thought Spain would ever look so unsurpassing \$ in a game #Worldcup #ESP	Very poor	Excellent
Spanien vs Niederlande 1-5 # WTF # omg # Worldcup lol	Spain vs Netherlands 1-5 #WTF #omg #Worldcup lol	Excellent	Excellent
kann ich dich bitte mit einem Klumpen Cheddar-Käse bekanntmachen ! los England ! # Worldcup	can i please get you acquainted with a lump of cheddar cheese! Let's go England! #Worldcup	Good	Excellent

ST	MT Output	Adequacy	Fluency
OK eine kleine Unaufmerksamkeit : - mal sehen , wie tor-hungrig das Team sein wird : - es geht nur darum , wie sehr man es will # Worldcup	OK a little inattention: - let's see how goal-hungry the team will be: - it's all about how bad you want it #Worldcup	Excellent	Excellent
Herzinfarkt ist die einzige Option # Worldcup	Heart attack is the only option #Worldcup	Excellent	Excellent
unnütz ! England nutzlos ! # eng # Worldcup	useless! England useless! #tight #worldcup	Excellent	Excellent
unglückliches Brasilien , es war Klose	unlucky Brazil it was Klose	Excellent	Good
warum hat der Schiedsrichter aufgehört dort zuspiesen ? es ist keine Kopfverletzung . es lässt sich darüber streiten , ob das überhaupt eine Verletzung war ! # Worldcup	why did the ref stop playing there? it's not a head injury. it can be argued about whether this was even an injury! #Worldcup	Good	Excellent
spanisches wurde das Tor nicht gegeben ; Abseits # Worldcup # sucks # ESPNED	The goal wasn't given in Spanish; offside #Worldcup #sucks #ESPNED	Fair	Good
wenn man in diesem # Worldcup nur zwei Tore macht , sollte man wahrscheinlich damit rechnen bald nach Hause zu fahren . # ITAURU	If you only score two goals in this #Worldcup, you should probably expect to drive home soon. #ITAURU	Good	Excellent
warte was ? nicht , wenn er in den Ruhestand gehen muss # Gerrard # Worldcup	Wait for what? not when he needs to retire #Gerrard #Worldcup	Poor	Excellent
verdammte # Worldcup hier und # Worldcup da. hab es jetzt schon satt	damn #worldcup here and #worldcup there. tired of it already	Excellent	Excellent
zum ersten Mal hat mich das zur Toilette gehen gekriegt und ich habe das Tor verpasst ! # Worldcup # England	for the first time this got me going to the bathroom and I missed the gate! #Worldcup #England	Very poor	Excellent
USA Man of the Match ~ NO ONE ! # scmotm # WorldCupProblems # Worldcup # ESPN	USA Man of the Match ~ NO ONE ! # scmotm # WorldCupProblems # Worldcup # ESPN	Excellent	Excellent
Messis Eilmeldung an den verletzten Neymar http : / / t.co / QlAKzDb5e	Messi's breaking message to injured Neymar http: / / t. co / QlAKzDb5e	Poor	Excellent

ST	MT Output	Adequacy	Fluency
lol , ich habe dieses Spiel nur geguckt , um mich über das Team , das verliert , lustig zu machen.Ich war darüber noch nie so begeistert , TBH . ☹️ # Worldcup	lol I just watched this game to make fun of the team losing. Never been so excited about this TBH. ☹️ #Worldcup	Poor	Good
# WK2014 : van Persie : die Niederländer müssen eine harte Zeit erwarten http : // t.co / gREqhKADkY	#WK2014: van Persie: Dutch must expect a hard time http: // t. co / gREqhKADKY	Excellent	Good
süßes Italien hat gerade einen kleines Zittern # avicii # Worldcup	cute Italy just got a little tremor #avicii #worldcup	Excellent	Excellent
Howard Webb ist ein schrecklicher Schiri # Worldcup	Howard Webb is a terrible ref #Worldcup	Excellent	Excellent
darf ich der erste sein der sagt , Totalschaden für Spanien ! # Spainvschile # Worldcup	May I be the first to say total damage to Spain! #Spainvschile #Worldcup	Fair	Good
die Leute sagen , es ist sinnlos das @ england Spiel morgen zu gucken . ja , sie sind raus aus dem # Worldcup aber ich warte immer noch darauf	people say it's pointless to watch the @ england game tomorrow. yes they are out of the #worldcup but I'm still waiting for it	Excellent	Excellent
@ MiseryBear Armer kleine Bär . vielleicht könnten du und @ officialkylie einander helfen , das Elend zu bedauern # Worldcup	@ MiseryBear poor little bear. maybe you and @ official kylie could help each other regret the misery #Worldcup	Fair	Excellent
sind Brasilien und England verkleidet ? # Worldcup	are brazil and england disguised? #Worldcup	Excellent	Excellent

Google Translate – Αξιολόγηση ακρίβειας και ευχέρειας

ST	MT Output	Adequacy	Fluency
es ist total deprimierend , zu sehen , dass \$ mein # Mexico den # Worldcup verlassen hat , aber niemand kann leugnen , sie haben gekämpft wie die Löwen ! ! ! ! ! # VamosMexico	it's totally depressing to see that \$ my # Mexico has left the # World Cup, but no one can deny that they fought like the lions! ! ! ! ! # VamosMexico	Good	Excellent
verdammte Scheiße # Brazil , ich weiß , ihr habt vergessen , einige Stadien fertig zubauen , aber habt ihr auch vergessen \$ euer Team zu trainieren ? # Worldcup # BrazilvsMexico	Damn shit # Brazil, I know you forgot to finish building some stadiums, but did you forget \$ to train your team too? # Worldcup # BrazilvsMexico	Excellent	Excellent
@ themoonexpress Lass mich raten ... du bist kein Fan des südamerikanischen Tanzes ? # sarcasm # Worldcup	@ themoonexpress Let me guess ... you are not a fan of the South American dance? # sarcasm # Worldcup	Excellent	Fair
ich brauche eine Wiederholung mit Torlinientechnik für dieses # robben Tor , Ich bin nicht sicher , ob er drin war # Worldcup	I need a goal-line repetition for this # robben goal, I'm not sure if it was # Worldcup	Poor	Excellent
kein Abseits von Higuain ! ☹ der verdammte Schiedsrichter hat es verdorben ! # Worldcup # ARG	no apart from Higuain! ☹ the damn referee spoiled it! # Worldcup # ARG	Very poor	Good
ich denke \$ es ist zu früh für schlechten Ballismus ... # Worldcup	I think \$ it's too early for bad ballism ... # Worldcup	Excellent	Excellent
der einzige Grund , warum ich mich nach diesem langweiligen Spiel nicht umgebracht habe , ist das # USA in einer Stunde spielt . # Worldcup	the only reason I didn't kill myself after this boring game is because #USA is playing in an hour. # Worldcup	Excellent	Excellent
bringt einen Gedanken für die Königin Maxima der Niederlande auf , da ich argentinisch bin , sage ich , es kann sein , dass sie nach dem Halbfinale keine Königin ist ! : wooooo Orange ! Weltmeisterschaft	brings up a thought for Queen Maxima of the Netherlands, since I'm Argentine, I say, it may be that she won't be a queen after the semifinals! : wooooo orange! World Championship	Poor	Good
nach einer 5-1 Niederlage gegen die Holländer , muss sich Spanien so fühlen http : // t.co / urc1aV7uJ9 # Worldcup	after a 5-1 loss to the Dutch, Spain must feel this way http: // t.co / urc1aV7uJ9 # Worldcup	Excellent	Excellent

ST	MT Output	Adequacy	Fluency
und die Geschichte wiederholt sich mit den Schweizern , sie entscheiden sich , sich nicht zu wehren .. Zing # Worldcup # SUIvsFRA	and history repeats itself with the Swiss, they decide not to fight back .. Zing # Worldcup # SUIvsFRA	Excellent	Excellent
es ist so seltsam zu sehen , dass # MEX rot und schwarz als \$ ihre alternativen Farben tragen . # Worldcup	it's so strange to see # MEX wearing red and black as \$ their alternate colors. # Worldcup	Excellent	Excellent
Frankreich verschießt einen Elfmeter , trotzdem noch 2-0 für Frankreich # Worldcup	France misses a penalty, but still 2-0 for France #Worldcup	Excellent	Excellent
wir glauben , \$ unser # Belgium Chef ist ein wenig verwirrt darüber , bei wem er mitfiebern soll in dem # BELvsUSA Spiel . # Worldcup http : / / t.co / zm0sVFGJXb	we think \$ our # Belgium boss is a little confused about who to cheer for in the # BELvsUSA game. # Worldcup http: / / t.co / zm0sVFGJXb	Good	Excellent
# Siri besitzt einige Frechheit . ☹ . # pummeled # brazil # germany # Worldcup http : / / t.co / EGjMM6fF9u	# Siri has some cheek. ☹. # pummeled # brazil # germany # Worldcup http: / / t.co / EGjMM6fF9u	Excellent	Excellent
erste Wasserpause in der Geschichte der Weltmeisterschaft ... Nicht sicher , wie ich darüber denke . # Worldcup # USAvsPOR	first water break in World Cup history ... not sure how I think about that. # Worldcup # USAvsPOR	Excellent	Excellent
# Worldcup & #124; Ghana leidet unter ihrer 2 - 1 Niederlage gegen die USA : Ghanas schwarze Stars verloren \$ ihr erstes Spiel der Weltmeisterschaft 2014 : ... http : / / t.co / GLVO5G8wBE	# Worldcup & # 124; Ghana suffers from their 2-1 loss to the USA: Ghana's black stars lost \$ their first game of the 2014 World Cup: ... http: / / t.co / GLVO5G8wBE	Fair	Excellent
tut mir leid Brasilien veranstalte nächstes Mal keine Weltmeisterschaft # Worldcup Host # SSFootball	I'm sorry Brazil don't host a World Cup # Worldcup Host # SSFootball next time	Excellent	Excellent
ich hätte nie gedacht , dass Spanien jemals in einem Spiel so unüberzeugend \$ aussehen würden # Worldcup # ESP	I never thought that Spain would ever look so unconvincing in a game \$ # Worldcup # ESP	Excellent	Excellent
Spanien vs Niederlande 1-5 # WTF # omg # Worldcup lol	Spain vs Netherlands 1-5 # WTF # omg # Worldcup lol	Excellent	Excellent
kann ich dich bitte mit einem Klumpen Cheddar-Käse bekanntmachen ! los England ! # Worldcup	can I introduce you to a lump of cheddar cheese, please! go England! # Worldcup	Excellent	Excellent

