2

ΓΛΩΣΣΕΣ ΣΕ ΚΙΝΔΥΝΟ: ΜΙΑ ΠΡΟΚΛΗΣΗ ΓΙΑ ΤΗ ΓΛΩΣΣΙΚΗ ΤΕΧΝΟΛΟΓΙΑ

Είμαστε μάρτυρες μιας ψηφιακής επανάστασης η οποία επηρεάζει δραματικά την επικοινωνία και την κοινωνία. Οι πρόσφατες εξελίξεις στην ψηφιακή τεχνολογία των πληροφοριών και των επικοινωνιών αρκετές φορές συγκρίνονται με την εφεύρεση της τυπογραφίας από τον Γουτεμβέργιο. Τι μας λέει αυτή η αναλογία για το μέλλον της ευρωπαϊκής κοινωνίας της πληροφορίας και ειδικότερα για το μέλλον των γλωσσών μας;

> Σήμερα είμαστε μάρτυρες μιας ψηφιακής επανάστασης που μπορεί να συγκριθεί με την εφεύρεση της τυπογραφίας από τον Γουτεμβέργιο.

Μετά την εφεύρεση του Γουτεμβέργιου επιτεύχθηκαν πραγματικές καινοτομίες στην επικοινωνία και την ανταλλαγή γνώσεων με προσπάθειες όπως η μετάφραση της Βίβλου στην καθομιλουμένη από τον Λούθηρο. Στους αιώνες που ακολούθησαν, αναπτύχθηκαν πολιτισμικές τεχνικές για την καλύτερη προσέγγιση της επεξεργασίας του λόγου και της ανταλλαγής γνώσεων:

- Η ορθογραφική και γραμματική τυποποίηση ευρέως διαδεδομένων γλωσσών επέτρεψε την ταχεία διάδοση νέων επιστημονικών γνώσεων και ιδεών.
- Η ανάπτυξη επίσημων γλωσσών κατέστησε δυνατή την επικοινωνία των πολιτών εντός ορισμένων (συχνά πολιτικών) συνόρων.
- Η διδασκαλία και η μετάφραση γλωσσών επέτρεψε διαγλωσσικές ανταλλαγές.
- Η δημιουργία εκδοτικών και βιβλιογραφικών οδηγιών διασφάλισε την ποιότητα και τη διαθεσιμότητα έντυπου υλικού.
- Η δημιουργία διαφορετικών μέσων όπως οι εφημερίδες, το ραδιόφωνο, η τηλεόραση, τα βιβλία κ. ά. ικανοποίησε διάφορες επικοινωνιακές ανάγκες.

Τα τελευταία είκοσι χρόνια, η πληροφορική έχει βοηθήσει στην αυτοματοποίηση και τη διευκόλυνση πολλών διαδικασιών:

- οι ηλεκτρονικές εκδόσεις έχουν αντικαταστήσει την δακτυλογράφηση και τη στοιχειοθεσία,
- το Microsoft PowerPoint έχει αντικαταστήσει τον προβολέα διαφανειών,
- το ηλεκτρονικό ταχυδρομείο στέλνει και λαμβάνει έγγραφα ταχύτερα και από την τηλεομοιοτυπία (fax),
- το Skype προσφέρει οικονομικές τηλεφωνικές κλήσεις μέσω Διαδικτύου και υποστηρίζει εικονικές συσκέψεις,
- τα μορφότυπα κωδικοποίησης ήχου και βίντεο διευκολύνουν την ανταλλαγή πολυμεσικού περιεχομένου,
- οι μηχανές αναζήτησης προσφέρουν πρόσβαση σε ιστοσελίδες βασιζόμενες σε λέξεις κλειδιά,
- διαδικτυακές υπηρεσίες όπως το Google Translate παράγουν γρήγορες, κατά προσέγγιση μεταφράσεις,
- οι πλατφόρμες κοινωνικής δικτύωσης όπως το Facebook, το Twitter και το Google+ διευκολύνουν την επικοινωνία, τη συνεργασία και την ανταλλαγή πληροφοριών.

