

B. Μέσα Μαζικής Επικοινωνίας και Γλωσσική Τεχνολογία: Παρούσα κατάσταση και προοπτικές¹

Γεώργιος Κ. Μικρός
Πανεπιστήμιο Αθηνών

1. Γλωσσική Τεχνολογία

Ο όρος “Γλωσσική Τεχνολογία” (ΓΤ) είναι ο όρος για τις γλωσσικές δυνατότητες που ενσωματώνονται σε υπολογιστικές εφαρμογές οι οποίες χρησιμοποιούνται σε συστήματα πληροφορικής και επικοινωνιακής τεχνολογίας (EUROMAP 2001). Μέσω αυτής βελτιώνεται η χρηστικότητα και η προσβασιμότητα του ψηφιακού περιεχομένου και των υπηρεσιών και υποστηρίζεται ενεργά η πολυγλωσσία σε ευρωπαϊκό επίπεδο. Η ΓΤ μερικές φορές είναι αρκετά οικεία, π.χ. ο ορθογραφικός διορθωτής στον επεξεργαστή κειμένου σας, αλλά συχνά μπορεί να βρίσκεται ενσωματωμένη σε πολύπλοκες δικτυακές εφαρμογές, όπως οι μηχανές αναζήτησης στο Διαδίκτυο. Τόσο τα προϊόντα, όσο και οι υπηρεσίες της ΓΤ, δίνουν τη δυνατότητα στους ανθρώπους να επικοινωνούν πιο φυσικά, πιο αποτελεσματικά και σε μεγαλύτερο εύρος με τους υπολογιστές τους, αλλά πάνω απ’ όλα μεταξύ τους.

Η ΓΤ παρέχει καινοτόμες δυνατότητες αλληλεπίδρασης με το τεχνολογικό περιβάλλον μας. Η γλωσσική γνώση επενδύεται στην ανάπτυξη συστημάτων τα οποία μπορούν να αναγνωρίσουν ομιλία και γραφή, να κατανοήσουν κείμενα αρκετά καλά, ώστε να μπορούν να επιλέγουν πληροφορίες, να μεταφράζουν από μια γλώσσα σε άλλες, και να τα κατηγοριοποιούν βάσει της θεματικής τους.

Η ΓΤ αποτελεί την διεπαφή μεταξύ ποικίλων ανθρωπίνων δραστηριοτήτων και τεχνολογικής χρήσης. Σε αυτό το πλαίσιο μπορούμε να εξετάσουμε την ενσωμάτωσή της στις δραστηριότητες των ΜΜΕ, αφού πολλές από αυτές διευκολύνονται εξαιρετικά από την ύπαρξη σχετικών προϊόντων. Η θέση που θα υιοθετήσουμε στην παρούσα περίπτωση είναι διττή. Θα εξετάσουμε τα ΜΜΕ ως χρήστες της ΓΤ και θα δούμε μερικά χα-

ρακτηριστικά παραδείγματα εφαρμογών σε αυτά. Ωστόσο, σε ένα δεύτερο επίπεδο θα αναλύσουμε το ρόλο των ΜΜΕ ως αντικείμενο παρατήρησης και πεδίο δοκιμής των εργαλείων της ΓΤ και θα εξετάσουμε κατά πόσο η ΓΤ και ειδικότερα τα ΗΣΚ των ΜΜΕ μπορούν να φωτίσουν τη γλωσσική χρήση στις πραγματικές συνθήκες επικοινωνίας.

2. Τα ΜΜΕ ως χρήστες της ΓΤ

Τα ΜΜΕ αποτελούν έναν από τους σημαντικότερους χρήστες ΓΤ. Παρακάτω θα επιχειρήσουμε να σκιαγραφήσουμε τις σημαντικότερες εφαρμογές ΓΤ που σχετίζονται με ορισμένες από τις ανάγκες αυτών που εμπλέκονται με τη λειτουργία των ΜΜΕ. Κάποιες από αυτές είναι ώριμες τεχνολογίες και είναι ενσωματωμένες σε προϊόντα ευρείας κατανάλωσης, όπως τους επεξεργαστές κειμένου, ενώ άλλες είναι σε πειραματικό στάδιο και η λειτουργία τους τεκμηριώνεται σε ερευνητικό ακόμα επίπεδο.

Γραπτή Γλώσσα

- Ορθογραφική και γραμματική διόρθωση: Μια από τις πρώτες αλλά και πιο χρήσιμες εφαρμογές ΓΤ ήταν ο ορθογραφικός διορθωτής. Η δυνατότητα της απλής ορθογραφικής διόρθωσης έχει τον τελευταίο καιρό εμπλουτιστεί σημαντικά με αποτέλεσμα ο απλός ορθογραφικός διορθωτής να περιλαμβάνει δυνατότητα παροχής συνωνύμων και αντωνύμων λέξεων και σχετικών εκφράσεων. Επιπλέον, έχουν παρουσιαστεί για την ελληνική γλώσσα διορθωτές που μπορούν να κάνουν «έξυπνη» ορθογραφική διόρθωση επιτρέποντας τον έλεγχο σε μορφοσυντακτικό επίπεδο. Έτσι, για παράδειγμα στη φράση «έχει μεγάλη τείχη», το «τείχη» δεν θα διορθωθεί ποτέ από έναν κλασικό ορθογράφο, αφού δεν παραβιάζει κανέναν ορθογραφικό κανόνα. Ωστόσο, η νέα γενιά ορθογράφων θα εντοπίσει τη μορφοσυντακτική ασυμφωνία μεταξύ επιθέτου και ουσιαστικού από τη μία και ρήματος και ουσιαστικού από την άλλη και θα προτείνει τη σωστή ορθογραφικά λέξη «τύχη».
- Υφολογική διόρθωση: Ο διορθωτής ύφους ελέγχει