ST	MT Output	Adequacy	Fluency
OK eine kleine Unaufmerksamkeit : - mal sehen , wie torhungrig das Team sein wird : - es geht nur darum , wie sehr man es will # Worldcup	OK a little inattention: - let's see how goal-hungry the team will be: - it's all about how much you want it # Worldcup	Excellent	Excellent
Herzinfarkt ist die einzige Option # Worldcup	Heart attack is the only option # Worldcup	Excellent	Excellent
unnütz ! England nutzlos ! # eng # Worldcup	useless! England useless! # eng # Worldcup	Excellent	Excellent
unglückliches Brasilien , es war Klose	unhappy Brazil, it was Klose	Excellent	Excellent
warum hat der Schiedsrichter aufgehört dort zupspielen ? es ist keine Kopfverletzung . es lässt sich darüber streiten , ob das überhaupt eine Verletzung war ! # Worldcup	why did the referee stop playing there? it's not a head injury. it is debatable whether that was an injury at all! # Worldcup	Excellent	Excellent
spanisches wurde das Tor nicht gegeben ; Abseits # Worldcup # sucks # ESPNED	Spanish was not given the gate; Offside # Worldcup # sucks # ESPNED	Very poor	Good
wenn man in diesem # Worldcup nur zwei Tore macht , sollte man wahrscheinlich damit rechnen bald nach Hause zu fahren . # ITAURU	if you only score two goals in this #Worldcup, you should probably expect to go home soon. # ITAURU	Excellent	Excellent
warte was ? nicht , wenn er in den Ruhestand gehen muss # Gerrard # Worldcup	wait what not if he has to retire # Gerrard # Worldcup	Poor	Poor
verdammt # Worldcup hier und # Worldcup da. hab es jetzt schon satt	damn #worldcup here and #worldcup there. already tired of it	Excellent	Excellent
zum ersten Mal hat mich das zur Toilette gehen gekriegt und ich habe das Tor verpasst ! # Worldcup # England	for the first time this got me to the bathroom and I missed the gate! # Worldcup # England	Very poor	Excellent
USA Man of the Match ~ NO ONE ! # scmotm # WorldCupProblems # Worldcup # ESPN	USA Man of the Match ~ NO ONE! # scmotm # WorldCupProblems # Worldcup # ESPN	Excellent	Excellent
Messis Eilmnachricht an den verletzten Neymar http : / / t.co / QILAKzDb5e	Messi's urgent message to the injured Neymar http: / / t.co / QILAKzDb5e	Poor	Excellent

ST	MT Output	Adequacy	Fluency
lol , ich habe dieses Spiel nur geguckt , um mich über das Team , das verliert , lustig zu machen.Ich war darüber noch nie so begeistert , TBH . ☹️ # Worldcup	lol, I just watched this game to make fun of the team that was losing, I've never been so excited about it, TBH. ☹️ # Worldcup	Poor	Excellent
# WK2014 : van Persie : die Niederländer müssen eine harte Zeit erwarten http : / / t.co / gREqhKADkY	# WK2014: van Persie: the Dutch have to expect a hard time http: / / t.co / gREqhKADkY	Excellent	Excellent
süßes Italien hat gerade einen kleines Zittern # avicii # Worldcup	sweet italy just got a little tremor # avicii # worldcup	Excellent	Excellent
Howard Webb ist ein schrecklicher Schiri # Worldcup	Howard Webb is a terrible referee #Worldcup	Good	Excellent
darf ich der erste sein der sagt , Totalschaden für Spanien ! # Spainvschile # Worldcup	may I be the first to say, total loss for Spain! # Spainvschile # Worldcup	Excellent	Excellent
die Leute sagen , es ist sinnlos das @ england Spiel morgen zu gucken . ja , sie sind raus aus dem # Worldcup aber ich warte immer noch darauf	people say there is no point in watching @ england tomorrow. yes, they are out of the #worldcup but i'm still waiting for it	Good	Excellent
@ MiseryBear Armer kleine Bär . vielleicht könnten du und @ officialkylie einander helfen , das Elend zu bedauern # Worldcup	@ MiseryBear Poor little bear. maybe you and @officialkylie could help each other regret the misery #Worldcup	Fair	Excellent
sind Brasilien und England verkleidet ? # Worldcup	are Brazil and England in disguise? # Worldcup	Excellent	Excellent

Facebook Translate - Κατηγοριοποίηση λαθών

ST	MT output	No error	Accuracy	Omission	Mistranslation	Untranslated	Fluency	Punctuation	Grammar	Style	Awkward
11 novembre 2011: unisciti e condividi la tua vita con il mondo	11 Νοεμβρίου 2011: Μπείτε και μοιραστείτε τη ζωή σας με τον κόσμο	1									
Cosa starai facendo l'11 novembre del 2011?	Τι κάνεις στις 11 Νοεμβρίου 2011;									1	1
Condividilo con il resto del mondo tramite l'11Eleven Project [en, come tutti gli altri link, eccetto ove diversamente indicato] (Progetto 11Undici), aiutando così le Nazioni Unite a conseguire gli Obiettivi di Sviluppo del Millennio.	Μοιραστείτε το με τον υπόλοιπο κόσμο μέσω του 11 Eleven Project, [όπως όλοι οι άλλοι σύνδεσμοι, εκτός εάν αναφέρεται διαφορετικά] (Project 11 Eleven), βοηθώντας έτσι τα Ηνωμένα Έθνη να επιτύχουν Αναπτυξιακούς Στόχους της Χιλιετίας.		1	1							
Raccontaci qualcosa di te, della tua vita e del mondo che ti circonda.	Πείτε μας κάτι για τον εαυτό σας, για τη ζωή σας και τον κόσμο γύρω σας.	1									
La tua storia per noi è importante, perciò ti chiediamo di partecipare al progetto.	Η ιστορία σας είναι σημαντική για εμάς, γι' αυτό σας ζητάμε να συμμετέχετε στο έργο.	1									

ST	MT output	No error	Accuracy	Omission	Mistranslation	Untranslated	Fluency	Punctuation	Grammar	Style	Awkward
Dopo l' 11.11.2011, ti verrà chiesto di presentare una foto, un video o un file audio che entreranno a far parte di tre progetti speciali: un libro fotografico, una compilation musicale e un film documentario.	Μετά το 11.11.2011, θα σας ζητηθεί να υποβάλετε μια φωτογραφία, βίντεο ή αρχείο ήχου που θα γίνει μέρος τριών ειδικών έργων: ένα φωτογραφικό βιβλίο, μια μουσική συλλογή και μια ταινία ντοκιμαντέρ.						2		2		
I fondi raccolti grazie ai 3 progetti andranno a sostenere le associazioni benefiche che si impegnano ad aiutare l'umanità contribuendo al raggiungimento degli 8 Obiettivi di Sviluppo del Millennio delle Nazioni Unite.	Τα κονδύλια που αντλήθηκαν από τα 3 έργα θα υποστηρίξουν φιλανθρωπίες που προσπαθούν να βοηθήσουν την ανθρωπότητα συμβάλλοντας στους 8 Αναπτυξιακούς Στόχους της Χιλιετίας των Ηνωμένων Εθνών.		2	2							
11-11-11 Menti Creative Unite - 3 mesi alla partenza!	11-11-11 Ενωμένα Δημιουργικά Μυαλά - 3 μήνες για αρχή!		1		1						
Video di 11Eleven Project su Youtube	Βίντεο του 11 Έντεκα Πρότζεκτ στο YouTube		1		1						

ST	MT output	No error	Accuracy	Omission	Mistranslation	Untranslated	Fluency	Punctuation	Grammar	Style	Awkward
A partire dall'11 agosto ci si può iscrivere e registrare online per partecipare l'11 novembre documentando un pezzetto della tua vita.	Ξεκινώντας από τις 11 Αυγούστου μπορείτε να γραφτείτε και να εγγραφείτε online για να συμμετέχετε στις 11 Νοεμβρίου, τεκμηριώνοντας ένα κομμάτι της ζωής σας.		2		2						
Basta avere un video, foto o della musica, si può diventare 'Ambasciatore' diffondendo il messaggio nella propria città, dando così voce alla propria comunità.	Αρκεί να έχεις βίντεο , φωτογραφία ή μουσική, μπορείς να γίνεις 'Ambassador' διαδίδοντας το μήνυμα στην πόλη σου , δίνοντας έτσι φωνή στην κοινότητά σου .		3	2		1	5		5		
Puoi seguire 11Eleven Project su Facebook e Twitter per rimanere aggiornato sulle sue attività.	Μπορείτε να ακολουθήσετε το 11 Eleven Project στο Facebook και στο Twitter για να ενημερώνεστε για τις δραστηριότητές τους .		1		1						
Poi, l'11 novembre 2011 (11/11/11) accendi il tuo computer, cellulare, macchina fotografica o registratore e documenta, in qualsiasi lingua, una parte significativa	Στη συνέχεια στις 11 Νοεμβρίου 2011 (11/11/11) ανοίξτε τον υπολογιστή, κινητό, κάμερα ή μαγνητόφωνο και τεκμηριώστε ένα σημαντικό μέρος της ημέρας σας σε οποιαδή-		2	1	1		3		3		

ST	MT output	No error	Accuracy	Omission	Mistranslation	Untranslated	Fluency	Punctuation	Grammar	Style	Awkward
tiva della tua giornata.	ποτε γλώσσα.										
Dopo l'11 novembre, inoltraci video, fotografie o musica e il team di 11Eleven Project provvederà a montarne il contenuto per trasformarlo in un film.	Μετά τον Νοέμβριο του 11, τα προωθητικά βίντεο, φωτογραφίες ή μουσική σε εμάς και η ομάδα του 11 Eleven Project θα τοποθετήσει το περιεχόμενο για να μεταμορφωθεί σε ταινία.				3						
Il 21 Settembre 2012, il video documentario verrà proiettato nel corso della Giornata Internazionale dell'ONU per la Pace.	Στις 21 Σεπτεμβρίου 2012, το βίντεο ντοκιμαντέρ θα προβληθεί κατά τη διάρκεια της Διεθνούς Ημέρας Ειρήνης του ΟΗΕ.	1									
Global Voices sta collaborando con 11Eleven Project per far crescere la consapevolezza riguardo ai Millennium Development Goals (MDGs) [Obiettivi di Sviluppo del Millennio dell'ONU].	Η Global Voices συνεργάζεται με το 11 Έντεκα Πρότζεκτ για την ευαισθητοποίηση για τους Αναπτυξιακούς Στόχους της Χιλιετίας (ΑΣΧ) [Αναπτυξιακοί Στόχοι της Χιλιετίας].		1		1						

