

# Μηχανική Μετάφραση και Ελεγχόμενες Γλώσσες

Ελευθερία Φλώρου

*Imperial College London*

## Abstract:

The production of an unambiguous translation by means of machine translation dictates the use of Controlled Languages. Controlled Languages are used in order to restrict the characteristics of a natural language that result in grammatical, syntactic and structural ambiguities. However, there are still margins of improvement in the machine translation output whether Controlled Language rules are applied or not. This assumption is based on the results occurring after the evaluation process of the output. The pre- and post-editing phase are essential in order to obtain a translation result of higher quality. A Controlled Language, on the other hand, may prove to be useful regarding the amelioration of the text alignment results. This article provides a study on the translation results occurring by means of a machine translation system and a discussion of the problems follows. However, the intention of this study is to achieve: (a) an analysis of the translation results produced by translating a corpus with and without the use of Controlled English into Modern Greek and (b) a statistical analysis of the Source text and its machine translation results in the level of alignment with and without the use of Controlled Language rules. The alignment results of texts, in which controlled language rules are applied, proved to be satisfying and the hypothesis that machine translation, controlled language and text alignment could be related, though in a complex way, proved to be existing.

Keywords: machine translation, controlled language, text alignment, input, output, Modern Greek

## 1. Εισαγωγή

Στα πλαίσια της επικοινωνίας η μετάφραση θεωρείται πολύ χρήσιμο εργαλείο και διαδραματίζει έναν ουσιαστικό ρόλο. Ιδιαίτερα τα τελευταία χρόνια εξαιτίας της παγκοσμιοποίησης και της διεθνοποίησης της αγοράς, η ανάγκη για τη μετάφραση κειμένων μεγαλώνει καθημερινά. Ωστόσο, η διαδικασία της μετάφρασης είναι χρονοβόρα και αυτό έρχεται σε αντίθεση με την ανάγκη για αύξηση του όγκου των μεταφράσεων. Είναι αυτή η ανάγκη που ώθησε τη μεταφραστική βιομηχανία να στραφεί σε εναλλακτικούς τρόπους μετάφρασης πέραν του παραδοσιακού και πιο συγκεκριμένα στη μηχανική ή αυτόματη μετάφραση, κυρίως σε είδη κειμένων όπως τα τεχνικά, με σκοπό την εξοικονόμηση χρόνου και χρημάτων καθώς και την εξασφάλιση μεγαλύτερου όγκου μεταφράσεων. Ωστόσο, τα μεταφραστικά αποτελέσματα που προέρχονται από τη χρήση συστημάτων μηχανικής μετάφρασης, παρουσιάζουν πληθώρα προβλημάτων σε επίπεδο συντακτικογραμματικό και λεξιλογικό. Η δημιουργία των ελεγχόμενων γλωσσών (controlled languages) και η χρήση των κανόνων που τις χαρακτηρίζουν, σε συνδυασμό πάντα με τη μηχανική μετάφραση επιφέρουν κατά ένα μεγάλο ποσοστό τη βελτίωση του μεταφραστικού αποτελέσματος.

## 2. Μηχανική Μετάφραση

Κατά τις τελευταίες δεκαετίες γλωσσολόγοι, μεταφραστές και άλλοι επιστήμονες πραγματοποίησαν έρευνες σχετικές με τις δυνατότητες της μηχανικής μετάφρασης – με τον όρο μηχανική μετάφραση εννοείται η μετάφραση κειμένων από μια γλώσσα σε μια άλλη με τη βοήθεια ηλεκτρονικού υπολογιστή και προγραμμάτων

επεξεργασίας κειμένων χωρίς την παρέμβαση ανθρώπινου παράγοντα, σαφώς, η εισαγωγή του κειμένου γίνεται με τη βοήθεια ανθρώπου αλλά η διαδικασία της μετάφρασης γίνεται αποκλειστικά και μόνο από τον υπολογιστή – και αυτή η συστηματική μελέτη οδήγησε στη δημιουργία συστημάτων μηχανικής μετάφρασης.

Ωστόσο, η προσπάθεια που κατέβαλαν οι επιστήμονες σε σχέση με τα αποτελέσματα που επιτεύχθηκαν με τη χρήση της μηχανικής μετάφρασης δεν είναι ανάλογα. Είναι γεγονός ότι έχει σημειωθεί πρόοδος στον τομέα αυτό αλλά δεν πραγματοποιήθηκε σε τόσο γρήγορο ρυθμό όσο αναμενόταν. Τα αποτελέσματα της μηχανικής μετάφρασης είναι φτωχά σε ποιότητα, και σε επίπεδο ακρίβειας (accuracy) και σε επίπεδο κατανόησης κειμένου (intelligibility).