¹ Τμήματα του παρόντος άρθρου έχουν ανακοινωθεί στην «Ημερίδα για τα 10 χρόνια λειτουργίας του Τμήματος Επικοινωνίας και Μέσων Μαζικής Ενημέρωσης με θέμα: Δικτύωση, διάδοση της γνώσης και επικοινωνία στην Ελλάδα σήμερα» που διοργανώθηκε στο Πάντειο Πανεπιστήμιο (8 Νοεμβρίου 2001).

την καταλληλότητα του ύφους ενός κειμένου καθώς αυτό ετοιμάζεται μέσα από έναν επεξεργαστή κειμένου. Οι σύγχρονοι υφολογικοί διορθωτές αποτελούνται από δύο τμήματα: α) το πρώτο ελέγχει κατά πόσο χρησιμοποιούνται ορισμένα γλωσσικά στοιχεία η παρουσία των οποίων θεωρείται αρνητική για την υφολογική διαμόρφωση των κειμένων (π.χ. την κατάχρηση παθητικής φωνής) και β) το δεύτερο υπολογίζει ορισμένα υφομετρικά χαρακτηριστικά του κειμένου. Ενδεικτικά μπορούμε να αναφέρουμε τους δείκτες αναγνωσιμότητας που υπολογίζονται από την μέτρηση των συλλαβών ανά λέξη και τη μέτρηση λέξεων ανά πρόταση. Ορισμένοι υφολογικοί διορθωτές προχωρούν και σε συμβουλές βελτίωσης του ύφους, ενώ μια νέα γενιά υφολογικών διορθωτών μπορεί να επιλύσει αμφισημίες αναφοράς και να προσδιορίσει τους υφολογικούς κανόνες ανάλογα με το είδος του κειμένου και το θεματικό του περιεχόμενο (Johnson 1992).

- Μηχανική μετάφραση – Δίγλωσσα ηλεκτρονικά λεξικά – Αυτόματος εντοπισμός όρων: Η μηχανική μετάφραση ήταν πάντα ένας από τους σημαντικότερους στόχους της ΓΤ. Η αρχική αισιοδοξία που υπήρξε στην δεκαετία του '50 για τη δημιουργία συστημάτων αυτόματης μετάφρασης υψηλής ποιότητας γρήγορα αντικαταστάθηκε από τον σκεπτικισμό και την αποτυχία παρουσίασης ενός συστήματος που να μεταφράζει ζευγάρια γλωσσών με τρόπο που δεν θα χρειάζεται ανθρώπινη επιμέλεια έπειτα. Οι σύγχρονες τάσεις της μηχανικής μετάφρασης πλέον μπορούν να προσδιοριστούν από τις παρακάτω τάσεις (Καραγιάννης 1999: 22):
 - Χρησιμοποίηση της συντακτικής ανάλυσης στα κείμενα της γλώσσας πηγής έτσι ώστε να διευκολυνθεί η ανασύσταση της συντακτικής δομής στη γλώσσα στόχο και να δημιουργηθεί μετάφραση που θα έχει τη συντακτική δομή της γλώσσας στόχου και όχι της γλώσσας πηγής. Για να γίνει αυτό θα πρέπει τα κείμενα της γλώσσας πηγής να υποστούν προεπεξεργασία έτσι ώστε να περιέχουν τόσο μορφολογική όσο και συντακτική πληροφορία πριν δοθούν ως εισαγόμενο στο συντακτικό αναλυτή.
 - Αξιοποίηση τόσο συμβολικών, όσο και στατιστικών τεχνικών. Οι πρώτες συμπληρώνουν τις δεύτερες και οδηγούν σε υβριδικά συστήματα με αυξημένη αποδοτικότητα.

- Χρησιμοποίηση τεχνικών μετάφρασης βασισμένων σε παραδείγματα (example based translation). Υψηλής ποιότητας υπάρχουσες μεταφράσεις αποθηκεύονται σε μεταφραστικές μνήμες και ανακαλούνται από το σύστημα με διαδικασίες ταιριάσματος.
- Χρησιμοποίηση σημασιολογικών χαρακτηριστικών και μοντέλων για την άρση της αμφισημίας η οποία αποτελεί το κύριο εμπόδιο για τη διαδικασία της μηχανικής μετάφρασης. Ενσωμάτωση της αλληλεπιδραστικής δυνατότητας με το σύστημα σε περιπτώσεις που δεν μπορεί να επιλύσει την αμφισημία.