ST	MT output	No error	Accuracy	Omission	Mistranslation	Untranslated	Fluency	Punctuation	Grammar	Style	Awkward
A novembre, esattamente l'11/11/11, invitiamo i cittadini del mondo ad aiutarci a creare un ritratto condiviso dell'umanità con foto, video, audio e post che raccontino questa giornata.	Τον Νοέμβριο, ακριβώς 11/11/11 , καλούμε τους πολίτες του κόσμου να μας βοηθήσουν να δημιουργήσουμε ένα κοινόχρηστο πορτρέτο της ανθρωπότητας με φωτογραφίες, βίντεο, ήχο και δημοσιεύσεις που μαρτυρούν σήμερα .		3	1	2						
Non serve far riferimento agli MDG, l'importante è che TI rappresenti .	Δεν χρειάζεται να αναφέρετε τα ΑΣΧ, το σημαντικό είναι ότι εκπροσωπείτε .		2	1	1						
Per documentarti meglio sulle 8 sfide per lo sviluppo e le misure adottate per conseguirle entro il 2015, consulta la nostra Sezione Speciale.	Για καλύτερη τεκμηρίωση των 8 αναπτυξιακών προκλήσεων και μέτρων που λαμβάνονται για την επίτευξή τους μέχρι το 2015, δείτε το Ειδ		2	1	1		1	1			
Incoraggiamo infine tutti a scrivere un post sui propri blog sul tema "Il mondo che vorremmo avere tra 100 anni", enfatizzando l'importanza degli MDG.	Τέλος, ενθαρρύνουμε όλους να γράψουν μια ανάρτηση στα blogs τους με θέμα "Ο κόσμος που θα θέλαμε να έχουμε σε 100 χρόνια ", δίνοντας έμφαση στη σημασία των ΑΣΧ.		1			1					

ST	MT output	No error	Accuracy	Omission	Mistranslation	Untranslated	Fluency	Punctuation	Grammar	Style	Awkward
Linka al tuo post la nostra pagina Mil-lenium Develop-ment Goals e noi lo inseriremo in una lista di post dello Speciale.	Σύνδεσμος για την ανάρτησή σας στη σελίδα Millenium Development Goals και θα την βάλουμε σε λίστα με Ειδικές Δημοσιεύσεις		1		1						
Unisciti a noi, e l'11/11/11 celebri-amo assieme l'u-manità!	Ελάτε μαζί μας, και στις 11/11/11 ας γιορτάσουμε μαζί την ανθρωπότητα!									1	1

Google Translate - Κατηγοριοποίηση λαθών

ST	MT output	No error	Accuracy	Addition	Omission	Mistranslation	Untranslated	Fluency	Grammar
11 novembre 2011: unisciti e condividi la tua vita con il mondo	11 Νοεμβρίου 2011: εγγραφείτε και μοιραστείτε τη ζωή σας με τον κόσμο	1							
Cosa starai facendo l'11 novembre del 2011?	Τι θα κάνετε στις 11 Νοεμβρίου 2011;	1							
Condividilo con il resto del mondo tramite l'11Eleven Project [en, come tutti gli altri link, eccetto ove diversamente indicato] (Progetto 11Undici), aiutando così le Nazioni Unite a conseguire gli Obiettivi di Sviluppo del Millennio.	Μοιραστείτε το με τον υπόλοιπο κόσμο μέσω του έργου 11 Eleven, [όπως όλοι οι άλλοι σύνδεσμοι, εκτός αν αναφέρεται διαφορετικά] (Project 11 Eleven), βοηθώντας έτσι τα Ηνωμένα Έθνη να επιτύχουν τους Αναπτυξιακούς Στόχους της Χιλιετίας.	1							
Raccontaci qualcosa di te, della tua vita e del mondo che ti circonda.	Πείτε μας κάτι για τον εαυτό σας, τη ζωή σας και τον κόσμο γύρω σας.	1							
La tua storia per noi è importante, perciò ti chiediamo di partecipare al progetto.	Η ιστορία σας είναι σημαντική για εμάς, επομένως σας ζητάμε να συμμετάσχετε στο έργο.	1							
Dopo l' 11.11.2011, ti verrà chiesto di presentare una foto, un video o un file audio che entreranno a far parte di tre progetti speciali: un libro fotografico, una compilation musicale e un film documentario.	Μετά την 11.11.2011, θα σας ζητηθεί να υποβάλετε ένα αρχείο φωτογραφίας, βίντεο ή ήχου που θα γίνει μέρος τριών ειδικών έργων: ένα άλμπουμ φωτογραφιών, μια συλλογή μουσικής και μια ταινία ντοκιμαντέρ.							1	1

ST	MT output	No error	Accuracy	Addition	Omission	Mistranslation	Untranslated	Fluency	Grammar
I fondi raccolti grazie ai 3 progetti andranno a sostenere le associazioni benefiche che si impegnano ad aiutare l'umanità contribuendo al raggiungimento degli 8 Obiettivi di Sviluppo del Millennio delle Nazioni Unite.	Τα κεφάλαια που συγκεντρώθηκαν χάρη στα 3 έργα θα διατεθούν για τη στήριξη φιλανθρωπικών ενώσεων που έχουν δεσμευτεί να βοηθήσουν την ανθρωπότητα συμβάλλοντας στην επίτευξη των 8 Αναπτυξιακών Στόχων της Χιλιετίας των Ηνωμένων Εθνών.	1							
11-11-11 Menti Creative Unite - 3 mesi alla partenza!	11-11-11 Creative Minds Unite - 3 μήνες ακόμη!		1			1			
Video di 11Eleven Project su Youtube	11 Έντεχνο βίντεο στο Youtube		1			1			
A partire dall'11 agosto ci si puoi iscrivere e registrare online per partecipare l'11 novembre documentando un pezzetto della tua vita.	Από τις 11 Αυγούστου μπορείτε να εγγραφείτε και να εγγραφείτε online για να συμμετάσχετε στις 11 Νοεμβρίου τεκμηριώνοντας ένα κομμάτι της ζωής σας.		2			2			
Basta avere un video, foto o della musica, si può diventare 'Ambasciatore' diffondendo il messaggio nella propria città, dando così voce alla propria comunità.	Το μόνο που χρειάζεστε είναι ένα βίντεο, φωτογραφία ή μουσική, μπορείτε να γίνετε «Πρέσβης» διαδίδοντας το μήνυμα στην πόλη σας, δίνοντας έτσι μια φωνή στην κοινότητά σας.		2	1	1				
Puoi seguire 11Eleven Project su Facebook e Twitter per rimanere aggiornato sulle sue attività.	Μπορείτε να παρακολουθήσετε το 11Eleven Project στο Facebook και το Twitter για να μείνετε ενημερωμένοι για τις δραστηριότητές του.		1			1			

ST	MT output	No error	Accuracy	Addition	Omission	Mistranslation	Untranslated	Fluency	Grammar
Poi, l'11 novembre 2011 (11/11/11) accendi il tuo computer, cellulare, macchina fotografica o registratore e documenti, in qualsiasi lingua, una parte significativa della tua giornata.	Στη συνέχεια, στις 11 Νοεμβρίου 2011 (11/11/11) ενεργοποιήστε τον υπολογιστή, το κινητό τηλέφωνο, τη φωτογραφική μηχανή ή τη συσκευή εγγραφής και το έγγραφο , σε οποιαδήποτε γλώσσα, σημαντικό μέρος της ημέρας σας.		3		2	1			
Dopo l'11 novembre, inoltraci video, fotografie o musica e il team di 11Eleven Project provvederà a montarne il contenuto per trasformarlo in un film.	Μετά τις 11 Νοεμβρίου, στείλτε μας βίντεο, φωτογραφίες ή μουσική και η ομάδα του 11Eleven Project θα επεξεργαστεί το περιεχόμενο για να το μετατρέψει σε ταινία.	1							
Il 21 Settembre 2012, il video documentario verrà proiettato nel corso della Giornata Internazionale dell'ONU per la Pace.	Στις 21 Σεπτεμβρίου 2012, το ντοκιμαντέρ θα προβληθεί κατά τη διάρκεια της Διεθνούς Ημέρας για την Ειρήνη των Ηνωμένων Εθνών.	1							
Global Voices sta collaborando con 11Eleven Project per far crescere la consapevolezza riguardo ai Millenium Development Goals (MDGs) [Obiettivi di Sviluppo del Millennio dell'ONU].	Η Global Voices συνεργάζεται με το 11Eleven Project για την ευαισθητοποίηση σχετικά με τους Αναπτυξιακούς Στόχους της Χιλιετίας (MDGs).		1				1		
A novembre, esattamente l'11/11/11, invitiamo i cittadini del mondo ad aiutarci a creare un ritratto condiviso dell'umanità con foto, video, audio e post che raccontino questa giornata.	Τον Νοέμβριο, ακριβώς 11/11/11, καλούμε τους πολίτες του κόσμου να μας βοηθήσουν να δημιουργήσουμε ένα κοινό πορτραίτο της ανθρωπότητας με φωτογραφίες, βίντεο, ήχο και δημοσιεύσεις που λένε για αυτήν την ημέρα.		2		1	1			

ST	MT output	No error	Accuracy	Addition	Omission	Mistranslation	Untranslated	Fluency	Grammar
Non serve far riferimento agli MDG, l'importante è che TI rappresenti .	Δεν χρειάζεται να αναφερθείτε στους ΑΣΧ, το σημαντικό είναι ότι εκπροσωπείτε ΕΣΑΣ.	1							
Per documentarti meglio sulle 8 sfide per lo sviluppo e le misure adottate per conseguirle entro il 2015, consulta la nostra Sezione Speciale.	Για να μάθετε περισσότερα σχετικά με τις 8 αναπτυξιακές προκλήσεις και τα μέτρα που έχουν ληφθεί για την επίτευξή τους έως το 2015, συμβουλευτείτε την Ειδική Ενότητα.	1							
Incoraggiamo infine tutti a scrivere un post sui propri blog sul tema "Il mondo che vorremmo avere tra 100 anni", enfatizzando l'importanza degli MDG.	Τέλος, ενθαρρύνουμε όλους να γράψουν μια ανάρτηση ιστολογίου με θέμα «Ο κόσμος που θέλουμε να είχαμε σε 100 χρόνια», τονίζοντας τη σημασία των ΑΣΧ.		1			1			
Linka al tuo post la nostra pagina Millenium Development Goals e noi lo inseriremo in una lista di post dello Speciale.	Συνδέστε τη σελίδα Millenium Development Goals με την ανάρτησή σας και θα την βάλουμε σε μια λίστα ειδικών δημοσιεύσεων.	1							
Unisciti a noi, e l'11/11/11 celebriamo assieme l'umanità!	Ελάτε μαζί μας και στις 11/11/11 γιορτάζουμε μαζί την ανθρωπότητα!		1			1			

To Twitter στην υπηρεσία της υγείας

Κείμενο Αφιετηρίας

Egitto: Twitter per sensibilizzare su temi medico-sanitari

Twitter può essere usato per salvare delle vite o migliorare la salute?