Αν και αυτά τα εργαλεία και οι εφαρμογές είναι χρήσιμα, δεν είναι ακόμα ικανά να υποστηρίξουν μια βιώσιμη, πολυγλωσσική ευρωπαϊκή κοινωνία για όλους, όπου οι πληροφορίες και τα αγαθά θα μπορούν να διακινούνται ελεύθερα.

2.1 ΓΛΩΣΣΙΚΑ ΣΥΝΟΡΑ: ΕΜΠΟΔΙΟ ΣΤΗΝ ΕΥΡΩΠΑΪΚΗ ΚΟΙΝΩΝΙΑ ΤΗΣ ΠΛΗΡΟΦΟΡΙΑΣ

Δεν είμαστε σε θέση να προβλέψουμε πώς ακριβώς θα μοιάζει η μελλοντική κοινωνία της πληροφορίας. Υπάρχει όμως μεγάλη πιθανότητα η επανάσταση στην τεχνολογία επικοινωνιών να φέρει κοντά ανθρώπους που μιλάνε διαφορετικές γλώσσες, με νέους τρόπους. Το γεγονός αυτό ωθεί τους ανθρώπους προς την εκμάθηση νέων γλωσσών και ασκεί πίεση στους προγραμματιστές να δημιουργήσουν νέες τεχνολογικές εφαρμογές που να εξασφαλίζουν την αμοιβαία κατανόηση και την πρόσβαση σε διαμοιραζόμενη γνώση. Στην παγκόσμια οικονομία και στον παγκόσμιο χώρο πληροφοριών, περισσότερες γλώσσες, ομιλητές και περιεχόμενο αλληλεπιδρούν ταχύτερα με νέους τύπους μέσων. Η τρέχουσα δημοτικότητα των κοινωνικών μέσων (Wikipedia, Facebook, Twitter, YouTube, και προσφάτως το Google+) είναι μόνον η κορυφή του παγόβουνου.

> Η παγκόσμια οικονομία και ο ενιαίος χώρος πληροφοριών μας φέρνει αντιμέτωπους με περισσότερες γλώσσες, ομιλητές και περιεχόμενο.

Σήμερα μπορούμε να μεταδίδουμε gigabytes κειμένου σε ολόκληρο τον πλανήτη μέσα σε λίγα δευτερόλεπτα προτού αντιληφθούμε ότι αφορά μια γλώσσα που δεν καταλαβαίνουμε. Σύμφωνα με μια πρόσφατη έκθεση της Ευρωπαϊκής Επιτροπής, το 57% των χρηστών του Διαδικτύου στην Ευρώπη αγοράζουν εμπορεύματα και υπηρεσίες χρησιμοποιώντας γλώσσες οι οποίες δεν είναι οι μητρικές τους (τα αγγλικά είναι η πιο διαδεδομένη ξένη γλώσσα κι ακολουθούν τα γαλλικά, τα γερμανικά και τα ισπανικά). Το 55% των χρηστών διαβάζει περιεχόμενο σε κάποια ξένη γλώσσα, ενώ μόλις το 35% χρησιμοποιεί άλλη γλώσσα για να γράψει ηλεκτρονικά μηνύματα ή να κάνει σχόλια στο Διαδίκτυο [3]. Πριν από λίγα χρόνια, τα αγγλικά ίσως ήταν η lingua franca του Διαδικτύου — η πλειονότητα του περιεχομένου στο Διαδίκτυο ήταν στα αγγλικά — αλλά η κατάσταση έχει πλέον αλλάξει δραματικά. Η ποσότητα του διαδικτυακού περιεχομένου σε άλλες ευρωπαϊκές γλώσσες (καθώς και σε ασιατικές και μεσανατολικές) έχει υπερπολλαπλασιαστεί. Προκαλεί έκπληξη το γεγονός ότι αυτό το πανταχού παρόν ψηφιακό χάσμα λόγω των γλωσσικών συνόρων δεν έχει προσελκύσει ιδιαίτερα την προσοχή, παρόλο που θέτει ένα πολύ πιεστικό ερώτημα: ποιες ευρωπαϊκές γλώσσες θα κατορθώσουν να επιβιώσουν στη δικτυωμένη κοινωνία της πληροφορίας και της γνώσης και ποιες είναι καταδικασμένες να εξαφανιστούν;