- Ανάκληση πληροφορίας (information retrieval): Η ταχύτατη εξάπλωση του Διαδικτύου και η οργάνωσή των κειμένων του σε δαιδαλώδεις ιεραρχικές αλυσίδες μετέτρεψε την αποτελεσματική αναζήτηση πληροφορίας ως έναν από τους βασικότερους στόχους της πληροφορικής επιστήμης. Και ενώ η ανάπτυξη μηχανών αναζήτησης αποτελεί ένα πρώτο βήμα στην αντιμετώπιση αυτού του θέματος η ακρίβειά τους και η αποτελεσματικότητά τους δεν έχει φτάσει ακόμα σε επιθυμητά επίπεδα. Τόσο ο θόρυβος των αποτελεσμάτων, όσο και η αδυναμία χρήσης ερωτημάτων διατυπωμένων σε φυσική γλώσσα, επηρεάζει την απόδοση των μηχανών αναζήτησης και μειώνει την παραγωγικότητα των χρηστών τους. Η ΓΤ έχει αναπτύξει μια σειρά από τεχνικές που ενσωματώνουν τη γλωσσική γνώση σε τέτοια συστήματα και βελτιώνουν την πιθανότητα ανάκλησης σχετικής πληροφορίας με το ερώτημα.
- Αυτόματη κατηγοριοποίηση και δρομολόγηση εγγράφων: Μια εφαρμογή που επίσης έχει σημαντικές επιπτώσεις στην εργασία χρηστών που διαχειρίζονται γλωσσικές πληροφορίες είναι τα συστήματα αυτόματης κατηγοριοποίησης και δρομολόγησης εγγράφων. Τέτοια συστήματα επιτρέπουν την αυτόματη οργάνωση της εισερχόμενης ηλεκτρονικής πληροφορίας του χρήστη βάσει θεματικού περιεχομένου. Έτσι για παράδειγμα ο καθημερινός κατακλυσμός από μηνύματα η-ταχυδρομείου μπορεί να οργανωθεί αυτόματα σε θεματικούς πυρήνες πριν φτάσει στον χρήστη, διευκολύνοντας έτσι την ανάγνωσή τους. Μέσω δε της αυτόματης δρομολόγησης κειμένων ο χρήστης μπορεί να επιλέξει συγκεκριμένη θεματική που τον ενδιαφέρει και να δέ-

χεται καθημερινά κείμενα από το Διαδίκτυο που σχετίζονται με το θέμα που έχει προεπιλέξει. Και οι δύο τεχνολογίες αποτελούν προϊόντα της έρευνας στη ΓΤ και μπορούν να συνδυαστούν με μια σειρά από επιμέρους συστήματα όπως αυτά της αυτόματης εξαγωγής όρων από κείμενα.

Προφορική Γλώσσα

- Σύνθεση φωνής: Η σύνθεση φωνής αποτελεί μια σημαντική δυνατότητα που έχει προστεθεί στο H/Y τα τελευταία χρόνια. Η ύπαρξη τέτοιων συστημάτων και για την ελληνική γλώσσα επιτρέπει την εκφώνηση της ψηφιακά αποθηκευμένης κειμενικής πληροφορίας με εφαρμογές που κινούνται από την χρήση των H/Y από άτομα με ειδικές ανάγκες έως την εκφώνηση των κειμένων που έχει δημιουργήσει ο χρήστης για τον έλεγχο συντακτικών λαθών.
- Αναγνώριση ομιλίας: Η προσθήκη δυνατότητας αναγνώρισης ομιλίας στον H/Y έχει προφανή χρησιμότητα για όλο το φάσμα των χρηστών του. Μια τέτοια δυνατότητα θα καταργήσει τη χρήση ενδιάμεσων συσκευών επικοινωνίας με τον H/Y, όπως είναι το πληκτρολόγιο ή το ποντίκι και θα επιτρέψει την φυσική διάδραση συστήματος και ανθρώπου. Οι επιδόσεις σχετικών συστημάτων είναι αρκετά ικανοποιητικές όταν έχουν να αντιμετωπίσουν κλειστό σύνολο λέξεων που συνήθως περιορίζεται σε ένα μικρό αριθμό ονομάτων ή εκφώνηση αριθμών. Η πρόκληση για τέτοια συστήματα βρίσκεται στο επίπεδο της αναγνώρισης συνεχούς λόγου όπου η ποικιλία εκφοράς, ο περιβάλλον θόρυβος, αλλά και η ίδια η συντακτική πολυπλοκότητα του προφορικού μηνύματος μειώνουν σημαντικά την ακρίβεια αναγνώρισης τέτοιων συστημάτων.

Οι παραπάνω εφαρμογές της ΓΤ ανοίγουν σημαντικούς ορίζοντες σε αυτούς που εμπλέκονται με την διαχείριση, οργάνωση και παρουσίαση πληροφοριών στα ΜΜΕ και η γνωριμία με αυτές αποτελεί τη βασικότερη προϋπόθεση για την αποτελεσματική αξιοποίησή τους.

