

Γλώσσα και Γλωσσική Τεχνολογία στην Ελλάδα

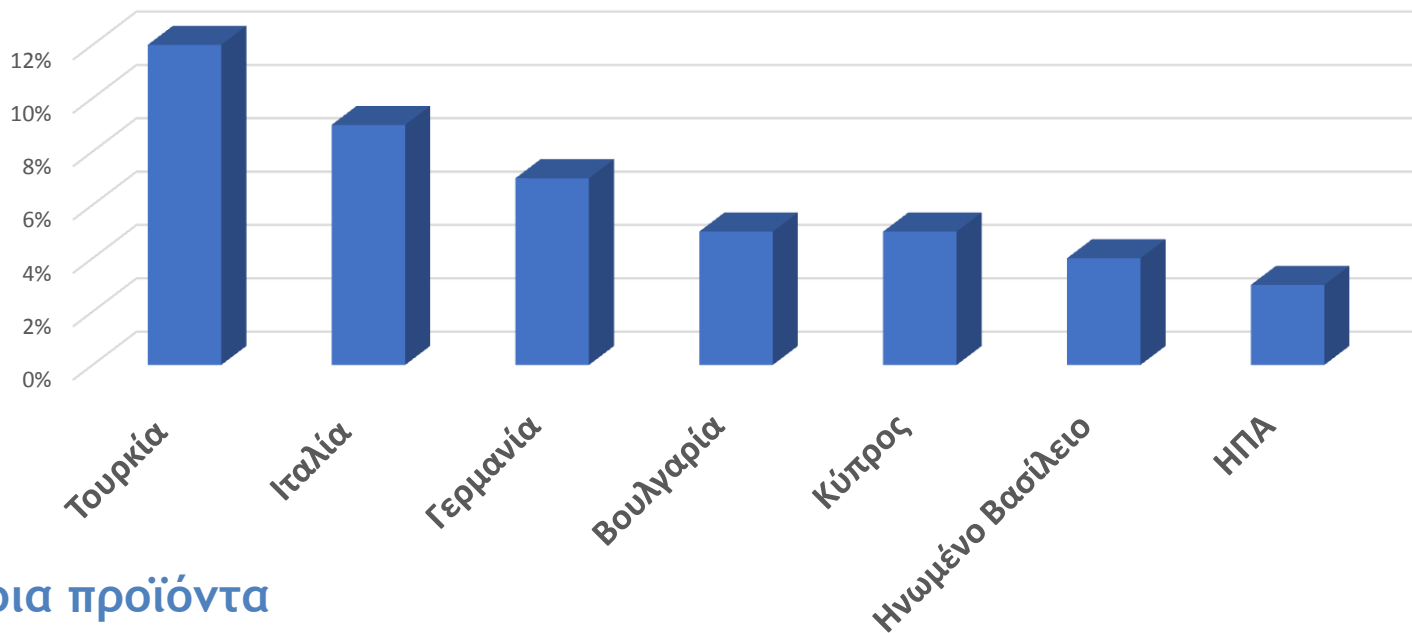
Μαρία Γαβριηλίδου
ΙΕΛ/ΕΚ «Αθηνά»
Κοινοπραξία «Ευρωπαϊκός Συντονισμός Γλωσσικών Πόρων»
Εθνικό Σημείο Επαφής για την Ελλάδα



- Η Γλωσσική Τεχνολογία είναι κρίσιμος παράγοντας για την επίτευξη ισότητας μεταξύ της Ελληνικής και των ευρύτερα χρησιμοποιούμενων, ομιλούμενων και διδασκόμενων γλωσσών
- Υπάρχει μία επίσημη γλώσσα στην Ελλάδα, αλλά...
- Μεγάλος αριθμός Ελλήνων πολιτών, στον ιδιωτικό και τον δημόσιο τομέα χρειάζεται υποστήριξη στις απαιτήσεις της πολυγλωσσικής επικοινωνίας στην καθημερινότητά του
- Η καθημερινότητα αυτή στη δημόσια διοίκηση, όπως και στις επιχειρήσεις είναι πολυγλωσσική, ειδικά σε διακρατικά θέματα, π.χ.
 - Κοινοτικά όργανα
 - Περιφερειακή ανάπτυξη και διακυβέρνηση
 - Οικονομία, απασχόληση και ζωή των πολιτών
 - Κινητικότητα και μεταφορές
 - Ενέργεια και κλιματική αλλαγή
 - Περιβάλλον και φυσικοί πόροι
 - Μετανάστευση και ασφάλεια