## **2.1 Ιδέες που αφορούν στη μηχανική μετάφραση**

Είναι αλήθεια πως η χρήση της Μηχανικής Μετάφρασης είναι παρεξηγημένη ακόμα και από άτομα τα οποία ανήκουν στο χώρο της μεταφραστικής βιομηχανίας. Πολλοί και ειδικότερα οι χρήστες του διαδικτύου αγνοούν το τί ακριβώς είναι η Μηχανική Μετάφραση και ποιος είναι ο ρόλος της και αυτό επιφέρει ως αποτέλεσμα τη δημιουργία λανθασμένων αντιλήψεων γύρω από τον τομέα αυτό. Η μηχανική μετάφραση δε φτιάχτηκε για να μεταφράζει Σαίξπηρ όπως αναφέρει ο Arnold (1994). Την ίδια στιγμή, η Μηχανική Μετάφραση προκαλεί φόβο σε μια μερίδα μεταφραστών οι οποίοι θεωρούν πως ελλοχεύει ο κίνδυνος να χάσουν τη δουλειά τους.

Αυτό που είναι σημαντικό, είναι να γίνει κατανοητό ότι με τον όρο Μηχανική Μετάφραση δεν εννοείται το πάτημα ενός κουμπιού του υπολογιστή και η παραγωγή ενός άσφουα μεταφρασμένου κειμένου. Σε καμία περίπτωση όμως δε θα μπορούσε κανείς να πει ότι από τα προβληματικά αποτελέσματα που έχουμε σήμερα προκύπτει το συμπέρασμα πως ο τομέας της Μηχανικής Μετάφρασης είναι άχρηστος. Πολλά συστήματα όπως το METEO που χρησιμοποιείται από το 1977 από το Canadian Meteorological Center, το SYSTRAN και το LOGOS, αποδεικνύουν το αντίθετο (Arnold, 1994).

## **2.2 Προβλήματα της μηχανικής μετάφρασης**

Τα προβλήματα τα οποία παρατηρούνται στα κείμενα που παράγονται από τις μεταφραστικές μηχανές είναι ποικίλα και κλιμακώνονται σε διάφορα επίπεδα. Η γραμματική αποτελεί ένα πολύ σημαντικό προβληματικό κομμάτι και πιο συγκεκριμένα όταν μιλάμε για γραμματική, παρατηρούνται προβλήματα ως προς τη μετάφραση της αναφοράς, της παθητικής φωνής, των μετοχών, της προστακτικής και του οριστικού και αόριστου άρθρου. Προβλήματα επίσης παρουσιάζονται στη σύνταξη, το λεξιλόγιο και ειδικότερα τις συντομογραφίες, τα αρκτικόλεξα, τα ομόγραφα και κυρίως τη μετάφραση της ορολογίας καθώς επίσης και στη στίξη. Αυτά θα μπορούσαμε με ασφάλεια να πούμε ότι αποδίδονται στην έλλειψη της ανθρώπινης γνώσης του υπολογιστή. Με άλλα λόγια ο υπολογιστής δε μπορεί να κατανοήσει έννοιες που για τον άνθρωπο είναι αυτονόητες (π.χ. Ο Γιώργος είναι έγκυος και περιμένει κορίτσι).

## **3.0 Ελεγχόμενες Γλώσσες**

Δεδομένου ότι η μηχανική μετάφραση παρουσιάζει αρκετά προβλήματα, θεωρήθηκε ότι είναι προτιμότερο από το να αναμένουμε μια νέα αρχιτεκτονική μηχανικής μετάφρασης να ελέγξουμε τη γλώσσα του κειμένου το οποίο εισάγεται στο σύστημα

μηχανικής μετάφρασης. Σύμφωνα με τον ορισμό του Kittredge “Η ελεγχόμενη γλώσσα (Controlled Language) αποτελεί μια περιορισμένη εκδοχή μιας φυσικής γλώσσας η οποία έχει υποστεί επεξεργασία με σκοπό την εκπλήρωση ενός συγκεκριμένου στόχου” (Kittredge 2003: 441-442). Οι Μαρκαντωνάτου & Μαΐστρος (2001) ορίζουν την ελεγχόμενη γλώσσα ως ‘ένα υποσύνολο/ υποδιαίρεση μιας φυσικής γλώσσας που προσδιορίζει έναν αριθμό περιορισμών στη μορφολογία, το λεξιλόγιο, τη σύνταξη και το ύφος του κειμένου’. Με άλλα λόγια, ελεγχόμενη γλώσσα θα χαρακτηρίζαμε τη γλώσσα που χρησιμοποιείται για τη συγγραφή ή διόρθωση ενός κειμένου που προορίζεται για μετάφραση, κατά τη συγγραφή ή διόρθωση του οποίου αποφεύγονται συγκεκριμένες συντακτικογραμματικές δομές και λέξεις που προκαλούν σύγχυση στον υπολογιστή.