3. Τα ΜΜΕ ως αντικείμενο έρευνας με τη βοήθεια της ΓΤ

Τα ΜΜΕ εκτός από δυνητικούς χρήστες των εφαρμογών ΓΤ, αποτελούν πολλές φορές αντικείμενο έρευ-

νας της. Παρακάτω θα προσδιορίσουμε γιατί η γλώσσα των ΜΜΕ ενδιαφέρει τους γλωσσολόγους και θα αναφέρουμε ορισμένα αποτελέσματα από έρευνες που χρησιμοποιώντας εργαλεία ΓΤ εξέτασαν τον λόγο των ΜΜΕ.

3.1 Τα ΗΣΚ και η σημασία τους

Τα τελευταία χρόνια η γλωσσολογία ως επιστήμη, αξιοποιώντας την αυξημένη δυνατότητα των H/Y να χειριστούν γλωσσικά δεδομένα, στράφηκε ξανά στην εξέταση της γλώσσας στο περιβάλλον χρήσης της. Αν και για ένα μεγάλο χρονικό διάστημα η θεωρία της γενετικής γραμματικής του Noam Chomsky επικέντρωσε την προσοχή των γλωσσολόγων στη γνωστική πλευρά της γλωσσικής παραγωγής, τα τελευταία χρόνια το βάρος έπεσε στην κοινωνική διάσταση της γλώσσας. Οι εργασίες των πρωτοπόρων κοινωνιογλωσσολόγων ήδη από την δεκαετία του '60 έδειξαν ότι η ανάλυση γλωσσικών δεδομένων από πλήθος ανθρώπων μπορεί να δώσει πληροφορίες στις οποίες ο θεωρητικός γλωσσολόγος δεν μπορούσε ποτέ να έχει πρόσβαση μέσω της ενδοσκόπησης. Δίχως την υιοθέτηση ακραίων θέσεων έγινε κατανοητό ότι η αναζήτηση της γλωσσικής ικανότητας σχετίζεται, τόσο με την αναζήτηση ικανών θεωρητικών σχημάτων ερμηνείας, όσο και εμπειρικής εξέτασης πλήθους γλωσσικών δεδομένων τα οποία έχουν παραχθεί από φυσικούς ομιλητές για να καλύψουν πραγματικές επικοινωνιακές ανάγκες. Για να γίνει εφικτή η ανάλυση τέτοιων δεδομένων προϋποτίθεται η συλλογή τους και η οργάνωσή τους με τέτοιο τρόπο ώστε να αντιπροσωπεύουν με την μεγαλύτερη δυνατή ακρίβεια τη γλωσσική χρήση σε πραγματικά περιβάλλοντα επικοινωνίας. Αυτή η προϋπόθεση μας οδηγεί στην έννοια των Ηλεκτρονικών Σωμάτων Κειμένων (ΗΣΚ), όπως την όρισε και ο Sinclair (1996) τα οποία αποτελούν μία από τις σημαντικότερες πηγές γλωσσικής ανάλυσης. Ακολουθώντας τους Biber, Conrad & Reppen (1998: 4) τα σημαντικότερα χαρακτηριστικά ενός ΗΣΚ είναι τα ακόλουθα:

- Παρέχει εμπειρικά γλωσσικά δεδομένα τα οποία έχουν συλλεχθεί με στατιστικά αντιπροσωπευτικό τρόπο έτσι ώστε η ανάλυσή τους να παρέχει αποτελέσματα τα οποία μπορούν να καλύπτουν τα χαρακτηριστικά της γλωσσικής κοινότητας από την οποία συλλέγονται.

- Οργανώνει τα κείμενα με πολυεπίπεδο τρόπο και τα κατηγοριοποιεί βάσει πολλών και διαφορετικών κριτηρίων όπως κειμενικό θέμα, κειμενικό γένος, χρονολογία συγγραφής, μέσο δημοσίευσης κ.ά.
- Επιτρέπει τη χρήση Η/Υ για την οργάνωση και την ανάλυση των κειμενικών δειγμάτων χρησιμοποιώντας κατάλληλη κωδικοποίηση των γλωσσικών δεδομένων.

Η ανάπτυξη ΗΣΚ αποτελεί σήμερα έναν από τους σημαντικότερους στόχους της ΓΤ καθώς με αυτά μπορεί:

- να εκπαιδεύσει συστήματα Επεξεργασίας Φυσικής Γλώσσας (ΕΦΓ) (συστήματα μηχανικής μετάφρασης, κειμενικής κατηγοριοποίησης κ.ά.)
- να δοκιμάσει τη λειτουργία συστημάτων ΕΦΓ (έλεγχος ακρίβειας συντακτικών και μορφολογικών χαρακτηριστών (parsers & taggers), συντακτικών αποσαφηνιστών (disambiguators), συστημάτων μηχανικής μετάφρασης κ.ά.)
- να δημιουργήσει πλήθος παράγωγων γλωσσικών πόρων (λεξικά γενικής και ειδικής γλώσσας, μορφολογικά λεξικά, λεξικά ορολογίας, δίγλωσσες και πολύγλωσσες αντιστοιχίσεις λεξικών μονάδων κ.ά.).