- Συνεργασία με γειτονικά κράτη
 - Εντός και εκτός ΕΕ (Αλβανία, Βοσνία, Βουλγαρία, Γεωργία, Ισραήλ, Ιταλία, Μαυροβούνιο, Ρουμανία, Σερβία, Τουρκία...)
- Ακόμη και στην καθημερινή ζωή, η πολυγλωσσία είναι ζωτικής σημασίας στην Ελλάδα, όπως και σε όλες της Ευρωπαϊκές χώρες:
 - Διεθνής συνεργασία δημόσιων υπηρεσιών σε σχέση με 'διεθνείς πολίτες' (ασφάλεια, ιθαγένεια, γάμος, επιχειρήσεις, νομικές υποθέσεις κ.λπ.)
 - Βιομηχανία και υπηρεσίες τουρισμού
 - Υπηρεσίες Εκπαίδευσης
 - Οικοσύστημα πολιτισμού



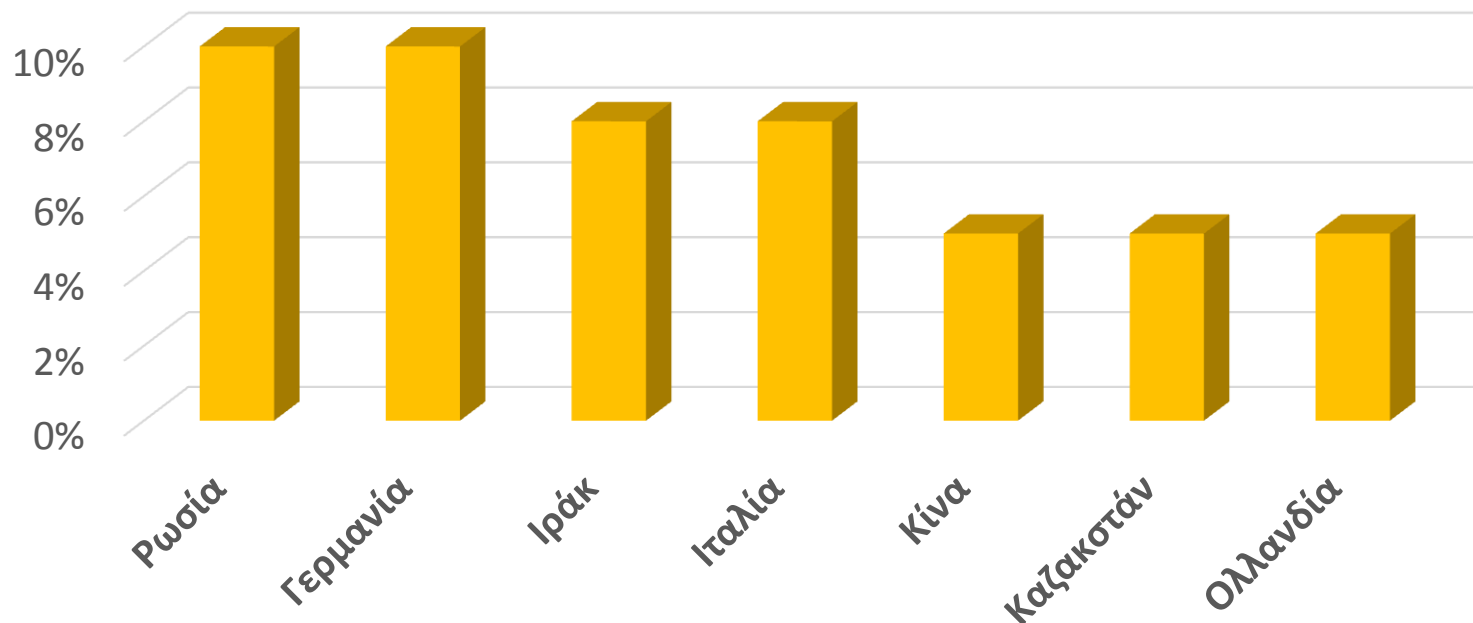
Κύρια προϊόντα

- Τρόφιμα και ποτά
- Βιομηχανικά προϊόντα
- Προϊόντα πετρελαίου
- Χημικά προϊόντα