2.2 ΟΙ ΓΛΩΣΣΕΣ ΜΑΣ ΚΙΝΔΥΝΕΥΟΥΝ

Παρόλο που η τυπογραφία βοήθησε στην ενίσχυση της ανταλλαγής πληροφοριών στην Ευρώπη, οδήγησε επίσης στον αφανισμό πολλών ευρωπαϊκών γλωσσών. Οι περιφερειακές και μειονοτικές γλώσσες σπανίως τυπώνονταν και γλώσσες όπως τα κορνουαλικά και τα δαλματικά περιορίστηκαν σε προφορικές μορφές μετάδοσης, οι οποίες με τη σειρά τους περιόρισαν το πεδίο χρήσης τους. Θα έχει και το Διαδίκτυο τις ίδιες επιπτώσεις στις γλώσσες μας;

Οι περίπου 80 γλώσσες της Ευρώπης είναι ένα από τα πολυτιμότερα και σημαντικότερα πολιτιστικά της περιουσιακά στοιχεία, καθώς και ζωτικό κομμάτι του μοναδικού της κοινωνικού μοντέλου [4]. Ενώ γλώσσες όπως τα αγγλικά και τα ισπανικά είναι πιθανότερο να επιβιώσουν στην αναδυόμενη ψηφιακή αγορά, πολλές ευρωπαϊκές γλώσσες θα μπορούσαν να καταλήξουν ήσσονος σημασίας σε μια διαδικτυωμένη κοινωνία. Αυτό θα αποδυνάμωνε την παγκό-

σμια θέση της Ευρώπης και θα εναντιωνόταν στον στρατηγικό στόχο της διασφάλισης της ίσης συμμετοχής κάθε Ευρωπαίου πολίτη ανεξαρτήτως γλώσσας.

> Η μεγάλη ποικιλία γλωσσών στην Ευρώπη είναι ένα από τα πολυτιμότερα και σημαντικότερα πολιτισμικά περιουσιακά της στοιχεία.

Σύμφωνα με μια έκθεση της UNESCO για την πολυγλωσσία, οι γλώσσες αποτελούν ένα ουσιαστικό μέσο για την απόλαυση θεμελιωδών δικαιωμάτων, όπως η πολιτική έκφραση, η εκπαίδευση και η συμμετοχή στην κοινωνία [5].

2.3 ΓΛΩΣΣΙΚΗ ΤΕΧΝΟΛΟΓΙΑ: ΜΙΑ ΒΑΣΙΚΗ ΤΕΧΝΟΛΟΓΙΑ ΠΡΟΣΒΑΣΙΜΟΤΗΤΑΣ

Στο παρελθόν οι επενδυτικές προσπάθειες για τη διατήρηση των γλωσσών επικεντρώνονταν στη γλωσσική εκπαίδευση και τη μετάφραση. Σύμφωνα με μια εκτίμηση, η ευρωπαϊκή αγορά μετάφρασης, διερμηνείας, λογισμικών τοπικοποίησης (localisation) και παγκοσμιοποίησης δικτυακών τόπων (website globalisation) ανερχόταν σε 8,4 δισεκατομμύρια ευρώ το 2008 με ρυθμό ανάπτυξης 10% ετησίως [6]. Κι όμως αυτός ο αριθμός καλύπτει ένα πολύ μικρό ποσοστό των τρεχουσών και των μελλοντικών αναγκών διαγλωσσικής επικοινωνίας. Η πιο πειστική λύση για τη διασφάλιση του εύρους και του βάθους της χρήσης της γλώσσας στην Ευρώπη του αύριο είναι η χρήση της κατάλληλης τεχνολογίας, ακριβώς όπως χρησιμοποιούμε τεχνολογία για να λύσουμε, μεταξύ άλλων, τις ανάγκες μας για μεταφορά, ενέργεια και πρόσβαση.