Durante gli scontri [come gli altri link eccetto ove diversamente indicato] dello scorso mese a piazza Tahrir, è stato creato l'account (@TahrirSupplies) per organizzare il rifornimento di provviste mediche, cibo e altri beni necessari agli ospedali da campo allestiti dai volontari nella piazza e dintorni.

Healthy Egypt Tips è un'altra iniziativa di un gruppo di volontari che ad agosto hanno aperto l'account (@HealthyEgypt) per sensibilizzare la popolazione sul tema della salute in generale.

Ecco un esempio dei messaggi diffusi:

@HealthyEgypt: Le ini-porzioni di caffè intervallate funzionano meglio di una grande tazza per tenervi svegli e concentrati.

Inoltre, il gruppo risponde alle domande della gente, sostiene altre iniziative e si occupa anche di educazione sessuale, in un Paese dove per ragioni culturali è spesso imbarazzante fare domande di questo tipo.

Chi siete e in cosa consiste la vostra iniziativa?

Siamo un gruppo di volontari, tutto è iniziato con me, Kareem Mossad, e mia moglie, Iman Ismail; poi si è unito un amico medico, Dr. Ahmed Ibrahim.

Nella propria descrizione, Healthy Egypt dice di condividere delle informazioni che sono al 100% credibili e verificabili.

Ci sentiamo quindi in dovere di chiedervi: siete dei dottori?

Come ottenete queste informazioni mediche, come le confermate, e come rispondete alle domande della gente?

Io e mia moglie siamo degli entusiasti attivisti sanitari.

Non siamo medici, ma date tempo abbiamo prestato servizio alla comunità.

Seguiamo un codice di condotta in base al quale ogni informazione medica che condividiamo con quanti ci seguono su Twitter viene doppiamente verificata, controllandola su diversi siti di medicina oppure verificandola con Ahmed Ibrahim, la terza persona che si occupa di quest'iniziativa, ed è un medico.

L'account è sponsorizzato da un ente commerciale o da una ONG?

Talvolta, ad esempio, fate riferimento a specifiche marche di prodotti.

Ne siete in qualche modo affiliati?

Non siamo ufficialmente sponsorizzati da nessun ente commerciale o ONG.

Tuttavia abbiamo ricevuto un paio di proposte, ma non siamo ancora pronti per questo passo.

Da quanto tempo operate online?

Prevedete una presenza oltre Twitter, ad esempio su Facebook, blog, ecc.?

Abbiamo iniziato ad agosto, verso la metà del Ramadan, e la risposta in numero di persone che seguono l'account è stata piuttosto impressionante.

Abbiamo acquisito i domini "healthyegypt.org e healthyegypt.net", e stiamo lavorando all'imminente lancio del sito.

I vostri tweet sono quasi sempre in inglese.

Non prevedete di scrivere in arabo?

Non pensate che l'uso dell'inglese finisca per limitare il pubblico di riferimento?

È assolutamente vero, talvolta scriviamo dei tweet in arabo, ma non in modo regolare.

Infatti il sito web sarà in entrambe le lingue per poter attrarre un maggior numero di visitatori.

Κείμενο Αναφοράς

Αίγυπτος: Χρησιμοποιώντας το Twitter για μια πιο υγιή ζωή

Μπορεί να χρησιμοποιηθεί το Twitter για να σώσει ανθρώπινες ζωές ή να βελτιώσει την υγεία τους;

Κατά τη διάρκεια των συμπλοκών που έλαβαν χώρα στην Πλατεία Ταχρίρ της Αιγύπτου τον προηγούμενο μήνα, δημιουργήθηκε ένας λογαριασμός στο Twitter (@TahrirSupplies) για την οργάνωση της μεταφοράς ιατρικών προμηθειών στα αυτοσχέδια νοσοκομεία που στήθηκαν από εθελοντές πάνω και γύρω από την πλατεία, καθώς και τροφίμων και άλλων αντικειμένων που χρειαζόνταν εκεί.

Από την άλλη πλευρά, μια διαφορετική πρωτοβουλία ήταν το Healthy Egypt Tips από ομάδα εθελοντών που ξεκίνησαν ένα λογαριασμό στο Twitter (@HealthyEgypt) τον Αύγουστο, με σκοπό να ευαισθητοποιήσουν τον κόσμο στην Αίγυπτο σχετικά με μικρά μυστικά για την υγεία.

Εδώ μια από τις συμβουλές τους:

@HealthyEgypt: Λίγες μικρές ποσότητες καφέ σε εργασιακά διαλείμματα έχουν καλύτερα αποτελέσματα απ' ό,τι μια μεγάλη κούπα, προκειμένου να μείνετε σε εγρήγορση και συγκεντρωμένοι.

Επίσης απαντούν σε ερωτήσεις του κόσμου, βοηθούν συμμετοχικές πρωτοβουλίες και διαδίδουν τη σεξουαλική διαπαιδαγώγηση, επίσης και όταν η κουλτούρα το θεωρεί ντροπή να γίνονται ερωτήσεις σεξουαλικού περιεχομένου.

Ποιοι είστε και πείτε μας περισσότερα για την πρωτοβουλία σας.

Είμαστε μια ομάδα εθελοντών, ξεκινήσαμε η σύζυγός μου, Iman Ismail, κι εγώ, ο Kareem Mossad, κι έπειτα προστέθηκε ο φίλος γιατρός Dr. Ahmed Ibrahim.

Γράφει στο βιογραφικό σας ότι μοιράζετε 100% αξιόπιστες και επαληθευμένες πληροφορίες.

Οπότε κι εμείς πρέπει να ρωτήσουμε: Είστε γιατροί;

Πώς λαμβάνετε κανονικά τις ιατρικές σας πληροφορίες, τις επιβεβαιώνετε και απαντάτε στις ερωτήσεις του κόσμου;

Η σύζυγός μου κι εγώ είμαστε ακτιβιστές και οπαδοί της υγείας.

Δεν είμαστε γιατροί, ωστόσο για κάποιο χρονικό διάστημα έχουμε προσφέρει τις υπηρεσίες μας στην κοινότητα.

Έχουμε ένα κώδικα συμπεριφοράς για κάθε ιατρική πληροφορία που παρέχουμε σε όσους μας ακολουθούν στο Twitter: πρέπει να την επιβεβαιώσουμε, ελέγχοντας παραπάνω από μια πηγές σε αξιόπιστες και γνωστές ιστοσελίδες υγείας ή να τις επαληθεύσουμε με τον

“Ahmed Ibrahim”, που είναι το τρίτο πρόσωπο (ιατρός) που βοηθά στην πρωτοβουλία αυτή.

Χρηματοδοτείται ο λογαριασμός από κάποια εμπορική επιχείρηση ή ΜΚΟ;

Παρατήρησα για παράδειγμα πως αναφέρεστε κάποιες φορές σε συγκεκριμένες μάρκες.

Σχετίζεστε με κάποιο τρόπο μαζί τους;

Δεν έχουμε επίσημα κανέναν χορηγό από εμπορική εταιρεία ή ΜΚΟ ακόμα.

Ωστόσο, δεχτήκαμε κάποιες προτάσεις, δεν ήμαστε όμως έτοιμοι για αυτό το βήμα ακόμη.

Πόσο καιρό ευαισθητοποιείτε τον κόσμο μέσω Διαδικτύου;

Έχετε σχέδια να διατηρήσετε την παρουσία σας και εκτός Twitter, όπως το Facebook, blogs κλπ.;

Ξεκινήσαμε τον Αύγουστο, περίπου στα μέσα του Ραμαζανιού, και η ανταπόκριση του κόσμου που ακολουθεί στο Twitter ήταν αρκετά εντυπωσιακή,

αγοράσαμε τα “www.healthyegypt.org και www.healthyegypt.net”, και επί του παρόντος είμαστε στη φάση προγραμματισμού της ιστοσελίδας μας, ώστε σύντομα να την ανεβάσουμε.

Επίσης παρατηρήσαμε ότι γράφετε σχεδόν όλα τα tweets στα αγγλικά.

Δε σκοπεύετε να γράφετε στα αραβικά;

Δεν πιστεύετε πως τα αγγλικά θα περιορίσουν το κοινό σας εντός Αιγύπτου κατά κάποιο τρόπο;

Έχετε απόλυτο δίκαιο, γράφουμε tweets στα αραβικά κάποιες φορές αλλά όχι τακτικά.

Είναι σίγουρα μες στα πλάνα μας με την ιστοσελίδα, η οποία θα είναι και στις δυο γλώσσες, ώστε να προσελκύσουμε περισσότερους επισκέπτες.