Συνολικά προς ΕΕ: 48%

Πηγή: <http://www.investingreece.gov.gr/>

Εισαγωγές (2014)



Κύρια προϊόντα

- Μηχανήματα
- Μεταφορικά μέσα
- Καύσιμα
- Χημικά προϊόντα
- Τρόφιμα

Πηγή: <http://www.investingreece.gov.gr/>

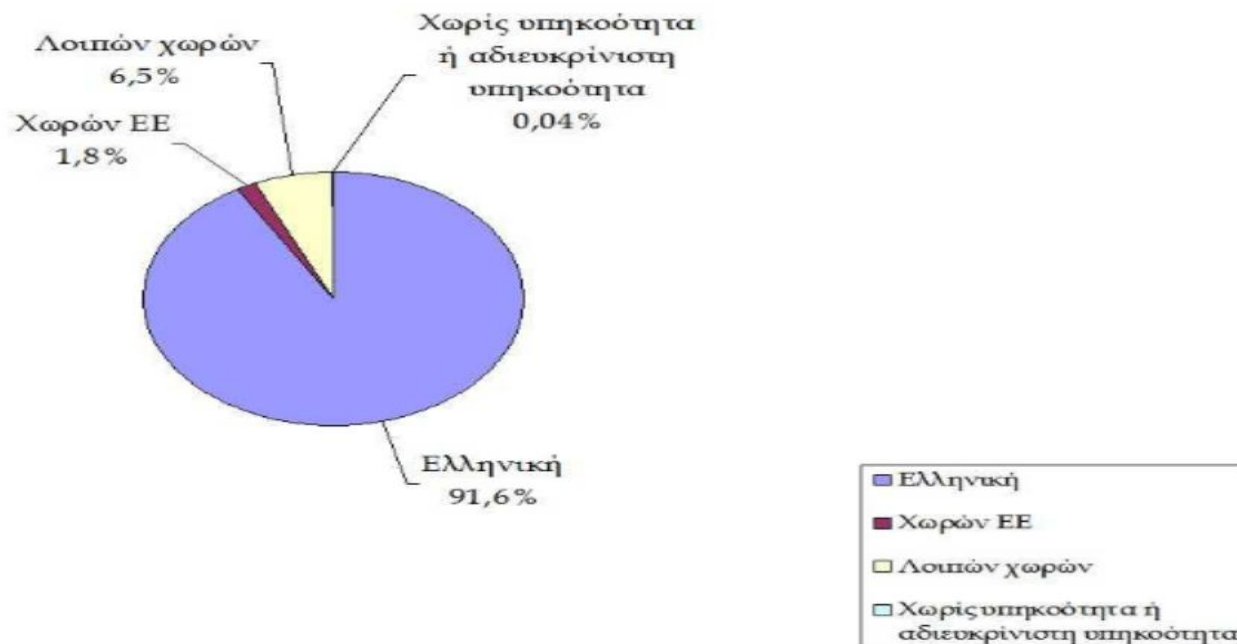
Citizenship	Census 2011	%
Greece	9,903,268	91.56
Albania	480,824	4.45
Bulgaria	75,915	0.70
Romania	46,523	0.43
Pakistan	34,177	0.31
Georgia	27,400	0.25
Ukraine	17,006	0.16
United Kingdom	15,386	0.14
Cyprus	14,446	0.13
Poland	14,145	0.13
Russia	13,807	0.12
India	11,333	0.10
Bangladesh	11,076	0.10
Germany	10,778	0.09
Egypt	10,455	0.09
Moldova	10,391	0.09
Philippines	9,804	0.09
Other	108,436	1
Total	10,815,197	100

911,929 αλλοδαποί στην Ελλάδα (απογραφή 2011)