Θα πρέπει σε αυτό το σημείο να τονίσουμε πως οι ελεγχόμενες γλώσσες χρησιμοποιούνται ως επί το πλείστον σε τεχνικά κείμενα και κυρίως από εταιρίες όπως μεγάλες αυτοκινητοβιομηχανίες και εταιρίες κατασκευής αεροπλάνων. Είναι σαφές ότι τα κείμενα πρέπει να δημιουργούνται από συγγραφείς τεχνικών κειμένων ή από μεταφραστές με ειδικευση και εμπειρία στις ελεγχόμενες γλώσσες κάτι το οποίο αναπτύσσεται ιδιαίτερα από τον Torrejón & Rico (2002: 107-116). Επίσης, σύμφωνα με τον Allen (1999) υπάρχει διαφορά μεταξύ των κειμένων που «προορίζονται για μετάφραση», δηλαδή των κειμένων που έχει αποφασισθεί πριν από τη συγγραφή τους ότι θα μεταφραστούν, και των κειμένων που «επιλέχθηκαν για μετάφραση» μετά την ολοκλήρωση της συγγραφής τους. Ο Allen (1999) επιχειρηματολογεί λέγοντας ότι για ένα κείμενο ή εγχειρίδιο που εξ αρχής παράγεται με σκοπό τη μετάφραση του, είναι ευκολότερο να διαχειριστεί κάποιος το προσωπικό, δηλαδή τη συγγραφική ομάδα που αποτελείται από τεχνικούς συγγραφείς και μεταφραστές, έτσι ώστε να εφαρμοστούν οι αρχές που διέπουν τα τεχνικά κείμενα και στη συνέχεια να συντελέσουν στη *μεταφρασιμότητα* (translatability) των κειμένων.

Σε αυτό το σημείο αξίζει να σημειωθεί πως η ύπαρξη των Ελεγχόμενων Γλωσσών χρονολογείται από το 1920 και ποικίλες Ελεγχόμενες Γλώσσες έχουν δημιουργηθεί και σχεδιαστεί. Ωστόσο, το μεγαλύτερο κομμάτι της έρευνας αφορά την Ελεγχόμενη Αγγλική Γλώσσα και σχετικά με την Ελληνική είναι πολύ περιορισμένη η έρευνα που έχει πραγματοποιηθεί έως σήμερα η οποία έχει παραμείνει σε πειραματικό στάδιο και αυτό κατά την προσωπική μου άποψη οφείλεται στην έλλειψη της βιομηχανικής ανάπτυξης στην Ελλάδα.

### 3.1 Κανόνες Ελεγχόμενων Γλωσσών

Όπως προαναφέρθηκε, μια ελεγχόμενη γλώσσα χρησιμοποιείται με σκοπό τη μείωση της *αμφισημίας* και την ενίσχυση της *μεταφρασιμότητας* ενός κειμένου. Ωστόσο, το μετάφρασμα είτε είναι αποτέλεσμα ανθρώπινης επεξεργασίας είτε μηχανικής βασίζεται στην εφαρμογή συγκεκριμένων κανόνων. Σε αυτό το σημείο κρίνεται απαραίτητο να αναλυθούν αυτοί οι κανόνες και περιορισμοί και να αναφερθεί ότι υπάρχουν τέσσερις (4) τύποι κανόνων σύμφωνα με την O’Brien (2003: 105-114) i) λεξιλογικοί ii) συντακτικοί iii) κειμενικοί και iv) πραγματολογικοί.

Ξεκινώντας, σύμφωνα με την O’Brien (2003: 105-114), μόνον ένα αποδεκτό λεξιλόγιο σε επίπεδο πολυσημίας θα πρέπει να χρησιμοποιείται. Προσπαθώντας να μεταφράσει κάποιος κείμενα με τη χρήση συστημάτων μηχανικής μετάφρασης συνιστάται η αποφυγή της χρήσης λέξεων με περισσότερες από μία σημασίες. Με άλλα λόγια, θα ήταν καλύτερο να χρησιμοποιείται μια λέξη πάντα με μια και μοναδική σημασία, και να επιλέγεται μια άλλη λέξη για την απόδοση κάποιας άλλης σημασίας. Για παράδειγμα η λέξη *right* στην Αγγλική γλώσσα μπορεί να σημαίνει

είτε *δεξιά* είτε *σωστό*. Στην περίπτωση των τεχνικών κειμένων συνιστάται να χρησιμοποιείται μόνο με την έννοια *δεξιά* και να χρησιμοποιείται η λέξη *correct* για να εκφράσει την έννοια *σωστό*. Αντιστοίχως, ο όρος *port* στα τεχνικά κείμενα αποδίδεται ως *θύρα* αλλά τα μεταφραστικά συστήματα τείνουν να τον αποδίδουν με τη λέξη *λιμάνι*.

Η χρήση της ορθογραφίας πρέπει να είναι η αρμόζουσα και να αποφεύγονται ορθογραφικά ή τυπογραφικά λάθη διότι διαφορετικά το σύστημα θα αντικαταστήσει τους λατινικούς χαρακτήρες με τους αντίστοιχους της ελληνικής γλώσσας, καθώς επίσης θα πρέπει να είναι σταθερή, δηλαδή να χρησιμοποιείται μονίμως μόνο ένας εκ των τρόπων με τους οποίους ενδέχεται να γράφεται μια λέξη όπως για παράδειγμα *τρένο* αλλά και *τραίνο* ή *analyze* και *analyse*.

Οι μετοχές που λήγουν σε *-ing* ή *-ed* απαγορεύονται. Οι μετοχές ενεστώτα δεν πρέπει να χρησιμοποιούνται και οι μετοχές αορίστου να χρησιμοποιούνται μόνον ως επίθετα.