Η γλωσσική τεχνολογία (που στοχεύει σε κάθε μορφή γραπτού κειμένου και προφορικού λόγου) βοηθά τους ανθρώπους να συνεργάζονται, να συναλλάσσονται, να μοιράζονται γνώσεις και να συμμετέχουν στον κοινωνικό και πολιτικό διάλογο ανεξάρτητα από γλωσσικούς φραγμούς και δεξιότητες χρήσης υπολογιστή. Συχνά λειτουργεί αόρατα μέσα σε σύνθετα συστήματα λογισμικού για να μας βοηθήσει:

- να βρούμε πληροφορίες με μια μηχανή αναζήτησης,
- να ελέγξουμε την ορθογραφία και τη γραμματική σε έναν επεξεργαστή κειμένου,
- να δούμε συστάσεις για προϊόντα σε ένα διαδικτυακό κατάστημα,
- να ακούσουμε φωνητικές οδηγίες από ένα σύστημα πλοήγησης αυτοκινήτου,
- να μεταφράσουμε ιστοσελίδες μέσω μιας διαδικτυακής υπηρεσίας.

Η γλωσσική τεχνολογία απαρτίζεται από πολλές βασικές εφαρμογές που καθιστούν δυνατές διαδικασίες στο πλαίσιο μιας μεγαλύτερης εφαρμογής. Ο στόχος των Λευκών Βίβλων του META-NET για τη γλώσσα είναι να εστιάσουν στο πόσο έτοιμες είναι αυτές οι βασικές τεχνολογίες για κάθε ευρωπαϊκή γλώσσα.

> Η Ευρώπη χρειάζεται αξιόπιστη και οικονομική Γλωσσική Τεχνολογία για όλες τις ευρωπαϊκές γλώσσες.

Για να διατηρήσουμε τη θέση μας στην πρώτη γραμμή της παγκόσμιας καινοτομίας, η Ευρώπη θα χρειαστεί γλωσσική τεχνολογία προσαρμοσμένη σε όλες τις ευρωπαϊκές γλώσσες, η οποία θα είναι αξιόπιστη, οικονομική και ολοκληρωμένη σε βασικά περιβάλλοντα λογισμικού. Χωρίς γλωσσική τεχνολογία δεν θα κατορθώσουμε στο προσεχές μέλλον να προσφέρουμε μια πραγματικά αποτελεσματική, διαδραστική, πολυμεσική και πολυγλωσσική εμπειρία στον χρήστη.

2.4 ΕΥΚΑΙΡΙΕΣ ΓΙΑ ΤΗ ΓΛΩΣΣΙΚΗ ΤΕΧΝΟΛΟΓΙΑ

Στον κόσμο της τυπογραφίας, τεχνολογική καινοτομία αποτέλεσε η γρήγορη αναπαραγωγή μιας εικόνας ενός κειμένου (σελίδας) χρησιμοποιώντας ένα κατάλληλο μηχανοκίνητο τυπογραφικό πιεστήριο. Άνθρωποι καλούνταν να επιτελέσουν το δύσκολο έργο της έρευνας, του διαβάσματος, της μετάφρασης και της συνοπτικής παρουσίασης της γνώσης. Χρειάστηκε να περιμένουμε μέχρι τον Έντισον για να καταγράψουμε τον προφορικό λόγο – και πάλι η τεχνολογία του απλά παρήγαγε αναλογικά αντίγραφα.