Δεδομένης της σημασίας των ΗΣΚ στην ανάπτυξη της ΓΤ και την διεύρυνση του ρόλου που κάθε γλώσσα μπορεί να παίξει στην αναδυόμενη Κοινωνία της Πληροφορίας το Ινστιτούτο Επεξεργασίας του Λόγου έχει αναπτύξει για την Ελληνική Γλώσσα τον Εθνικό Θησαυρό της Ελληνικής Γλώσσας (ΕΘΕΓ) ένα ΗΣΚ 30 εκ. λέξεων το οποίο φιλοδοξεί να αποτελέσει αντικείμενο ανάλυσης για τους απανταχού ερευνητές της Νεοελληνικής γλώσσας (Hatzigeorgiou et al. 2000). Το συγκεκριμένο ΗΣΚ είναι διαθέσιμο στο Διαδίκτυο από τη διεύθυνση: <http://corpus.ilsp.gr> και επιτρέπει την αναζήτηση λέξεων ή φράσεων καθώς και μορφολογικών κατηγοριών (π.χ. αναζήτηση της χρήσης γενικά των επιθέτων ή των ρημάτων στην πολιτική αρθρογραφία). Η σύσταση του συγκεκριμένου ΗΣΚ σε σχέση με το εκδοτικό μέσο είναι η ακόλουθη: Βιβλία 15,75%, Εφημερίδες 69,1%, Περιοδικά 6,97%, Διάφορα 8,27%.

Αν δούμε συγκριτικά τα ποσοστά θα παρατηρήσουμε ότι η γλώσσα των ΜΜΕ, και στην παρούσα περίπτωση οι εφημερίδες, καταλαμβάνουν σχεδόν τα 2/3 του συνολικού ΗΣΚ. Η συγκεκριμένη επιλογή βρίσκεται σε συμφωνία με αντίστοιχες κατηγοριοποιήσεις που ακολουθούν και άλλα ΗΣΚ. Αν εξετάσουμε τα σύνολο των ΗΣΚ γενικής γλώσσας θα παρατηρήσουμε ότι το σύνολό τους περιλαμβάνει δείγματα λόγου από τα ΜΜΕ, ενώ κάποια από αυτά χρησιμοποιούν μόνο αυτόν. Το αυξημένο ενδιαφέρον ενσωμάτωσης της γλώσσας των ΜΜΕ στα ΗΣΚ πηγάζει από τους ακόλουθους λόγους:

- Διαθεσιμότητα
- Ηλεκτρονική διαμόρφωση (τις περισσότερες φορές)
- Αντιπροσωπευτικότητα (αφού ο λόγος των ΜΜΕ αποτελεί μια από τις σημαντικότερες σε όγκο πηγές παραγωγής γλωσσικού υλικού)
 - Ερευνητικό ενδιαφέρον το οποίο εστιάζεται κυρίως στα ακόλουθα σημεία:
 - Η σχέση της γλώσσας των ΜΜΕ και η πολιτισμική πραγματικότητα που αυτή κατασκευάζει.
 - Η επίδραση της γλώσσας των ΜΜΕ στη δομή της γλώσσας εν γένει.
 - Η επιρροή της γλώσσας των ΜΜΕ στη διαμόρφωση στάσεων και αντιλήψεων στην κοινωνία μέσα από συγκεκριμένες επιλογές χρήσης.

3.2 Η γλώσσα των ΜΜΕ όπως αναλύεται μέσα από τη χρήση σχετικών ΗΣΚ

Μέσα από τη χρήση του ΕΘΕΓ και ειδικότερα επιλέγοντας τα κείμενα των ΜΜΕ μπορεί κανείς να ερευνήσει ιδιαίτερες ενδιαφέροντα ερωτήματα που για λόγους τυπολογίας μπορούμε γενικά να τα κατατάξουμε σε δύο κατηγορίες: α) Ερωτήματα που σχετίζονται με τη γλωσσική δομή και β) Ερωτήματα που διερευνούν την σχέση της γλωσσικής χρήσης με εξωγλωσσικές δομές, όπως η κοινωνία, το πολιτισμικό περιβάλλον κ.ά. Στη συνέχεια θα δώσουμε ορισμένα παραδείγματα πώς η εξέταση της γλώσσας στα ΜΜΕ μπορεί να φωτίσει, τόσο γραμματικά, όσο και κοινω-

νιοπολιτισμικά θέματα. Το μεγαλύτερο μέρος των ερευνητικών αποτελεσμάτων που θα παρουσιαστεί στη συνέχεια βασίζεται στην ανάλυση των δεδομένων του ΕΘΕΓ, ωστόσο κάποια από αυτά στηρίζονται και σε άλλα ΗΣΚ που έχουν συλλεχθεί από τον υποφαινόμενο για προσωπική ερευνητική χρήση.