Ένα από τα σημαντικότερα προβλήματα που παρατηρείται στα αποτελέσματα της μηχανικής μετάφρασης προκαλείται από τη χρήση των αντωνυμιών, με άλλα λόγια αναφερόμαστε στο πρόβλημα της *αναφοράς*. Σύμφωνα με τον Somers (2003: 518-519) η απόδοση της αναφοράς είναι υψίστης σημασίας για την επιτυχία του αποτελέσματος της Μηχανικής Μετάφρασης και αυτό είναι προφανές κατά τη μετάφραση προς γλώσσες που διακρίνουν τα γένη όπως τα Ελληνικά και τα Γαλλικά και άλλων που έχουν μηδενική αναφορά (*zero-anaphora*). Η χρήση αντωνυμιών όπως η αόριστη αντωνυμία “*one*” είναι απαγορευτική μιας και στα Ελληνικά θα μπορούσε να αποδοθεί ως *κάποιος*, *κάποια*, *κάποιο* ή *ένας*, *μία*, *ένα* και η επιλογή του γένους από το σύστημα να είναι τυχαία και όχι η πρόβουσα. Παράλληλα, και το φαινόμενο της *έλλειψης* θα πρέπει να αποφεύγεται.

Γενικότερα, η χρήση της παθητικής φωνής απαγορεύεται και πρέπει να γίνεται χρήση μόνο της ενεργητικής φωνής. Κατά τη δημιουργία των κειμένων πρέπει να αποφεύγονται οι συντομογραφίες και τα αρκτικόλεξα, καθώς επίσης οι προτάσεις πρέπει να χαρακτηρίζονται από απλότητα στη σύνταξη και σύνθετες ή προβληματικές συντακτικογραμματικές δομές να αποφεύγονται. Προτείνεται να εφαρμόζονται οι κανόνες στίξης με ακρίβεια και τέλος, οι προτάσεις πρέπει να είναι σύντομες και να μην υπερβαίνουν τις είκοσι (20) λέξεις, όπως επίσης και προτάσεις που αποτελούνται από τέσσερις (4) ή λιγότερες λέξεις να αποφεύγονται.

#### 4. Μεθοδολογία

Η συγκεκριμένη μελέτη δε βασίστηκε στη δημιουργία μιας νέας ελεγχόμενης γλώσσας αλλά στους υπάρχοντες κανόνες της Αγγλικής Ελεγχόμενης Γλώσσας σύμφωνα με την O'Brien. Σχετική θεωρία με σκοπό την κατανόηση της φύσης και της χρησιμότητας της Ελεγχόμενης Γλώσσας μελετήθηκε.

Το πείραμα πραγματοποιήθηκε με σκοπό τη μελέτη και καταγραφή των προβλημάτων στις μεταφράσεις διαφόρων κειμένων. Το ζευγάρι γλωσσών που εξετάστηκε είναι Αγγλικά > Ελληνικά. Τα κείμενα προέλευσης (*source texts*) ήταν στα Αγγλικά και η μετάφραση (*target texts*) στα Ελληνικά. Η κατηγορία στην οποία ανήκαν τα κείμενα – που σύστησαν ένα corpus 10.000 λέξεων – είναι αυτή των τεχνικών κειμένων και πιο συγκεκριμένα εγχειρίδια κινητών τηλεφώνων. Το πρώτο στάδιο του πειράματος μετά την επιλογή των κειμένων ήταν η μετάφρασή τους με τη βοήθεια του προγράμματος Systran χωρίς τη διαδικασία του *pre-editing*.

Η μετάφραση αυτή οδήγησε στον εντοπισμό των προβληματικών δομών και προτάσεων. Στη συνέχεια, στη δεύτερη φάση του πειράματος εφαρμόστηκαν οι κανόνες μιας ελεγχόμενης γλώσσας – όπως έχουν διατυπωθεί από την O'Brien – στο κείμενο προέλευσης (source text) και στη συνέχεια μετά την παραγωγή της μετάφρασης μελετήθηκαν τα αποτελέσματα δηλαδή, ακολούθησε η διαδικασία του post-editing.

Το τρίτο μέρος του πειράματος εξέτασε κατά πόσο η εφαρμογή των ελεγχόμενων γλωσσών διευκολύνει τη διαδικασία στοίχισης κειμένων (text alignment) με σκοπό τη δημιουργία παράλληλων σωμάτων κειμένων