Η γλωσσική τεχνολογία μπορεί πλέον να αυτοματοποιήσει τις ίδιες τις διεργασίες της μετάφρασης, της παραγωγής περιεχομένου και της διαχείρισης γνώσης για όλες τις ευρωπαϊκές γλώσσες. Μπορεί επίσης να εμπλουτίσει οικιακά ηλεκτρονικά συστήματα, μηχανήματα, οχήματα, υπολογιστές και ρομπότ με διεπαφές βασισμένες σε γραπτό ή προφορικό λόγο. Οι εμπορικές και βιομηχανικές εφαρμογές βρίσκονται ακόμη σε αρχικά στάδια ανάπτυξης, αλλά τα επιτεύγματα της Έρευνας & Ανάπτυξης δημιουργούν πραγματικές ευκαιρίες. Για παράδειγμα, η μηχανική μετάφραση είναι ήδη σχετικά ακριβής σε συγκεκριμένους τομείς. Επίσης, υπάρχουν πειραματικές εφαρμογές που προσφέρουν πολυγλωσσικές πληροφορίες και διαχείριση γνώσης, καθώς και παραγωγή περιεχομένου σε πολλές ευρωπαϊκές γλώσσες.

Όπως συμβαίνει με τις περισσότερες τεχνολογίες, οι πρώτες γλωσσικές εφαρμογές, όπως οι φωνητικές διεπαφές χρήστη και τα διαλογικά συστήματα, αναπτύχθηκαν για πολύ εξειδικευμένους τομείς και συχνά παρουσιάζουν περιορισμένη απόδοση. Υπάρχουν όμως τεράστιες επιχειρηματικές ευκαιρίες στον τομέα της εκπαίδευσης και της ψυχαγωγίας σχετικά με την ολοκλήρωση γλωσσικών τεχνολογιών σε παιχνίδια, χώρους πολιτιστικής κληρονομιάς, ψυχαγωγικά εκπαιδευτικά πακέτα, βιβλιοθήκες, περιβάλλοντα προσομοίωσης και προγράμματα επιμόρφωσης. Υπηρεσίες ενημέρωσης κινητής τηλεφωνίας, λογισμικό εκμάθησης γλωσσών μέσω Η/Υ, περιβάλλοντα εξ αποστάσεως μάθησης, εργαλεία αυτο-αξιολόγησης και λογισμικό εντοπισμού λογοκλοπής είναι μόνο μερικοί από τους τομείς εφαρμογής όπου η γλωσσική τεχνολογία μπορεί να διαδραματίσει σημαντικό ρόλο. Η δημοτικότητα των εφαρμογών κοινωνικών μέσων όπως το Twitter και το Facebook δείχνουν μια περαιτέρω ανάγκη για προηγμένες γλωσσικές τεχνολογίες που να μπορούν να παρακολουθούν τις αναρτήσεις, να συνοψίζουν συζητήσεις, να αναδεικνύουν τις τάσεις της κοινής γνώμης, να ανιχνεύουν συναισθηματικές αντιδράσεις, να εντοπίζουν παραβιάσεις πνευματικών δικαιωμάτων ή να ανιχνεύουν παράνομες χρήσεις.

> Η γλωσσική τεχνολογία βοηθά στην υπέρβαση της "αναπηρίας" που επιφέρει η γλωσσική ποικιλότητα.