Facebook Translate – Αξιολόγηση ακρίβειας και ευχέρειας

ST	MT Output	Adequacy	Fluency
Egitto: Twitter per sensibilizzare su temi medico-sanitari	Αίγυπτος: Twitter για την ευαισθητοποίηση του κόσμου σε ιατρικά και θέματα υγείας	Excellent	Good
Twitter può essere usato per salvare delle vite o migliorare la salute?	Μπορεί το Twitter να χρησιμοποιηθεί για να σώσει ζωές ή να βελτιώσει την υγεία;	Excellent	Excellent
Durante gli scontri [come gli altri link eccetto ove diversamente indicato] dello scorso mese a piazza Tahrir, è stato creato l'account (@TahrirSupplies) per organizzare il rifornimento di provviste mediche, cibo e altri beni necessari agli ospedali da campo allestiti dai volontari nella piazza e dintorni.	Κατά τη διάρκεια των συμπλοκών [όπως και άλλοι σύνδεσμοι εκτός εάν αναφέρεται διαφορετικά] τον προηγούμενο μήνα στην πλατεία Ταχρίρ, δημιουργήθηκε ο λογαριασμός (@TahrirSupplies) για να οργανωθεί η προμήθεια ιατρικών προμηθειών, τροφίμων και άλλων αγαθών που απαιτούνται για τα νοσοκομεία της κατασκήνωσης από εθελοντές η πλατεία και η γύρω περιοχή.	Fair	Fair
Healthy Egypt Tips è un'altra iniziativa di un gruppo di volontari che ad agosto hanno aperto l'account (@HealthyEgypt) per sensibilizzare la popolazione sul tema della salute in generale.	To Healthy Egypt Tips είναι μια ακόμη πρωτοβουλία από ομάδα εθελοντών που άνοιξαν τον λογαριασμό τον Αύγουστο (@HealthyEgypt) για να ευαισθητοποιηθούν για το γενικό θέμα υγείας.	Good	Good
Ecco un esempio dei messaggi diffusi:	Εδώ είναι ένα παράδειγμα των ευρέως διαδεδομένων μηνυμάτων:	Excellent	Excellent
@HealthyEgypt: Le ini-porzioni di caffè intervallate funzionano meglio di una grande tazza per tenervi svegli e concentrati.	@HealthyEgypt: Διαλειμματικές μερίδες καφέ λειτουργούν καλύτερα από ένα μεγάλο φλιτζάνι για να είστε ξύπνιοι και συγκεντρωμένοι.	Fair	Excellent

ST	MT Output	Adequacy	Fluency
Inoltre, il gruppo risponde alle domande della gente, sostiene altre iniziative e si occupa anche di educazione sessuale, in un Paese dove per ragioni culturali è spesso imbarazzante fare domande di questo tipo.	Επίσης, η ομάδα απαντά στα ερωτήματα των ανθρώπων, υποστηρίζει άλλες πρωτοβουλίες και ασχολείται επίσης με την σεξουαλική εκπαίδευση, σε μια χώρα όπου για πολιτιστικούς λόγους είναι συχνά αμήχανο να κάνεις τέτοιες ερωτήσεις.	Good	Excellent
Chi siete e in cosa consiste la vostra iniziativa?	Ποιος είσαι και σε τι συνίσταται η πρωτοβουλία σου;	Good	Excellent
Siamo un gruppo di volontari, tutto è iniziato con me, Kareem Mossad, e mia moglie, Iman Ismail; poi si è unito un amico medico, Dr. Ahmed Ibrahim.	Είμαστε ομάδα εθελοντών, όλα ξεκίνησαν από εμένα, Kareem Mossad, και τη σύζυγό μου, Iman Ismail, μετά ιατρικός φίλος, Dr. Αχμέτ Ιμπραήμ.	Poor	Poor
Nella propria descrizione, Healthy Egypt dice di condividere delle informazioni che sono al 100% credibili e verificabili.	Στη δική της περιγραφή, η Healthy Egypt αναφέρει ότι μοιράζονται πληροφορίες που είναι 100 % αξιόπιστες και επαληθεύσιμες.	Excellent	Excellent
Ci sentiamo quindi in dovere di chiedervi: siete dei dottori?	Επομένως αισθανόμαστε υποχρεωμένοι να σας ρωτήσουμε: είστε γιατροί;	Excellent	Excellent
Come ottenete queste informazioni mediche, come le confermate, e come rispondete alle domande della gente?	Πώς παίρνετε αυτές τις ιατρικές πληροφορίες, πώς τις επιβεβαιώνετε και πώς απαντάτε στις ερωτήσεις των ανθρώπων;	Excellent	Excellent
Io e mia moglie siamo degli entusiasti attivisti sanitari.	Η σύζυγός μου και εγώ είμαστε ενθουσιώδεις ακτιβιστές υγείας.	Excellent	Good
Non siamo medici, ma date tempo abbiamo prestato servizio alla comunità.	Δεν είμαστε γιατροί, αλλά δώστε χρόνο που υπηρετήσαμε την κοινότητα.	Good	Excellent

ST	MT Output	Adequacy	Fluency
Seguiamo un codice di condotta in base al quale ogni informazione medica che condividiamo con quanti ci seguono su Twitter viene doppiamente verificata, controllandola su diversi siti di medicina oppure verificandola con Ahmed Ibrahim, la terza persona che si occupa di quest'iniziativa, ed è un medico.	Ακολουθούμε έναν κώδικα δεοντολογίας σύμφωνα με τον οποίο κάθε ιατρική πληροφορία που μοιραζόμαστε με όσους μας ακολουθούν στο Twitter επαληθεύεται διπλά, ελέγχοντας τον σε διαφορετικές ιατρικές ιστοσελίδες ή ελέγχοντας τον με τον Ahmed Ibrahim, τρίτο υπεύθυνο αυτής της πρωτοβουλίας, και είναι ένας γιατρός..	Excellent	Good
L'account è sponsorizzato da un ente commerciale o da una ONG?	Ο λογαριασμός είναι χορηγός εμπορικής οντότητας ή ΜΚΟ;	Poor	Good
Talvolta, ad esempio, fate riferimento a specifiche marche di prodotti.	Μερικές φορές, για παράδειγμα, αναφέρονται συγκεκριμένες μάρκες προϊόντων.	Excellent	Excellent
Ne siete in qualche modo affiliati?	Είσαι σε καμία περίπτωση συνδεδεμένος με αυτό;	Fair	Excellent
Non siamo ufficialmente sponsorizzati da nessun ente commerciale o ONG.	Δεν είμαστε επισήμως χορηγούμενοι από κανένα εμπορικό φορέα ή ΜΚΟ.	Excellent	Excellent
Tuttavia abbiamo ricevuto un paio di proposte, ma non siamo ancora pronti per questo passo.	Ωστόσο, έχουμε λάβει μερικές προτάσεις, αλλά ακόμα δεν είμαστε έτοιμοι για αυτό το βήμα.	Excellent	Excellent
Da quanto tempo operate online?	Πόσο καιρό λειτουργείτε online;	Good	Excellent
Prevedete una presenza oltre Twitter, ad esempio su Facebook, blog, ecc.?	Προβλέψτε μια παρουσία στο Twitter, για παράδειγμα στο Facebook, blogs κτλ.?	Poor	Fair
Abbiamo iniziato ad agosto, verso la metà del Ramadan, e la risposta in numero di persone che seguono l'account è stata piuttosto impressionante.	Ξεκινήσαμε τον Αύγουστο, γύρω στη μέση του Ραμαζανιού, και η ανταπόκριση σε αριθμό ανθρώπων που ακολουθούν τον λογαριασμό ήταν αρκετά εντυπωσιακή.	Excellent	Excellent
Abbiamo acquisito i domini "healthyegypt.org e healthyegypt.net", e stiamo lavorando all'imminente lancio del sito.	Έχουμε αποκτήσει domains "healthyegypt.org και υγιή. Δίχτυ", και δουλεύουμε για την επερχόμενη εκτόξευση του site.	Poor	Good

ST	MT Output	Adequacy	Fluency
I vostri tweet sono quasi sempre in inglese.	Τα tweets σας είναι σχεδόν πάντα στα αγγλικά.	Excellent	Excellent
Non prevedete di scrivere in arabo?	Δεν σχεδιάζετε να γράψετε στα Αραβικά;	Excellent	Excellent
Non pensate che l'uso dell'inglese finisca per limitare il pubblico di riferimento?	Δεν νομίζετε ότι η αγγλική luse καταλήγει να περιορίζει το κοινό αναφοράς;	Poor	Excellent
È assolutamente vero, talvolta scriviamo dei tweet in arabo, ma non in modo regolare.	Είναι απολύτως αλήθεια, μερικές φορές γράφουμε tweets στα αραβικά, αλλά όχι κανονικά.	Good	Excellent
Infatti il sito web sarà in entrambe le lingue per poter attrarre un maggior numero di visitatori.	Πράγματι η ιστοσελίδα θα βρίσκεται και στις δύο γλώσσες για να προσελκύσει περισσότερους επισκέπτες.	Excellent	Excellent

Google Translate – Αξιολόγηση ακρίβειας και ευχέρειας

ST	MT Output	Adequacy	Fluency
Egitto: Twitter per sensibilizzare su temi medico-sanitari	Αίγυπτος: Twitter για ευαισθητοποίηση σε θέματα υγείας και υγείας	Fair	Good
Twitter può essere usato per salvare delle vite o migliorare la salute?	Μπορεί το Twitter να χρησιμοποιηθεί για να σώσει ζωές ή να βελτιώσει την υγεία;	Excellent	Excellent
Durante gli scontri [en, come gli altri link eccetto ove diversamente indicato] dello scorso mese a piazza Tahrir, è stato creato l'account (@TahrirSupplies) per organizzare il rifornimento di provviste mediche, cibo e altri beni necessari agli ospedali da campo allestiti dai volontari nella piazza e dintorni.	Κατά τη διάρκεια των συγκρούσεων [en, όπως και οι άλλοι σύνδεσμοι, εκτός αν αναφέρεται διαφορετικά] τον περασμένο μήνα στην πλατεία Tahrir, ο λογαριασμός (@TahrirSupplies) δημιουργήθηκε για να οργανώσει την προμήθεια ιατρικών προμηθειών, τροφίμων και άλλων απαραίτητων αγαθών στα ιδρύματα των νοσοκομείων. από εθελοντές στην πλατεία και τα περίχωρα.	Good	Good
Healthy Egypt Tips è un'altra iniziativa di un gruppo di volontari che ad agosto hanno aperto l'account (@HealthyEgypt) per sensibilizzare la popolazione sul tema della salute in generale.	Το Healthy Egypt Tips είναι μια άλλη πρωτοβουλία μιας ομάδας εθελοντών που άνοιξαν έναν λογαριασμό (@HealthyEgypt) τον Αύγουστο για να ευαισθητοποιήσουν την υγεία γενικά.	Good	Excellent
Ecco un esempio dei messaggi diffusi:	Ακολουθεί ένα παράδειγμα διάδοσης των μηνυμάτων:	Fair	Excellent
@HealthyEgypt: Le ini-porzioni di caffè intervallate funzionano meglio di una grande tazza per tenervi svegli e concentrati.	@HealthyEgypt: Τα ενδιάμεσα μέρη του καφέ λειτουργούν καλύτερα από ένα μεγάλο φλιτζάνι για να σας κρατήσουν ξύπνιοι και εστιασμένους.	Fair	Excellent