## **5. Κείμενο εισαγωγής & pre-editing και μεταφραστικό αποτέλεσμα & post-editing**

Ο συνδυασμός της ελεγχόμενης γλώσσας, των συστημάτων μηχανικής μετάφρασης και της διαδικασίας του pre-editing και του post-editing σε συνδυασμό με την συνεχώς αυξανόμενη χρήση της μεταφραστικής μνήμης, των εργαλείων CAT, των συστημάτων διαχείρισης ορολογίας κ.α. απεικονίζουν μια νέα πραγματικότητα στον τομέα της μετάφρασης (Torrejón & Rico 2002: 107-116 ). Η διαδικασία και η παραγωγή μετάφρασης με τη χρήση συστημάτων μηχανικής μετάφρασης – συστημάτων αυτόματης μετάφρασης – καθιστά αναγκαία, με σκοπό την επίτευξη του καλύτερου μεταφραστικού αποτελέσματος, τη φάση του pre-editing. Σε αυτή τη φάση, το κείμενο πηγή υφίσταται επεξεργασία βελτίωσης, δηλαδή αντικαθίστανται όλες οι σύνθετες συντακτικογραμματικές και λεξιλογικές δομές με άλλες πιο απλές και πιο κατανοητές για το σύστημα μηχανικής μετάφρασης που χρησιμοποιείται προσπαθώντας για το καλύτερο μεταφραστικό αποτέλεσμα. Ωστόσο, το κείμενο στόχος χρειάζεται τη διαδικασία του post-editing ώστε να εντοπισθούν και να διορθωθούν οι πιθανές γλωσσολογικές ατέλειες που οφείλονται στις αδυναμίες του συστήματος μηχανικής μετάφρασης. Στη συνέχεια δίνονται επιπλέον πληροφορίες για τη διαδικασία του pre-editing και του post-editing.

### **5.1 Pre-editing**

Όταν αναφερόμαστε στον όρο pre-editing εννοούμε την επεξεργασία σε γλωσσολογικό επίπεδο στην οποία υποβάλλεται το κείμενο ούτως ώστε να διευκολυνθεί η διαδικασία της μηχανικής μετάφρασης. Σύμφωνα με τον Shuttleworth

‘Η συγγραφή ή η ολοκλήρωση της φάσης του pre-editing ενός κειμένου που προορίζεται για μηχανική μετάφραση έτσι ώστε να συμμορφώνεται στις παραμέτρους μιας ελεγχόμενης γλώσσας, ενισχύει και βελτιώνει σημαντικά την απόδοση των περισσότερων συστημάτων μηχανικής μετάφρασης, καθώς εξαλείφει πολλές από τις αμφισημίες που ένα κείμενο προέλευσης (Source Text) υπό άλλες συνθήκες θα περιείχε και τις οποίες ο υπολογιστής θα ήταν αδύνατον να αναλύσει και να επεξεργαστεί’.

Shuttleworth (1997: 20)

### **5.2 Post-editing**

Από την άλλη πλευρά, η διαδικασία του post-editing, σύμφωνα με τους Torrejón και Rico (2002: 107-116) είναι μια δραστηριότητα που σχετίζεται με τη Μηχανική Μετάφραση και συνδέεται άμεσα με την ανάγκη και τη ζήτηση που υπάρχει για πρόσβαση στην πληροφορία. Η υπάρχουσα ανάγκη για μεγάλο όγκο μεταφράσεων, όπως προαναφέρθηκε, ώθησε τη μεταφραστική βιομηχανία στη μηχανική μετάφραση και τα χαμηλής ποιότητας μεταφραστικά αποτελέσματα της μηχανικής μετάφρασης, με τη σειρά τους, ενίσχυσαν την ανάγκη για τη διαδικασία του post-editing. Σύμφωνα με τον Frank Austerhüll (2001: 166) στο βιβλίο του *Electronic Tools for Translators* «ακόμη και αν το κείμενο πηγή έχει υποβληθεί στη διαδικασία του pre-editing ή ακόμη κι αν είναι γραμμένο σε ελεγχόμενη γλώσσα, το κείμενο στόχος, πρέπει να υποβληθεί στη διαδικασία του post-editing στοχεύοντας στο καλύτερο μετάφρασμα».

## **6. Μελέτη & Παραδείγματα**

Ο εντοπισμός προβληματικών δομών στα κείμενα και η εφαρμογή των κανόνων ελεγχόμενης γλώσσας είχαν ως αποτέλεσμα να κατηγοριοποιηθούν οι δομές αυτές και να επιτευχθεί μέσα από διάφορες εφαρμογές το αποτέλεσμα-στόχος με τη χρήση του συστήματος μηχανικής μετάφρασης.

Σε προτάσεις όπως το συγκεκριμένο παράδειγμά: Press Menu – αν και η μετάφραση του για έναν άνθρωπο μεταφραστή θα ήταν αυτονόητη – παρατηρήθηκε πως ο υπολογιστής αδυνατεί να αποδώσει σωστά το νόημα. Το αποτέλεσμα της μετάφρασης από τον υπολογιστή είναι το ακόλουθο: Επιλογές Τύπου. Όπως παρατηρήθηκε ο υπολογιστής μπερδεύει την προστακτική του ρήματος press = πιέζω με το ουσιαστικό press = τύπος (έντυπα έγγραφα). Η κατηγορία αυτών των λέξεων που στα Ελληνικά ονομάζονται ομόγραφα (homographs) προκαλεί τεράστια σύγχυση στον υπολογιστή και καθιστά αδύνατη την ορθή μετάφραση της προστακτικής όταν αυτή συμπίπτει με τον αντίστοιχο τύπο του ουσιαστικού.