Η γλωσσική τεχνολογία αποτελεί τεράστια ευκαιρία για την Ευρωπαϊκή Ένωση. Μπορεί να βοηθήσει στην αντιμετώπιση του σύνθετου ζητήματος της πολυγλωσσίας στην Ευρώπη – το γεγονός ότι διαφορετικές γλώσσες συνυπάρχουν φυσικά σε ευρωπαϊκές επιχειρήσεις, οργανισμούς και σχολεία. Αλλά οι πολίτες χρειάζεται να επικοινωνούν υπερπηδώντας τα γλωσσικά σύνορα, καθώς διασχίζουν από άκρη σε άκρη την Ευρωπαϊκή Κοινή Αγορά, και η γλωσσική τεχνολογία μπορεί να βοηθήσει στην υπέρβαση αυτού του τελευταίου φραγμού, υποστηρίζοντας παράλληλα την ελεύθερη και απρόσκοπτη χρήση των διαφόρων γλωσσών. Κοιτάζοντας ακόμα πιο μπροστά, η καινοτόμος ευρωπαϊκή πολύγλωσση γλωσσική τεχνολογία θα αποτελέσει σημείο αναφοράς για τους παγκόσμιους εταίρους μας, όταν θα ξεκινήσουν να οργανώνουν τις δικές τους πολύγλωσσες κοινότητες. Η γλωσσική τεχνολογία μπορεί να θεωρηθεί ως μια μορφή "υποστηρικτικής" τεχνολογίας που βοηθά στην υπέρβαση της "αναπηρίας" της γλωσσικής ποικιλότητας και κάνει τις γλωσσικές κοινότητες πιο προσβάσιμες τη μία στην άλλη.

Τέλος, ένα δυναμικό πεδίο έρευνας είναι η χρήση της γλωσσικής τεχνολογίας σε επιχειρήσεις διάσωσης σε περιοχές

καταστροφών, όπου το ζήτημα της επίδοσης ενός συστήματος μπορεί να είναι ζήτημα ζωής και θανάτου: μελλοντικά ευφυή ρομπότ με διαγλωσσικές ικανότητες θα είναι σε θέση να σώζουν ζωές.

2.5 ΠΡΟΚΛΗΣΕΙΣ ΠΟΥ ΑΝΤΙΜΕΤΩΠΙΖΕΙ Η ΓΛΩΣΣΙΚΗ ΤΕΧΝΟΛΟΓΙΑ

Αν και η γλωσσική τεχνολογία έχει σημειώσει σημαντική πρόοδο τα τελευταία χρόνια, ο τρέχων ρυθμός τεχνολογικής προόδου και καινοτομίας είναι πολύ αργός. Ευρέως διαδεδομένες τεχνολογίες όπως οι ορθογραφικοί και γραμματικοί διορθωτές σε κειμενογράφους είναι συνήθως μονόγλωσσοι και είναι διαθέσιμοι μόνο για λίγες γλώσσες.

Ο τρέχων ρυθμός της τεχνολογικής προόδου είναι πολύ αργός.

Οι διαδικτυακές υπηρεσίες μηχανικής μετάφρασης, παρ' ότι είναι σε θέση να παραγάγουν σχετικά αποδεκτή προσέγγιση του περιεχομένου ενός εγγράφου, βρίθουν προβλημάτων όταν απαιτούνται μεταφράσεις υψηλής ακρίβειας και πληρότητας. Εξαιτίας της πολυπλοκότητας της ανθρώπινης γλώσσας, η μοντελοποίηση των γλωσσών μας και η δοκιμή του μοντέλου στον πραγματικό κόσμο είναι μια μακρά, δαπανηρή υπόθεση που απαιτεί δεσμεύσεις συνεχούς χρηματοδότησης. Η Ευρώπη πρέπει επομένως να διατηρήσει τον πρωτοποριακό της ρόλο στην αντιμετώπιση των τεχνολογικών προκλήσεων μιας πολύγλωσσης κοινότητας, επινοώντας νέες μεθόδους για την επιτάχυνση της ανάπτυξης από τη μία άκρη του χάρτη ως την άλλη. Αυτές θα μπορούσαν να περιλαμβάνουν τόσο τις εξελίξεις στην επιστήμη των υπολογιστών όσο και τεχνικές όπως είναι ο πληθοπορισμός (crowdsourcing).