Inoltre, il gruppo risponde alle domande della gente, sostiene altre iniziative e si occupa anche di educazione sessuale, in un Paese dove per ragioni culturali è spesso imbarazzante fare domande di questo tipo.	Επιπλέον, η ομάδα απαντά στις ερωτήσεις των ανθρώπων, υποστηρίζει άλλες πρωτοβουλίες και ασχολείται επίσης με τη σεξουαλική εκπαίδευση, σε μια χώρα όπου για πολιτιστικούς λόγους είναι συχνά ενοχλητικό να κά- νουμε τέτοιες ερωτήσεις.	Excellent	Excellent
Chi siete e in cosa consiste la vostra iniziativa?	Ποιος είσαι και από τι αποτελείται η πρωτοβουλία σου;	Good	Excellent
Siamo un gruppo di volontari, tutto è iniziato con me, Kareem Mossad, e mia moglie, Iman Ismail; poi si è unito un amico medico, Dr. Ahmed Ibrahim.	Είμαστε μια ομάδα εθελοντών, όλα ξεκίνησαν μαζί μου, η Kareem Mossad, και η σύζυγός μου, Iman Ismail. στη συνέχεια προσχώρησε σε έναν ιατρικό φίλο, τον Δρ. Ahmed Ibrahim.	Very Poor	Poor
Nella propria descrizione, Healthy Egypt dice di condividere delle informazioni che sono al 100% credibili e verificabili.	Στην περιγραφή τους, η Healthy Egypt αναφέρει ότι μοι- ράζονται πληροφορίες που είναι 100% αξιόπιστες και επαληθεύσιμες.	Good	Excellent
Ci sentiamo quindi in dovere di chiedervi: siete dei dottori?	Ως εκ τούτου, αισθανόμαστε υποχρεωμένοι να σας ρωτή- σουμε: είστε γιατροί;	Excellent	Excellent
Come ottenete queste informazioni mediche, come le conferma- te, e come rispondete alle domande della gente?	Πώς λαμβάνετε αυτές τις ιατρικές πληροφορίες, πώς τις επιβεβαιώνετε και πώς απαντάτε στις ερωτήσεις των ανθρώπων;	Excellent	Excellent
Io e mia moglie siamo degli entusiasti attivisti sanitari.	Η γυναίκα μου και εγώ είμαστε ενθουσιώδεις ακτιβιστές υγείας.	Excellent	Good
Non siamo medici, ma date tempo abbiamo prestato servizio alla comunità.	Δεν είμαστε γιατροί, αλλά δεδομένου του χρόνου που έχουμε υπηρετήσει την κοινότητα.	Fair	Excellent

Seguiamo un codice di condotta in base al quale ogni informazione medica che condividiamo con quanti ci seguono su Twitter viene doppiamente verificata, controllandola su diversi siti di medicina oppure verificandola con Ahmed Ibrahim, la terza persona che si occupa di quest'iniziativa, ed è un medico.	Ακολουθούμε έναν κώδικα συμπεριφοράς σύμφωνα με τον οποίο τυχόν ιατρικές πληροφορίες που μοιραζόμαστε με εκείνους που μας ακολουθούν στο Twitter επαληθεύονται διπλά, ελέγχοντάς τον σε διαφορετικούς ιατρικούς ιστότοπους ή επαληθεύοντάς τον με τον Ahmed Ibrahim, τον τρίτο υπεύθυνο αυτής της πρωτοβουλίας και είναι γιατρός.	Good	Good
L'account è sponsorizzato da un ente commerciale o da una ONG?	Ο λογαριασμός χρηματοδοτείται από εμπορική οντότητα ή ΜΚΟ;	Excellent	Good
Talvolta, ad esempio, fate riferimento a specifiche marche di prodotti.	Μερικές φορές, για παράδειγμα, αναφέρεστε σε συγκεκριμένες μάρκες προϊόντων.	Excellent	Excellent
Ne siete in qualche modo affiliati?	Είστε συνδεδεμένοι με αυτό με οποιονδήποτε τρόπο;	Good	Excellent
Non siamo ufficialmente sponsorizzati da nessun ente commerciale o ONG.	Δεν χρηματοδοτούμε επίσημα καμία εμπορική οντότητα ή ΜΚΟ.	Poor	Excellent
Tuttavia abbiamo ricevuto un paio di proposte, ma non siamo ancora pronti per questo passo.	Ωστόσο, λάβαμε μερικές προτάσεις, αλλά δεν είμαστε ακόμη έτοιμοι για αυτό το βήμα.	Excellent	Excellent
Da quanto tempo operate online?	Πόσο καιρό ήσασταν συνδεδεμένοι;	Poor	Excellent
Prevedete una presenza oltre Twitter, ad esempio su Facebook, blog, ecc.?	Περιμένετε παρουσία πέρα από το Twitter, για παράδειγμα στο Facebook, τα blogs κ.λπ.;	Good	Excellent
Abbiamo iniziato ad agosto, verso la metà del Ramadan, e la risposta in numero di persone che seguono l'account è stata piuttosto impressionante.	Ξεκινήσαμε τον Αύγουστο, στα μέσα του Ραμαζανιού, και η ανταπόκριση σε αριθμό ατόμων που ακολουθούν τον λογαριασμό ήταν αρκετά εντυπωσιακή.	Excellent	Excellent
Abbiamo acquisito i domini "healthyegypt.org e healthyegypt.net", e stiamo lavorando all'imminente lancio del sito.	Έχουμε αποκτήσει τους τομείς "sehategypt.org και sehategypt.net" και εργαζόμαστε για την επικείμενη κυκλοφορία του ιστότοπου.	Good	Good
I vostri tweet sono quasi sempre in inglese.	Τα tweet σας είναι σχεδόν πάντα στα Αγγλικά.	Excellent	Excellent

Non prevedete di scrivere in arabo?	Δεν σκοπεύετε να γράψετε στα Αραβικά;	Excellent	Excellent
Non pensate che l'uso dell'inglese finisca per limitare il pubblico di riferimento?	Δεν νομίζετε ότι η χρήση των Αγγλικών περιορίζει το κοινό-στόχο;	Excellent	Excellent
È assolutamente vero, talvolta scriviamo dei tweet in arabo, ma non in modo regolare.	Αυτό είναι απολύτως αλήθεια, μερικές φορές κάνουμε tweet στα αραβικά, αλλά όχι σε τακτική βάση.	Excellent	Excellent
Infatti il sito web sarà in entrambe le lingue per poter attrarre un maggior numero di visitatori.	Στην πραγματικότητα, ο ιστότοπος θα είναι και στις δύο γλώσσες για να προσελκύσει μεγαλύτερο αριθμό επισκεπτών.	Excellent	Excellent

Facebook Translate - Κατηγοριοποίηση λαθών

ST	MT output	No error	Accuracy	Omission	Mistranslation	Untranslated	Fluency	Punctuation	Grammar	Style	Awkward
Egitto: Twitter per sensibilizzare su temi medico-sanitari	Αίγυπτος: Twitter για την ευαισθητοποίηση του κόσμου σε ιατρικά και θέματα υγείας						1		1	1	1
Twitter può essere usato per salvare delle vite o migliorare la salute?	Μπορεί το Twitter να χρησιμοποιηθεί για να σώσει ζωές ή να βελτιώσει την υγεία;	1									
Durante gli scontri [come gli altri link eccetto ove diversamente indicato] dello scorso mese a piazza Tahrir, è stato creato l'account (@TahrirSupplies) per organizzare il rifornimento di provviste mediche, cibo e altri beni necessari agli ospedali da campo allestiti dai volontari nella piazza e dintorni.	Κατά τη διάρκεια των συμπλοκών [όπως και άλλοι σύνδεσμοι εκτός εάν αναφέρεται διαφορετικά] τον προηγούμενο μήνα στην πλατεία Ταχρίρ, δημιουργήθηκε ο λογαριασμός (@TahrirSupplies) για να οργανωθεί η προμήθεια ιατρικών προμηθειών, τροφίμων και άλλων αγαθών που απαιτούνται για τα νοσοκομεία της κατασκήνωσης από εθελοντές η πλα-		1		1		3		3		

ST	MT output	No error	Accuracy	Omission	Mistranslation	Untranslated	Fluency	Punctuation	Grammar	Style	Awkward
	τεία και η γύρω περιοχή.										
Healthy Egypt Tips è un'altra iniziativa di un gruppo di volontari che ad agosto hanno aperto l'account (@HealthyEgypt) per sensibilizzare la popolazione sul tema della salute in generale.	Το Healthy Egypt Tips είναι μια ακόμη πρωτοβουλία από ομάδα εθελοντών που άνοιξαν τον λογαριασμό τον Αύγουστο (@HealthyEgypt) για να ευαισθητοποιηθούν για το γενικό θέμα υγείας.		1		1						
Ecco un esempio dei messaggi diffusi:	Εδώ είναι ένα παράδειγμα των ευρέως διαδεδομένων μηνυμάτων:									1	1