Σε περιπτώσεις όπως η προαναφερόμενη, η εφαρμογή των κανόνων μιας ελεγχόμενης γλώσσας καθιστά τον υπολογιστή ικανό να αποδώσει τον όρο με ακρίβεια. Στη συγκεκριμένη περίπτωση η λύση δόθηκε με την αποφυγή της χρήσης της προστακτικής. Αντί για προστακτική, ο μεταφραστής χρησιμοποίησε το βοηθητικό ρήμα should. Η πρόταση: You should press menu μεταφράζεται ως: Πρέπει να πιέσετε τις επιλογές. Όπως φαίνεται η σύγχυση των δυο μορφών δηλαδή του ρήματος και του ουσιαστικού από τον υπολογιστή δεν υπάρχει πλέον.

Στην προσπάθειά του το σύστημα να αποδώσει τις συντομογραφίες αποτυγχάνει. Καταγράφει λοιπόν τους λατινικούς χαρακτήρες που αποτελούν το αρκτικόλεξο ή τη συντομογραφία με τους αντίστοιχους ελληνικούς χαρακτήρες και όχι με την αντίστοιχη ελληνική συντομογραφία. Για παράδειγμα, C.N.S. > Κ.Ν.Σ., δηλαδή Κεντρικό Νευρικό Σύστημα, αν και σε αυτή την περίπτωση το αποτέλεσμα είναι το επιθυμητό επειδή οι τύποι συμπίπτουν αλλά σε περιπτώσεις όπως οι ακόλουθες το αποτέλεσμα δεν είναι το επιθυμητό: e-mail > ε-ταχυδρομείο και LPS > “λπς”. Η λύση στην προκειμένη περίπτωση είναι η αντικατάσταση της συντομογραφίας με τον πλήρη όρο δηλαδή, electronic mail και loopset αντίστοιχως.

## **7. Στοιχισμός κειμένων (Text Alignment)**

Μετά την εφαρμογή των κανόνων ελεγχόμενης γλώσσας στο σώμα κειμένων που συστάθηκε και την καταγραφή των αποτελεσμάτων, η μελέτη επικεντρώθηκε και σε

ένα άλλο ερώτημα: κατά πόσο η εφαρμογή των κανόνων της ελεγχόμενης γλώσσας εφόσον επιφέρει βελτίωση σε γλωσσολογικό επίπεδο θα μπορούσε να βελτιώσει και τα αποτελέσματα στοίχισης. Σύμφωνα με τους Oakes & McEnery (2000: 2):

«στόχος της στοίχισης των προτάσεων είναι να ανακαλύψουμε ακριβώς ποια πρόταση του κειμένου πηγής αντιστοιχεί με την πρόταση του κειμένου στόχου ή ποιες προτάσεις σε ένα κείμενο της γλώσσας A αντιστοιχούν σε ποια ή σε ποιες προτάσεις της γλώσσας B, όπου η γλώσσα A είναι το πρωτότυπο κείμενο και η γλώσσα B είναι η μετάφραση του κειμένου της γλώσσας A. Η στοίχιση των προτάσεων είναι συχνά ο προπομπός για τον εντοπισμό αντιστοιχιών μεταξύ των λέξεων μιας γλώσσας A και μιας B».

Oakes & McEnery (2000: 2):

Η στοίχιση, γενικότερα, «προϋποθέτει τη δημιουργία συνδυαστικών δεσμών μεταξύ δύο ή περισσότερων κειμένων» (Bowker & Pearson 2002: 95). Θα μπορούσε κάποιος να πει ότι στοίχιση είναι η διαδικασία όπου τα τμήματα του κειμένου πηγής στοιχίζονται με τα αντίστοιχα τους τμήματα στο κείμενο στόχος. Ιδιαίτερα, η διαδικασία στοίχισης προϋποθέτει τη σύγκριση και χρήση κειμένων τα οποία συνήθως αποτελούν αμοιβαίες μεταφράσεις. Η στοίχιση αποτελεί τη διάκριση των δύο κειμένων στα μέρη από τα οποία αποτελείται. Η διάκριση αυτή μπορεί να έχει πολλά επίπεδα. Τα κείμενα μπορούν να στοιχηθούν σε επίπεδα κειμένων, παραγράφων, προτάσεων, ακόμη και σε επίπεδο λέξεων ή χαρακτήρων. Ωστόσο, υπάρχουν δύο βασικές προσεγγίσεις της στοίχισης από υπολογιστική άποψη: α) η γλωσσολογική τεχνική που στηρίζεται σε πληροφορίες σχετικές με τις εμπλεκόμενες γλώσσες και β) η στατιστική τεχνική που χρησιμοποιεί ποσοτικά κριτήρια. Η τελευταία χρησιμοποιήθηκε για την εξακρίβωση της υπόθεσης ότι η εφαρμογή των κανόνων ελεγχόμενης γλώσσας βελτιώνει εκτός από το αποτέλεσμα της μηχανικής μετάφρασης και το αποτέλεσμα της στοίχισης.