2.6 ΚΑΤΑΚΤΗΣΗ ΤΗΣ ΓΛΩΣΣΑΣ ΑΠΟ ΑΝΘΡΩΠΟΥΣ ΚΑΙ ΜΗΧΑΝΗΜΑΤΑ

Για να απεικονίσουμε το πώς οι υπολογιστές χειρίζονται τη γλώσσα και γιατί είναι δύσκολο να τους προγραμματίσουμε ώστε να την χρησιμοποιούν, ας ρίξουμε μια γρήγορη ματιά στον τρόπο που οι άνθρωποι μαθαίνουν την πρώτη και τη δεύτερη γλώσσα, κι έπειτα θα εξετάσουμε το πώς λειτουργούν τα συστήματα γλωσσικής τεχνολογίας.

Ο άνθρωπος κατακτά γλωσσικές δεξιότητες με δύο διαφορετικούς τρόπους: μαθαίνοντας παραδείγματα και μαθαίνοντας τους υποκείμενους γλωσσικούς κανόνες.

Οι άνθρωποι κατακτούν γλωσσικές δεξιότητες με δύο διαφορετικούς τρόπους. Τα μωρά μαθαίνουν μια γλώσσα ακούγοντας τους γονείς τους, τα αδέρφια τους και άλλα μέλη της οικογένειάς τους να μιλάνε σε πραγματικές συνθήκες. Από την ηλικία περίπου των δύο ετών, τα παιδιά λένε τις πρώτες τους λέξεις και σχηματίζουν μικρές φράσεις. Αυτό είναι εφικτό μόνο επειδή ο άνθρωπος έχει γενετική προδιάθεση να μιμείται κι έπειτα να εκλογικεύει τα όσα ακούει.

Η εκμάθηση μιας δεύτερης γλώσσας σε μεγαλύτερη ηλικία απαιτεί μεγαλύτερη προσπάθεια, κυρίως επειδή το παιδί δεν περιβάλλεται από μια γλωσσική κοινότητα φυσικών ομιλητών. Στο σχολείο οι ξένες γλώσσες συνήθως διδάσκονται μέσω γραμματικής, λεξιλογίου και ορθογραφίας με τη χρήση ασκήσεων που περιγράφουν τη γλωσσική γνώση εκπεφρασμένη σε αφηρημένους κανόνες, πίνακες και παραδείγματα.

Οι δύο κύριοι τύποι συστημάτων γλωσσικής τεχνολογίας "κατακτούν" γλωσσικές ικανότητες με παρόμοιο τρόπο. Οι στατιστικές προσεγγίσεις (ή "βασισμένες σε δεδομένα") αποκτούν γλωσσική γνώση από τεράστιες συλλογές δειγμάτων πραγματωμένου λόγου. Ενώ αρκεί η χρήση κειμένου σε μια μόνον γλώσσα για την εκπαίδευση, π. χ. ενός ορ-

θογράφου, απαιτούνται παράλληλα κείμενα σε δύο (ή περισσότερες) γλώσσες για την εκπαίδευση ενός συστήματος μηχανικής μετάφρασης. Ο αλγόριθμος μηχανικής μάθησης "μαθαίνει" τρόπους μετάφρασης λέξεων, σύντομων φράσεων και ολόκληρων προτάσεων.

Αυτή η στατιστική προσέγγιση απαιτεί εκατομμύρια προτάσεων και η ποιότητα της επίδοσής της αυξάνεται ανάλογα με την όγκο των κειμένων που έχει αναλύσει. Αυτός είναι ο λόγος για τον οποίο οι πάροχοι των μηχανών αναζήτησης ενδιαφέρονται να συγκεντρώσουν όσο περισσότερο γραπτό υλικό μπορούν. Ο ορθογραφικός έλεγχος σε κειμενογράφους αλλά και υπηρεσίες όπως το Google Search και το Google Translate βασίζονται σε στατιστικές προσεγγίσεις. Το μεγάλο πλεονέκτημα της στατιστικής είναι ότι η μηχανή μαθαίνει γρήγορα μέσα από διαρκείς σειρές αλλεπάλληλων κύκλων εκμάθησης, αν και η ποιότητα μπορεί να ποικίλλει.