ST	MT output	No error	Accuracy	Omission	Mistranslation	Untranslated	Fluency	Punctuation	Grammar	Style	Awkward
@HealthyEgypt: Le ini-porzioni di caffè intervallate funzionano meglio di una grande tazza per tenervi svegli e concentrati.	@HealthyEgypt: Διαλειμματικές μερίδες καφέ λειτουργούν καλύτερα από ένα μεγάλο φλιτζάνι για να είστε ξύπνιοι και συγκεντρωμένοι.		2		2		1		1		
Inoltre, il gruppo risponde alle domande della gente, sostiene altre iniziative e si occupa anche di educazione sessuale, in un Paese dove per ragioni culturali è spesso imbarazzante fare domande di questo tipo.	Επίσης, η ομάδα απαντά στα ερωτήματα των ανθρώπων, υποστηρίζει άλλες πρωτοβουλίες και ασχολείται επίσης με την σεξουαλική εκπαίδευση, σε μια χώρα όπου για πολιτιστικούς λόγους είναι συχνά αμήχανο να κάνεις τέτοιες ερωτήσεις.		1		1					2	2
Chi siete e in cosa consiste la vostra iniziativa?	Ποιος είσαι και σε τι συνίσταται η πρωτοβουλία σου;		2		2						
Siamo un gruppo di volontari, tutto è iniziato con me, Kareem Mossad, e mia moglie, Iman Ismail; poi si è unito un amico medico, Dr.	Είμαστε ομάδα εθελοντών, όλα ξεκίνησαν από εμένα, Kareem Mossad, και τη σύζυγό μου, Iman Ismail, μετά ιατρι-		4	2	1	1	2		2		

ST	MT output	No error	Accuracy	Omission	Mistranslation	Untranslated	Fluency	Punctuation	Grammar	Style	Awkward
Ahmed Ibrahim.	κός φίλος, Dr. Αχμέτ Ιμπραήμ.										
Nella propria descrizione, Healthy Egypt dice di condividere delle informazioni che sono al 100% credibili e verificabili.	Στη δική της περιγραφή, η Healthy Egypt αναφέρει ότι μοιράζονται πληροφορίες που είναι 100 % αξιόπιστες και επαληθεύσιμες.	1									
Ci sentiamo quindi in dovere di chiedervi: siete dei dottori?	Επομένως αισθανόμαστε υποχρεωμένοι να σας ρωτήσουμε: είστε γιατροί;	1									
Come ottenete queste informazioni mediche, come le confermate, e come rispondete alle domande della gente?	Πώς παίρνετε αυτές τις ιατρικές πληροφορίες, πώς τις επιβεβαιώνετε και πώς απαντάτε στις ερωτήσεις των ανθρώπων;	1									
Io e mia moglie siamo degli entusiasti attivisti sanitari.	Η σύζυγός μου και εγώ είμαστε ενθουσιώδεις ακτιβιστές υγείας .						1			1	
Non siamo medici, ma date tempo abbiamo prestato servizio alla comunità.	Δεν είμαστε γιατροί, αλλά δώστε χρόνο που υπηρετήσαμε την κοινότητα.		1			1					

ST	MT output	No error	Accuracy	Omission	Mistranslation	Untranslated	Fluency	Punctuation	Grammar	Style	Awkward
Seguiamo un codice di condotta in base al quale ogni informazione medica che condividiamo con quanti ci seguono su Twitter viene doppiamente verificata, controllandola su diversi siti di medicina oppure verificandola con Ahmed Ibrahim, la terza persona che si occupa di quest'iniziativa, ed è un medico.	Ακολουθούμε έναν κώδικα δεοντολογίας σύμφωνα με τον οποίο κάθε ιατρική πληροφορία που μοιραζόμαστε με όσους μας ακολουθούν στο Twitter επαληθεύεται διπλά, ελέγχοντας τον σε διαφορετικές ιατρικές ιστοσελίδες ή ελέγχοντας τον με τον Ahmed Ibrahim, τρίτο υπεύθυνο αυτής της πρωτοβουλίας, και είναι ένας γιατρός..						2		2	1	1
L'account è sponsorizzato da un ente commerciale o da una ONG?	Ο λογαριασμός είναι χορηγός εμπορικής οντότητας ή ΜΚΟ;		1		1						
Talvolta, ad esempio, fate riferimento a specifiche marche di prodotti.	Μερικές φορές, για παράδειγμα, αναφέρονται συγκεκριμένες μάρκες προϊόντων.	1									
Ne siete in qualche modo affiliati?	Είσαι σε καμία περίπτωση συνδεδεμένος με αυτό;		3		3						

ST	MT output	No error	Accuracy	Omission	Mistranslation	Untranslated	Fluency	Punctuation	Grammar	Style	Awkward
Non siamo ufficialmente sponsorizzati da nessun ente commerciale o ONG.	Δεν είμαστε επίσημως χορηγούμενοι από κανένα εμπορικό φορέα ή ΜΚΟ.	1									
Tuttavia abbiamo ricevuto un paio di proposte, ma non siamo ancora pronti per questo passo.	Ωστόσο, έχουμε λάβει μερικές προτάσεις, αλλά ακόμα δεν είμαστε έτοιμοι για αυτό το βήμα.	1									
Da quanto tempo operate online?	Πόσο καιρό λειτουργείτε online;		1		1						
Prevedete una presenza oltre Twitter, ad esempio su Facebook, blog, ecc.?	Προβλέψτε μια παρουσία στο Twitter, για παράδειγμα στο Facebook, blogs κτλ.?		2		2		1		1		
Abbiamo iniziato ad agosto, verso la metà del Ramadan, e la risposta in numero di persone che seguono l'account è stata piuttosto impressionante.	Ξεκινήσαμε τον Αύγουστο, γύρω στη μέση του Ραμαζανιού, και η ανταπόκριση σε αριθμό ανθρώπων που ακολουθούν τον λογαριασμό ήταν αρκετά εντυπωσιακή.	1									
Abbiamo acquisito i domini "healthyegypt.org e healthyegypt.net", e stiamo lavorando	Έχουμε αποκτήσει domains "healthyegypt.org και υγιή. Δίκτυο ", και δουλεύουμε		3	1	2		2	1	1		

ST	MT output	No error	Accuracy	Omission	Mistranslation	Untranslated	Fluency	Punctuation	Grammar	Style	Awkward
all'imminente lancio del sito.	για την επερχόμε- νη εκτόξευση του site.										
I vostri tweet sono quasi sempre in inglese.	Τα tweets σας είναι σχεδόν πάντα στα αγγλικά.	1									
Non prevedete di scrivere in arabo?	Δεν σχεδιάζετε να γράψετε στα Αραβικά;	1									
Non pensate che l'uso dell'inglese finisca per limitare il pubblico di riferimento?	Δεν νομίζετε ότι η αγγλική luse καταλήγει να περιορίζει το κοινό αναφοράς;			1							
È assolutamente vero, talvolta scriviamo dei tweet in arabo, ma non in modo regolare.	Είναι απολύτως αλήθεια, μερικές φορές γράφουμε tweets στα αραβικά, αλλά όχι κανονικά.		1		1						
Infatti il sito web sarà in entrambe le lingue per poter attrarre un maggior numero di visitatori.	Πράγματι η ιστοσελίδα θα βρίσκεται και στις δύο γλώσσες για να προσελκύσει περισσότερους επισκέπτες.		1		1						

Google Translate - Κατηγοριοποίηση λαθών

ST	MT output	No error	Accuracy	Omission	Mistranslation	Fluency	Punctuation	Grammar	Style	Awkward
Egitto: Twitter per sensibilizzare su temi medico-sanitari	Αίγυπτος: Twitter για ευαισθητοποίηση σε θέματα υγείας και υγείας		1		1	2		2		
Twitter può essere usato per salvare delle vite o migliorare la salute?	Μπορεί το Twitter να χρησιμοποιηθεί για να σώσει ζωές ή να βελτιώσει την υγεία;									
Durante gli scontri [come gli altri link eccetto ove diversamente indicato] dello scorso mese a piazza Tahrir, è stato creato l'account (@TahrirSupplies) per organizzare il rifornimento di provviste mediche, cibo e altri beni necessari agli ospedali da campo allestiti dai volontari nella piazza e dintorni.	Κατά τη διάρκεια των συγκρούσεων [όπως και οι άλλοι σύνδεσμοι, εκτός αν αναφέρεται διαφορετικά] τον περασμένο μήνα στην πλατεία Tahrir, ο λογαριασμός (@TahrirSupplies) δημιουργήθηκε για να οργανώσει την προμήθεια ιατρικών προμηθειών, τροφίμων και άλλων απαραίτητων αγαθών στα ιδρύματα των νοσοκομείων . από εθελοντές στην πλατεία και τα περίχωρα.		1		1	1	1		1	1
Healthy Egypt Tips è un'altra iniziativa di un gruppo di volontari che ad agosto hanno aperto l'account (@HealthyEgypt) per sensibilizzare la popolazione sul tema della salute in generale.	Το Healthy Egypt Tips είναι μια άλλη πρωτοβουλία μιας ομάδας εθελοντών που άνοιξαν έναν λογαριασμό (@HealthyEgypt) τον Αύγουστο για να ευαισθητοποιήσουν την υ-		2	2						

ST	MT output	No error	Accuracy	Omission	Mistranslation	Fluency	Punctuation	Grammar	Style	Awkward
	γεία γενικά.									
Ecco un esempio dei messaggi diffusi:	Ακολουθεί ένα παράδειγμα διάδοσης των μηνυμάτων :		1		1					
@HealthyEgypt: Le ini-porzioni di caffè intervallate funzionano meglio di una grande tazza per tenervi svegli e concentrati.	@HealthyEgypt: Τα ενδιάμεσα μέρη του καφέ λειτουργούν καλύτερα από ένα μεγάλο φλιτζάνι για να σας κρατήσουν ξύπνιοι και εστιασμένοι .		2		2				1	1
Inoltre, il gruppo risponde alle domande della gente, sostiene altre iniziative e si occupa anche di educazione sessuale, in un Paese dove per ragioni culturali è spesso imbarazzante fare domande di questo tipo.	Επιπλέον, η ομάδα απαντά στις ερωτήσεις των ανθρώπων, υποστηρίζει άλλες πρωτοβουλίες και ασχολείται επίσης με τη σεξουαλική εκπαίδευση, σε μια χώρα όπου για πολιτιστικούς λόγους είναι συχνά ενοχλητικό να κάνουμε τέτοιες ερωτήσεις.					1		1		
Chi siete e in cosa consiste la vostra iniziativa?	Ποιος είσαι και από τι αποτελείται η πρωτοβουλία σου ;		2		2					