### **7.1 Ανάλυση των αποτελεσμάτων στοίχισης μετά την εφαρμογή ελεγχόμενης γλώσσας**

Μετά τη στοίχιση των μεταφράσεων των εγχειριδίων με τη βοήθεια μηχανικής μετάφρασης στα οποία δεν είχαν εφαρμοστεί οι κανόνες της ελεγχόμενης γλώσσας, παρατηρήθηκε πως σε ένα σύνολο περίπου 1000 προτάσεων 101 από αυτές εμφάνισαν πρόβλημα στη στοίχιση. Οι προτάσεις αυτές οδήγησαν στη σύσταση ενός corpus 101 προτάσεων στις οποίες στη συνέχεια εφαρμόστηκαν οι κανόνες ελεγχόμενης γλώσσας με σκοπό τη βελτίωση του μεταφραστικού αποτελέσματος. Ωστόσο, σε αυτό το σημείο είναι σημαντικό να αναφερθεί πως το μέγεθος του δείγματος είναι περιορισμένο και πως ενδεχομένως ένα μεγαλύτερο δείγμα να μπορούσε να εγγυηθεί πιο ασφαλή αποτελέσματα.

Σε ένα σύνολο 101 προτάσεων, όπου τα πρώτα αποτελέσματα στοίχισης αποδείχθηκαν προβληματικά, 77 δηλαδή, 76.2 % μετά την εφαρμογή των κανόνων της ελεγχόμενης γλώσσας στοιχίστηκαν σωστά. Ωστόσο, 25 προτάσεις δηλαδή, 24.5% παρουσίασαν προβλήματα στη στοίχισή τους τα οποία κατά ένα μεγάλο ποσοστό οφείλονται στην κακή στίξη. Σχετικά με τη στοίχιση κειμένων (text alignment), παρατηρήθηκε η κατά 76,2% μείωση του ποσοστού λάθους σε επίπεδο

στοίχισης με την εφαρμογή κανόνων ελεγχόμενης γλώσσας σε κείμενα μεταφρασμένα με Μηχανική Μετάφραση. Όπως γίνεται κατανοητό οι ελεγχόμενες γλώσσες σε συνδυασμό με τη μηχανική μετάφραση έχουν επιπλέον εφαρμογές πέραν των αμιγώς γλωσσολογικών. Επιδίωξη αυτής της προσπάθειας ήταν να ερευνηθεί η πιθανότητα ότι ένα υψηλότερης ποιότητας μεταφραστικό αποτέλεσμα δύναται να επιφέρει καλύτερα αποτελέσματα και μεγαλύτερο ποσοστό επιτυχίας σε επίπεδο στοίχισης.

## 8. Συνοψίζοντας

Η χρήση της μηχανικής μετάφρασης ακόμη αμφισβητείται και ο λόγος είναι πως συνεχίζει να παρουσιάζει προβλήματα ως προς το αποτέλεσμά της σε επίπεδο γραμματικό, συντακτικό και λεξιλογικό. Σε καμία περίπτωση όμως δεν μπορεί κάποιος να πει πως οι προβληματικές συντακτικογραμματικές δομές που παρατηρούνται στα κείμενα που μεταφράζονται με τη βοήθεια συστημάτων μηχανικής μετάφρασης καθιστούν τη μηχανική μετάφραση άχρηστη ή τα αποτελέσματά της δεν επιδέχονται βελτίωσης.

Η βελτίωση του μεταφραστικού αποτελέσματος που προήλθε από τη χρήση συστημάτων μηχανικής μετάφρασης επήλθε με τη χρήση και εφαρμογή των ελεγχόμενων γλωσσών, οι οποίες καλυτερεύουν αισθητά τα αποτελέσματα της μετάφρασης. Σαφώς, παραμένει αναγκαία η φάση του pre-editing – της οποίας αναπόσπαστο μέρος είναι η εφαρμογή των κανόνων μιας ελεγχόμενης γλώσσας – και του post-editing, με σκοπό τον εντοπισμό τυχόν προβληματικών δομών μετά την ολοκλήρωση της διαδικασίας της μετάφρασης. Με τη μηχανική μετάφραση και τις ελεγχόμενες γλώσσες έχουμε εξοικονόμηση χρόνου, εξοικονόμηση χρήματος και μεγαλύτερη παραγωγή και όγκο μετάφρασης. Ο συνδυασμός της μηχανικής μετάφρασης και της ελεγχόμενης γλώσσας προκύπτει έπειτα από σχετική έρευνα πως εκτός από βελτίωση στο μεταφραστικό αποτέλεσμα συμβάλλει και στη βελτίωση των αποτελεσμάτων της στοίχισης των κειμένων και στη δημιουργία παράλληλων σωμάτων κειμένων (parallel corpora).

## Βιβλιογραφικές αναφορές

- Allen, J., “CL15: Different Types of Controlled Languages”, <http://www.tc-forum.org/topiccl/cl15diff.htm> [Last accessed 13/06/06]
- Arnold, D., Balkan, L., Meijer, S., Humphreys, R. Lee, Sadler, L. (1994) *Machine Translation: An Introductory Guide* <http://www.essex.ac.uk/linguistics/clmt/MTBook/> [Last accessed 16/06/06]
- Austermühl, F., (2001) *Electronic Tools for Translators*, Manchester: St Jerome Publishing, p.166
- Bowker, L., Pearson, J., (2002) *Working with Specialized Language*, London: Routledge, p.95
- Esselink, B., (2000) *A Practical Guide to Localization*, John Benjamins: Amsterdam/Philadelphia
- Kittredge, R. I., (2003) ‘Sublanguages and Controlled Languages’, in Ruslan Mitkov (ed.) *Oxford Handbook of Computational Linguistics*, Oxford: OUP, p.p. 441-442
- Markantonatou, S., Maistros, Y., Vassiliou, M., Karkaletsis, V., (2001) *Controlled Greek: Specifications and Implementation*, To be included in the Proceedings of the 5th International Conference in Greek Linguistics, Université Paris 5, Sorbonne, 13-15 September 2001
- Markantonatou, S., Maistros, Y., Vassiliou, M., Karkaletsis, V., (2002) *Evaluating Specifications for Controlled Greek*, To be included in the Proceedings of EAMT/CLAW 2003, Dublin, Ireland, 15-17 May 2003