3.2.1 Ερωτήματα σχετικά με την κατάσταση της διγλωσσίας στη σύγχρονη χρήση της γλώσσας στα ΜΜΕ

Ένα από τα πρώτα ερωτήματα που θελήσαμε να διερευνήσουμε ήταν κατά πόσο η πάλη δημοτικής – καθαρεύουσας έχει οδηγηθεί σε μια ισορροπία. Η γλώσσα των ΜΜΕ ως παράγοντας προβολής και καθιέρωσης της γλωσσικής νόρμας είναι η αφετηρία για να εξετάσει κανείς κατά πόσο αυτό που προβάλλεται ως γλωσσικό στάνταρντ ενέχει και σε ποιο βαθμό διλήμματα διγλωσσικής φύσης. Το παραπάνω ερευνητικό ερώτημα τέθηκε υπό τη μορφή εργασίας στους φοιτητές του Γ' έτους του Τμήματος ΜΜΕ του Πάντειου Πανεπιστημίου². Η ποσοτική ανάλυση των ΗΣΚ που συλλέχθηκαν από 3 εφημερίδες ευρείας κυκλοφορίας (ΤΟ ΒΗΜΑ, Ελευθεροτυπία, Καθημερινή) τεκμηρίωσε μια σημαντική υπεροχή των «δημοτικών» τύπων σε σχέση με αυτούς που παραπέμπουν σε αρχαιότερους γλωσσικές μορφές. Περαιτέρω διερεύνηση της χρήσης των λόγιων στοιχείων στη γλώσσα του Τύπου (Μικρός et al. 1995) έδειξε ότι υπάρχει αύξηση της χρήσης τους, όταν το θέμα του άρθρου είναι περισσότερο «σοβαρό». Τα ποσοτικά στοιχεία από την ανάλυση αυτών των ΗΣΚ υποστηρίζουν την άποψη ότι η διγλωσσία έχει πλέον υποχωρήσει στην χρήση της ΝΕ γλώσσας και τα αρχαϊκά στοιχεία αξιοποιούνται από τον ομιλητή με τρόπο υφολογικό για να πλουτίσουν την εκφραστική του παλέτα και να υπογραμμίσουν συγκεκριμένες επικοινωνιακές περιστάσεις.

3.2.2 Εξέταση της συγχρονικής ισχύος γραμματικών κανόνων

Η ύπαρξη της γραμματικής του Μ. Τριανταφυλλίδη υπήρξε μια από τις σημαντικότερες συμβολές στην κωδικοποίηση και επικράτηση της Δημοτικής γλώ-

σσας στο ελληνικό κράτος. Η γραμματική αυτή αποτελεί ακόμα και σήμερα τον ακρογωνιαίο λίθο της γλωσσικής διδασκαλίας στην πρωτοβάθμια και τη δευτεροβάθμια εκπαίδευση. Αν και αποτελεί έργο αναφοράς για την ελληνική γλώσσα θα πρέπει να θυμόμαστε ότι είναι η προσπάθεια αποτύπωσης της γλώσσας που ομιλείτο στις αρχές του αιώνα μας. Με δεδομένο ότι η γλώσσα αλλάζει ταχύτατα και σε πολλά επίπεδα ταυτόχρονα, θα πρέπει να εντοπίσουμε τα σημεία εκείνα που έχει συντελεστεί μια γλωσσική αλλαγή και να επικαιροποιήσουμε την γραμματική περιγραφή. Σε διαφορετική περίπτωση θα βρεθούμε στην άκομψη θέση να διδάσκουμε πράγματα τα οποία δεν ισχύουν και δεν βρίσκουν έρεισμα στο γλωσσικό αισθητήριο των ομιλητών. Με τη βοήθεια των ΗΣΚ μπορέσαμε να εντοπίσουμε έναν πολύ γνωστό κανόνα της γραμματικής του Τριανταφυλλίδη ο οποίος έχει υποστεί αλλαγή και θα πρέπει να διαφοροποιηθεί η διδασκαλία του. Πρόκειται για τον κανόνα του τελικού -ν ο οποίος τόσο στον προφορικό (Μικρός 1994, 1999), όσο και στον γραπτό λόγο (Μικρός & Καραγιάννης, 2001) έχει αλλάξει και λειτουργεί με διαφορετικό τρόπο από αυτόν που αρχικά είχε περιγραφεί στη γραμματική του ΜΤ.

3.2.3 Ποσοτικές μετρήσεις για την ελληνική γλώσσα

Η χρήση των ΗΣΚ επέτρεψε την διεξαγωγή μιας σειράς από μετρήσεις στην χρήση της Νέας Ελληνικής γλώσσας που παλαιότερα δεν ήταν εφικτές. Έτσι τα τελευταία χρόνια είμαστε σε θέση να γνωρίζουμε ποιες είναι οι πιο συχνές λέξεις και λήμματα της ΝΕ, ποιο είναι το μέσο μήκος της λέξης και της πρότασης στα ΝΕ τόσο στον Τύπο, όσο και σε λογοτεχνικά κείμενα, ποιες είναι οι πιο συχνές γραμματικές κατηγορίες και κατά πόσο το ποσοστό κλιτών και άκλιτων λέξεων μεταβάλλεται βάσει του κειμενικού είδους που εξετάζουμε (Hatzigeorgiou, Mikros & Carayannis 2001).