ST	MT output	No error	Accuracy	Omission	Mistranslation	Fluency	Punctuation	Grammar	Style	Awkward
Siamo un gruppo di volontari, tutto è iniziato con me, Kareem Mossad, e mia moglie, Iman Ismail; poi si è unito un amico medico, Dr. Ahmed Ibrahim.	Είμαστε μια ομάδα εθελοντών, όλα ξεκίνησαν μαζί μου, η Kareem Mossad, και η σύζυγός μου, Iman Ismail. στη συνέχεια προσχώρησε σε έναν ιατρικό φίλο, τον Δρ. Ahmed Ibrahim.		3		3	3		3		
Nella propria descrizione, Healthy Egypt dice di condividere delle informazioni che sono al 100% credibili e verificabili.	Στην περιγραφή τους, η Healthy Egypt αναφέρει ότι μοιράζονται πληροφορίες που είναι 100% αξιόπιστες και επαληθεύσιμες.					1		1		
Ci sentiamo quindi in dovere di chiedervi: siete dei dottori?	Ως εκ τούτου, αισθανόμαστε υποχρεωμένοι να σας ρωτήσουμε: είστε γιατροί;	1								
Come ottenete queste informazioni mediche, come le confermate, e come rispondete alle domande della gente?	Πώς λαμβάνετε αυτές τις ιατρικές πληροφορίες, πώς τις επιβεβαιώνετε και πώς απαντάτε στις ερωτήσεις των ανθρώπων;	1								
Io e mia moglie siamo degli entusiasti attivisti sanitari.	Η γυναίκα μου και εγώ είμαστε ενθουσιώδεις ακτιβιστές υγείας.					1		1		
Non siamo medici, ma date tempo abbiamo prestato servizio alla comunità.	Δεν είμαστε γιατροί, αλλά δεδομένου του χρόνου που έχουμε υπηρετήσει την κοινότητα.		1		1					

ST	MT output	No error	Accuracy	Omission	Mistranslation	Fluency	Punctuation	Grammar	Style	Awkward
Seguiamo un codice di condotta in base al quale ogni informazione medica che condividiamo con quanti ci seguono su Twitter viene doppiamente verificata, controllandola su diversi siti di medicina oppure verificandola con Ahmed Ibrahim, la terza persona che si occupa di quest'iniziativa, ed è un medico.	Ακολουθούμε έναν κώδικα συμπεριφοράς σύμφωνα με τον οποίο τυχόν ιατρικές πληροφορίες που μοιραζόμαστε με εκείνους που μας ακολουθούν στο Twitter επαληθεύονται διπλά, ελέγχοντάς τον σε διαφορετικούς ιατρικούς ιστότοπους ή επαληθεύοντάς τον με τον Ahmed Ibrahim, τον τρίτο υπεύθυνο αυτής της πρωτοβουλίας και είναι γιατρός.		2		2				1	1
L'account è sponsorizzato da un ente commerciale o da una ONG?	Ο λογαριασμός χρηματοδοτείται από εμπορική οντότητα ή ΜΚΟ;		1	1						
Talvolta, ad esempio, fate riferimento a specifiche marche di prodotti.	Μερικές φορές, για παράδειγμα, αναφέρεστε σε συγκεκριμένες μάρκες προϊόντων.	1								
Ne siete in qualche modo affiliati?	Είστε συνδεδεμένοι με αυτό με οποιονδήποτε τρόπο;		1		1					
Non siamo ufficialmente sponsorizzati da nessun ente commerciale o ONG.	Δεν χρηματοδοτούμε επίσημα καμία εμπορική οντότητα ή ΜΚΟ.		1		1					
Tuttavia abbiamo ricevuto un paio di proposte, ma non siamo ancora pronti per questo passo.	Ωστόσο, λάβαμε μερικές προτάσεις, αλλά δεν είμαστε ακόμη έτοιμοι για αυτό το βήμα.	1								
Da quanto tempo operate online?	Πόσο καιρό ήσασταν συνδεδεμένοι ;		1		1					

ST	MT output	No error	Accuracy	Omission	Mistranslation	Fluency	Punctuation	Grammar	Style	Awkward
Prevedete una presenza oltre Twitter, ad esempio su Facebook, blog, ecc.?	Περιμένετε παρουσία πέρα από το Twitter, για παράδειγμα στο Facebook, τα blogs κ.λπ.;		1		1					
Abbiamo iniziato ad agosto, verso la metà del Ramadan, e la risposta in numero di persone che seguono l'account è stata piuttosto impressionante.	Ξεκινήσαμε τον Αύγουστο, στα μέσα του Ραμαζανιού, και η ανταπόκριση σε αριθμό ατόμων που ακολουθούν τον λογαριασμό ήταν αρκετά εντυπωσιακή.	1								
Abbiamo acquisito i domini "healthyegypt.org e healthyegypt.net", e stiamo lavorando all'imminente lancio del sito.	Έχουμε αποκτήσει τους τομείς "sehategypt.org και sehategypt.net" και εργαζόμαστε για την επικείμενη κυκλοφορία του ιστότοπου.		1		1	1	1			
I vostri tweet sono quasi sempre in inglese.	Τα tweet σας είναι σχεδόν πάντα στα Αγγλικά.	1								
Non prevedete di scrivere in arabo?	Δεν σκοπεύετε να γράψετε στα Αραβικά;	1								
Non pensate che l'uso dell'inglese finisca per limitare il pubblico di riferimento?	Δεν νομίζετε ότι η χρήση των Αγγλικών περιορίζει το κοινό-στόχο;	1								
È assolutamente vero, talvolta scriviamo dei tweet in arabo, ma non in modo regolare.	Αυτό είναι απολύτως αλήθεια, μερικές φορές κάνουμε tweet στα αραβικά, αλλά όχι σε τακτική βάση.	1								
Infatti il sito web sarà in entrambe le lingue per poter attrarre un maggior numero di visitatori.	Στην πραγματικότητα, ο ιστότοπος θα είναι και στις δύο γλώσσες για να προσελκύσει μεγαλύτερο αριθμό επισκεπτών.	1								

Κώδικας Tokenizer για την ελληνική γλώσσα

```
# MTUOC_tokenizer_gr
# Copyright (C) 2020 Antoni Oliver, Konstantinos Dagklis
#
# This program is free software: you can redistribute it and/or modify
# it under the terms of the GNU General Public License as published by
# the Free Software Foundation, either version 3 of the License, or
# (at your option) any later version.
#
# This program is distributed in the hope that it will be useful,
# but WITHOUT ANY WARRANTY; without even the implied warranty of
# MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the
# GNU General Public License for more details.
#
# You should have received a copy of the GNU General Public License
# along with this program. If not, see <https://www.gnu.org/licenses/>.

from nltk import word_tokenize
from nltk.tokenize.toktok import ToktokTokenizer
from nltk.tokenize.treebank import TreebankWordTokenizer, TreebankWordDetokenizer
import sys
import io
input_stream = io.TextIOWrapper(sys.stdin.buffer, encoding='utf-8', errors="replace")

def tokenize(segment):
    global subs
    tokens = tok.tokenize(segment)
    tokenized=" ".join(tokens)
    for sub in subs:
        tokenized=tokenized.replace(sub[0],sub[1])
        tokenized=tokenized.replace(sub[0].upper(),sub[1].upper())
        tokenized=tokenized.replace(sub[0].capitalize(),sub[1].capitalize())
    tokenized=' '.join(tokenized.split())
    return(tokenized)

def detokenize(segment):
    #generic
    for det in subs:
        if not det[1]==det[1].replace(" ",""):
            segment=segment.replace(det[1],det[1].replace(" ",""))
            segment=segment.replace(det[1].upper(),det[1].replace(" ","").upper())
            segment=segment.replace(det[1].capitalize(),det[1].replace(" ","").capitalize())
    tokens=segment.split(" ")
    desegment=detok.detokenize(tokens)
    return(desegment)
tok = ToktokTokenizer()
detok = TreebankWordDetokenizer()

subs=[
    ("αντ' ", "αντ' "),
```

```

("απ' ", "απ' "),
("αφ' ", "αφ' "),
("γλ' ", "γλ' "),
("δώσ' ", "δώσ' "),
("επ' ", "επ' "),
("εφ' ", "εφ' "),
("θα' ", "θα' "),
("καθ' ", "καθ' "),
("κατ' ", "κατ' "),
("μ' ", "μ' "),
("μεθ' ", "μεθ' "),
("μετ' ", "μετ' "),
("ν' ", "ν' "),
("να' ", "να' "),
("πάρ' ", "πάρ' "),
("σ' ", "σ' "),
("στ' ", "στ' "),
("τ' ", "τ' "),
("υπ' ", "υπ' "),
("υφ' ", "υφ' "),
("φέρ' ", "φέρ' ")

]

nosplit={}

if __name__ == "__main__":
    if len(sys.argv)>1:
        if sys.argv[1]=="-h" or sys.argv[1]=="--help":
            print("MTUOC Tokenizer and Detokenizer for Greek:")
            print("Tokenize text file:")
            print("    python3 MTUOC_tokenizer_gr.py < inputfile > tokenizedfile")
            print("    python3 MTUOC_tokenizer_gr.py tokenize < inputfile > tokenizedfile")
            print("Tokenize text file:")
            print("    python3 MTUOC_tokenizer_gr.py detokenize < inputfile > tokenizedfile")
            sys.exit()

        if len(sys.argv)>1:
            action=sys.argv[1]
        else:
            action="tokenize"
        for line in sys.stdin:
            line=line.strip()
            #normalization of apostrophe
            line=line.replace("'", "")
            line=line.replace('\'', '')
            line=line.replace('\'', '')
            line=line.replace('.', ' . ')
            line=line.replace('/', ' /')

```

```
line=line.replace("\\","\\")
line=line.replace("€"," € ")
line=line.replace("£"," £ ")
line=line.replace("¢"," ¢ ")
line=line.replace("¥"," ¥ ")
line=line.replace("#"," # ")
```

```
if action=="tokenize":
    output=tokenize(line)
    print(output)
elif action=="detokenize":
    output=detokenize(line)
    print(output)
```