Oakes, M., McEnery, T., (2000), 'Bilingual Text alignment – an overview', in Simon Philip Botley, Anthony Mark McEnery and Andrew Wilson (eds.) *Multilingual Corpora in Teaching and Research*, Amsterdam - Atlanta, GA 2000. pp. 1-37.

O'Brien, S., *Controlling Controlled English: An Analysis of Several Controlled Language Rule Sets*, in "Controlled Language Translation", Proceedings of EAMT-CLAW 2003, Dublin City University, 15<sup>th</sup> – 17<sup>th</sup> May 2003, p.p.105-114

Somers, H. (2003), 'Machine Translation: Latest developments' in Ruslan Mitkov (ed.) *Oxford Handbook of Computational Linguistics*, Oxford: OUP, p.p.518-519)

Shuttleworth M., Cowie M., (1997), *Dictionary of Translation Studies*, Manchester: St Jerome Publishing, p.20

Torrejon, E., Rico, C., (2002) "Controlled translation: A new teaching scenario tailor-made for the translation industry", [6th EAMT Workshop Teaching Machine Translation, Manchester](#), 107-116

## Υλικό της έρευνας

Nokia 2270

[http://www.google.com/search?hs=IJc&hl=el&client=firefox-a&rls=org.mozilla%3Ael%3Aofficial\\_s&as\\_qdr=all&q=Nokia+6200+filetype%3Adoc&btnG=%CE%91%CE%BD%CE%B1%CE%B6%CE%AE%CF%84%CE%B7%CF%83%CE%B7&lr=](http://www.google.com/search?hs=IJc&hl=el&client=firefox-a&rls=org.mozilla%3Ael%3Aofficial_s&as_qdr=all&q=Nokia+6200+filetype%3Adoc&btnG=%CE%91%CE%BD%CE%B1%CE%B6%CE%AE%CF%84%CE%B7%CF%83%CE%B7&lr=)

Nokia 6200

[http://www.google.com/search?hl=el&client=firefox-a&rls=org.mozilla%3Ael%3Aofficial\\_s&as\\_qdr=all&q=Configuration+of+Sony+Ericsson+S700i+to+Access&btnG=%CE%91%CE%BD%CE%B1%CE%B6%CE%AE%CF%84%CE%B7%CF%83%CE%B7&lr](http://www.google.com/search?hl=el&client=firefox-a&rls=org.mozilla%3Ael%3Aofficial_s&as_qdr=all&q=Configuration+of+Sony+Ericsson+S700i+to+Access&btnG=%CE%91%CE%BD%CE%B1%CE%B6%CE%AE%CF%84%CE%B7%CF%83%CE%B7&lr)

[Configuration of Sony Ericsson T230 to Access MTNL on GPRS](#)

[Configuration of Sony Ericsson P900 to Access MTNL on GPRS](#)

[Configuration of Sony Ericsson S700i to Access MTNL on GPRS](#)

[Configuration of Sony Ericsson S700i to Access MTNL on GPRS](#)

[Configuration of Sony Ericsson P910i to Access MTNL on GPRS](#)

[http://www.google.com/search?hs=ce7&hl=el&client=firefox-a&rls=org.mozilla%3Ael%3Aofficial\\_s&q=wap\\_gprs+Configuration&btnG=%CE%91%CE%BD%CE%B1%CE%B6%CE%AE%CF%84%CE%B7%CF%83%CE%B7&lr](http://www.google.com/search?hs=ce7&hl=el&client=firefox-a&rls=org.mozilla%3Ael%3Aofficial_s&q=wap_gprs+Configuration&btnG=%CE%91%CE%BD%CE%B1%CE%B6%CE%AE%CF%84%CE%B7%CF%83%CE%B7&lr)

Configuration of Siemens S45/ ME45

Configuration of Samsung S500 / E700 / E100

Configuration of Nokia 3510 / 3300

Configuration of Sony Ericsson T610 / T630 / Z200 / Z600

Configuration of Ericsson T65/ T200 to Access GPRS

Configuration of Sony Ericsson K700i & K500i to Access GPRS

Configuration of Sony Ericsson S700i to Access wap over GPRS

Configuration of Samsung S300 / V200 / S200 / C100 / S300M / X430 / P400

Configuration of Nokia 8310/ 6510/ 8910/ 7210/ 6610/ 6100/ 3530 / 7250i / 6108 / 3100 / 3120 / 3200 / 3108 to Access GPRS