3.2.4 Εξωγλωσσικοί παράγοντες της ποικιλίας στη γλωσσική χρήση των ΜΜΕ

Η ποσοτική ανάλυση των ΗΣΚ και μια σειρά από εξειδικευμένες στατιστικές προσεγγίσεις μας επιτρέπουν

² Στο σημείο αυτό θα ήθελα να ευχαριστήσω την Λέκτορα Γλωσσολογίας του Τμήματος ΜΜΕ του Πάντειου Πανεπιστημίου κ. Μαρία Κακαβούλια για τη αμέριστη στήριξη της στο όλο εγχείρημα. Η συγκεκριμένη εργαστηριακή άσκηση συνδιοργανώθηκε με τον υπογράφο στο πλαίσιο του εξαμηνιαίου μαθήματός της στους φοιτητές του Γ' έτους (ακ. έτος 2000 – 2001).

να αναλύσουμε τη γλωσσική ποικιλία και να εντοπίσουμε τους εξωγλωσσικούς εκείνους παράγοντες που επηρεάζουν τη γλωσσική συμπεριφορά και ευνοούν ή αποθαρρύνουν τη χρήση συγκεκριμένων γλωσσικών τύπων που βρίσκονται σε ποικιλία. Έτσι είναι εφικτή η εξήγηση των γλωσσικών επιλογών και η αναγωγή της χρήσης τους σε συγκεκριμένες θεωρητικές προτάσεις για την οργάνωση της γλωσσικής δομής. Η εξέταση ενός ΗΣΚ από εκφωνήσεις ειδήσεων σε δύο ραδιοφωνικούς σταθμούς της πρωτεύουσας (Mikros 1997) έκανε εφικτή την επιβεβαίωση της θεωρίας «Προσαρμογή στο Ακροατήριο – Audience Design» (Bell 1992) και την ερμηνεία των εμπειρικών ευρημάτων μέσα από τη συγκεκριμένη θεωρητική προσέγγιση. Έτσι φάνηκε πώς το «κύρος» του ραδιοφωνικού σταθμού και η «εικόνα» που ο εκφωνητής διατηρεί για το ακροατήριό του, επιδρούν στην ομιλία του και καθορίζουν τις γλωσσικές του επιλογές.

Η επίδραση εξωγλωσσικών κατηγοριών μπορεί να εντοπιστεί και στις έννοιες κειμενικό είδος και κειμενικό γένος. Το γεγονός ότι τα κείμενα διαφοροποιούνται, τόσο ανάλογα με τη θεματική τους, όσο και ανάλογα με την επικοινωνιακή λειτουργία τους, διαφοροποιεί τη γλωσσική χρήση σε αυτά. Ενδεικτικά είναι τα παραδείγματα από τη μελέτη του κανόνα του τελικού -ν (Μικρός & Καραγιάννης 2001), όπου φαίνεται ακριβώς αυτή η εξάρτηση του κειμενικού γένους και του κειμενικού τύπου στα ποσοστά εφαρμογής του κανόνα.

4. Συμπεράσματα

Από την παραπάνω επιλεκτική παρουσίαση τεχνολογιών και ερευνών πάνω στη γλωσσική παραγωγή των ΜΜΕ γίνεται φανερό ότι η Κοινωνία της Πληροφορίας εντείνει τους δεσμούς της ΓΤ και των ΜΜΕ. Η σχέση που αναπτύσσεται είναι αυτή ενός ανατροφοδοτούμενου συστήματος το οποίο εφόσον αρχίσει να λειτουργεί θα οδηγήσει σε ταχεία ανάπτυξη και τους δύο τομείς. Η σύγκλιση των δύο αυτών κόσμων και η κατανόηση της χρησιμότητας της συνεργασίας τους σε επίπεδο θεωρητικής και εφαρμοσμένης έρευνας θα βελτιώσει τη λειτουργία των ΜΜΕ, αυτοματοποιώντας ένα σημαντικό κομμάτι χρονοβόρων διαδικασιών στις οποίες εμπλέκονται οι εργαζόμενοι σε αυτά. Η ΓΤ από την άλλη μεριά μέσα από τη συνεργασία με τα

ΜΜΕ θα διαμορφώσει μια ερευνητική ατζέντα προανατολισμένη στις ανάγκες τους και τις ιδιαίτερες απαιτήσεις που δημιουργούνται από τη διαχείριση γλωσσικού υλικού πολυμεσικά παρεχόμενου.