Η δεύτερη προσέγγιση της γλωσσικής τεχνολογίας, και συγκεκριμένα της μηχανικής μετάφρασης, είναι η δημιουργία συστημάτων βασισμένων σε κανόνες. Ειδικοί στους τομείς της γλωσσολογίας, της υπολογιστικής γλωσσολογίας και της πληροφορικής καλούνται πρώτα να κωδικοποιήσουν γραμματικές αναλύσεις (κανόνες μετάφρασης) και να συγκεντρώσουν καταλόγους λεξιλογίου (λεξικά). Αυτό είναι εξαιρετικά χρονοβόρο και απαιτεί εντατική εργασία. Ορισμένα από τα καλύτερα συστήματα μηχανικής μετάφρασης βάσει κανόνων αναπτύσσονται εδώ και πάνω από μία εικοσαετία. Το μεγάλο πλεονέκτημα των συστημάτων αυτών είναι ότι οι ειδικοί έχουν ουσιαστικό έλεγχο στην επεξεργασία της γλώσσας. Αυτό δίνει τη δυνατότητα να διορθώνονται συστηματικά τα λάθη στο λογισμικό και να υπάρχει αναλυτικός σχολιασμός προς τον χρήστη, ειδικά όταν συστήματα βασισμένα σε κανόνες χρησιμοποιούνται για την εκμάθηση γλώσσας. Εξαιτίας, ωστόσο, του υψηλού κόστους αυτής της διαδικασίας, η βασισμένη σε κανόνες τεχνολογία προς το παρόν έχει αναπτυχθεί μόνο για τις περισσότερο χρησιμοποιούμενες γλώσσες.

Καθώς τα πλεονεκτήματα και οι αδυναμίες των στατιστικών και των βασισμένων σε κανόνες συστημάτων τείνουν να αλληλοσυμπληρώνονται, η έρευνα εστιάζει πλέον σε υβριδικές προσεγγίσεις που συνδυάζουν τις δύο μεθοδολογίες. Εντούτοις, αυτές οι προσεγγίσεις μέχρι σήμερα ήταν λιγότερο πετυχημένες σε βιομηχανικές εφαρμογές από ό,τι στο ερευνητικό εργαστήριο.

Όπως είδαμε σε αυτό το κεφάλαιο, πολλές ευρέως διαδεδομένες εφαρμογές στη σημερινή κοινωνία της πληροφορίας βασίζονται σε πολύ μεγάλο βαθμό στη γλωσσική τεχνολογία. Εξαιτίας της πολύγλωσσης κοινότητάς της, αυτό ισχύει ιδιαίτερα για την οικονομία και την τεχνολογία της Ευρώπης. Αν και η γλωσσική τεχνολογία έχει σημειώσει σημαντική πρόοδο τα τελευταία χρόνια, υπάρχουν ακόμα τεράστιες δυνατότητες για βελτίωση της ποιότητας των συστημάτων γλωσσικής τεχνολογίας. Ακολούθως, θα περιγράψουμε το ρόλο των Ελληνικών στην ευρωπαϊκή κοινωνία της πληροφορίας και θα κάνουμε μια εκτίμηση της σημερινής κατάστασης της γλωσσικής τεχνολογίας για την ελληνική γλώσσα.

http://www.springer.com/978-3-642-28935-4

The Greek Language in the Digital Age
Rehm, G.; Uszkoreit, H. (Eds.)
2012, VI, 79 p. 24 illus. in color., Softcover
ISBN: 978-3-642-28935-4