Βιβλιογραφικές αναφορές

- Bell, A. 1992. *The language of news media*. Oxford: Blackwell.
- Biber, D., Conrad, S. & R. Reppen. 1998. *Corpus linguistics. Investigating language structure and use*. Cambridge: Cambridge University Press.
- EUROMAP (Ευκαιρίες προώθησης της Τεχνολογίας Ανθρώπινου Λόγου στην Ευρώπη). Τι είναι γλωσσική τεχνολογία; (06 – 04 –2001). ΙΕΛ. (διαθέσιμο από <http://www.ilsp.gr/hopeframe.html>).
- Hatzigeorgiu N., Gavrilidou M., Piperidis S., Carayannis G., Papakostopoulou A., Spiliotopoulou A., Vacalopoulou A., Labropoulou P., Mantzari E., Papageorgiou H., Demiros I. 2000. Design and implementation of the online ILSP Greek Corpus. *Proceedings of the LREC 2000 Conference, Athens 2000, 1737-1742*.
- Hatzigeorgiu, N., Mikros, G. & Carayannis, G. 2001. "Word length, word frequencies and Zipf's law in the Greek language". *Journal of Quantitative Linguistics*, Vol. 8.
- Johnson, E. 1992. "The ideal grammar and style checker". *TEXT Technology*, 2.4: 3-4.
- Καραγιάννης, Γ. 2000. «Γενική επισκόπηση των συστημάτων μηχανικής μετάφρασης EUROTRA και SYSTRAN». Στο Σταύρου, Μ. & Μ. Τζεβελέκου (επιμ.), 19 –39.
- Μικρός, Γ. 1994. "Ο κανόνας του τελικού -ν: Κοινωνιογλωσσολογική και υφολογική ποικιλότητα", *Μελέτες για την Ελληνική Γλώσσα: Τιμητική προσφορά στον Καθηγητή Μ. Σετάτο, Πρακτικά της 15ης Ετήσιας Συνάντησης του Τομέα Γλωσσολογίας της Φιλοσοφικής Σχολής του Αριστοτελείου Πανεπιστημίου Θεσσαλονίκης*, 1995, σσ. 155-166.
- Mikros, G. 1997. "Radio news and phonetic variation in Modern Greek". *Greek Linguistics '95. Proceedings of the 2nd International Conference on Greek Linguistics*, 1997, Τόμος Ι, σσ. 35-44.
- Μικρός, Γ. 1999. *Κοινωνιογλωσσολογική προσέγγιση φωνολογικών προβλημάτων της Νέας Ελληνικής: Φωνητική ποικιλία έρρινων συμφώνων*. Αδημοσίευτη Διδακτορική Διατριβή, Εθνικό και Καποδιστριακό Πανεπιστήμιο Αθηνών, Αθήνα.

- Μικρός, Γ., Γαβριηλίδου Μ., Λαμπροπούλου Π. & Δούκας, Γ. 1995. "Χθες ή χτες; Μια ποσοτική μελέτη φωνητικών και μορφολογικών στοιχείων σε κείμενα της Νέας Ελληνικής". *Μελέτες για την Ελληνική Γλώσσα, Πρακτικά της 16ης Ετήσιας Συνάντησης του Τομέα Γλωσσολογίας της Φιλοσοφικής Σχολής του Αριστοτελείου Πανεπιστημίου Θεσσαλονίκης*, 1996, σσ. 645-656.
- Μικρός, Γ. & Γ. Καραγιάννης, 2001. "Ποσοτική ανάλυση της χρήσης του κανόνα του τελικού -ν σε κείμενα της Νέας Ελληνικής". Χειρόγραφο, Αθήνα.
- Sinclair, J. McH. 1996. "Preliminary recommendations on corpus typology." EAGLES Document TCWG-CTYP/P (διαθέσιμο από <http://www.ilc.pi.cnr.it/EAGLES/corpuSTYP/corpuSTYP.html>).
- Σταύρου, Μ. & Τζεβελέκου, Μ. (επιμ.), 2000. *Η μηχανική μετάφραση και η ελληνική γλώσσα*. Αθήνα: Καστανιώτης.

Γ. Σχεδίαση και υλοποίηση ενός συστήματος αυτόματου εντοπισμού μαθητών με πιθανά μαθησιακά προβλήματα

Αθανάσιος Πρωτόπαπας, Δημήτριος Νικολόπουλος, Χρήστος Σκαλούμπακας, Μιχαήλ Καραμάνης, Άννα Κριμπά

*Ινστιτούτο Επεξεργασίας του Λόγου
Αρτέμιδος 6 & Επιδαύρου
151 25 Παράδεισος Αμαρουσίου*

*Παιδαγωγικό Ινστιτούτο
Αιγαίου Πελάγους 1-3
153 41 Αγ. Παρασκευή*

Abstract: In the framework of the project "added value services in the Greek web for schools" of the Hellenic Paedagogical Institute, the Institute for Language & Speech Processing has developed a software application to automatically screen students for learning disabilities, focussing on the typical dyslexic profile. Eight component tasks assess reading and spelling ability, nonverbal and phonetic auditory perception, and phonological and verbal memory. Following pilot application, validation, and standardisation, the software can become available for educators throughout the Greek school system, requiring only minimal computing power and personnel training. The system aims to provide reliable recommendations for referrals to learning disabilities specialists (not diagnosis or comprehensive assessment). Validation will be conducted on a national representative sample of students, and statistical analysis of the results will be applied to select tasks and items to maximise screening reliability and validity.

Στην παρούσα ανακοίνωση παρουσιάζεται η δοκιμαστική έκδοση του λογισμικού με τίτλο «εΜαΔυς» που παρήχθη στο πλαίσιο του ΕΠΕΑΕΚ από το Ινστιτούτο Επεξεργασίας του Λόγου ως μέρος του έργου «υπηρεσίες προστιθέμενης αξίας στο ελληνικό δίκτυο σχολείων» του Παιδαγωγικού Ινστιτούτου (Paradopoulos, Karamanis, & Roussos, 2000).

1. Στόχοι

Το λογισμικό εΜαΔυς έχει στόχο να συμβάλλει στον εντοπισμό των μαθητών με πιθανά μαθησιακά προβλήματα. Δηλαδή, μετά από κατάλληλη στάθμιση, η