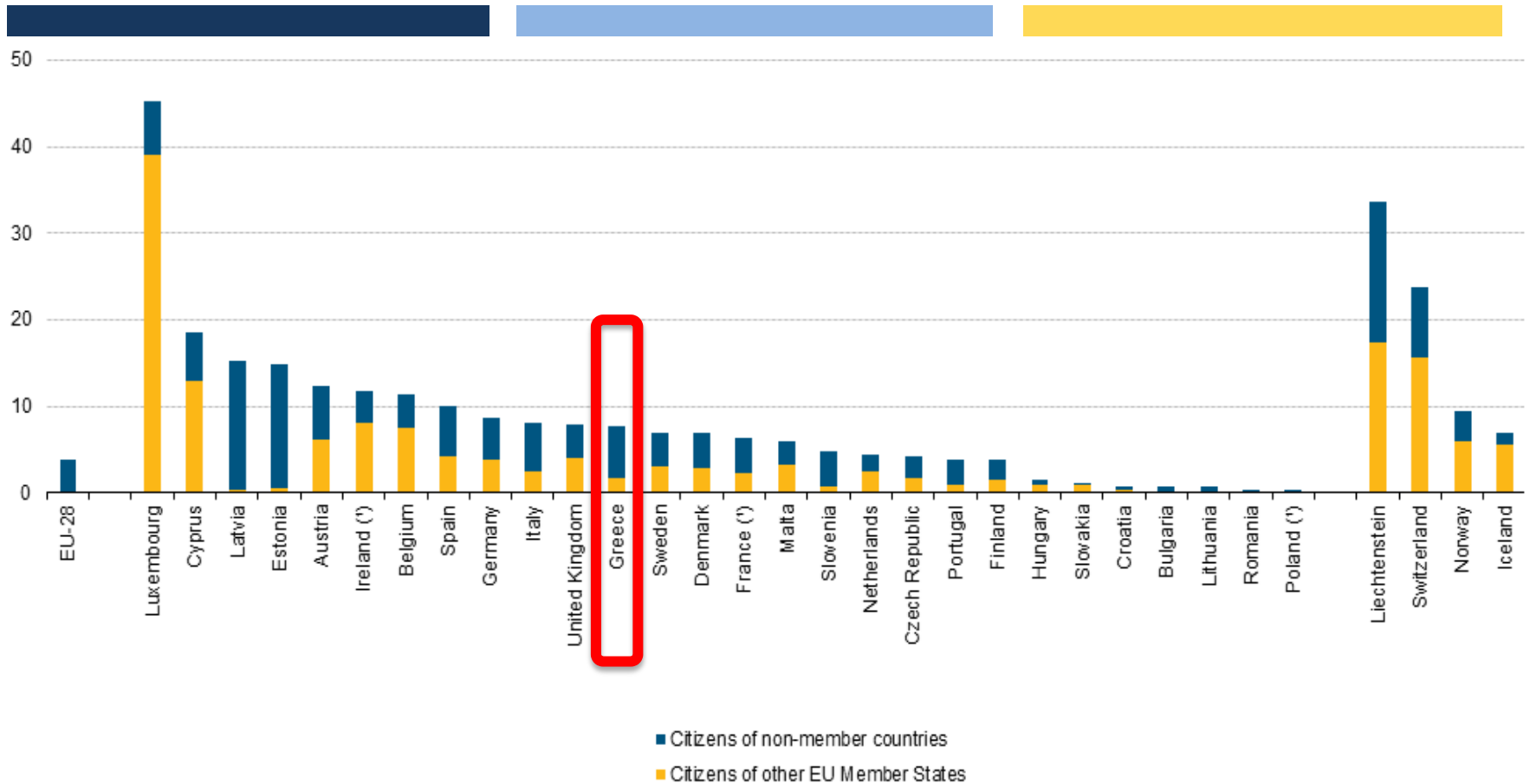
Πηγή: ΕΛΣΤΑΤ

http://www.statistics.gr/portal/page/portal/ESYE/BUCKET/General/nws_SAM01_GR.PDF

Μόνιμος πληθυσμός κατά υπηκοότητα



Πολίτες ΕΕ/εκτός ΕΕ στην Ελλάδα



(*) Provisional.

Source: Eurostat (online data code: migr_pop1ctz)

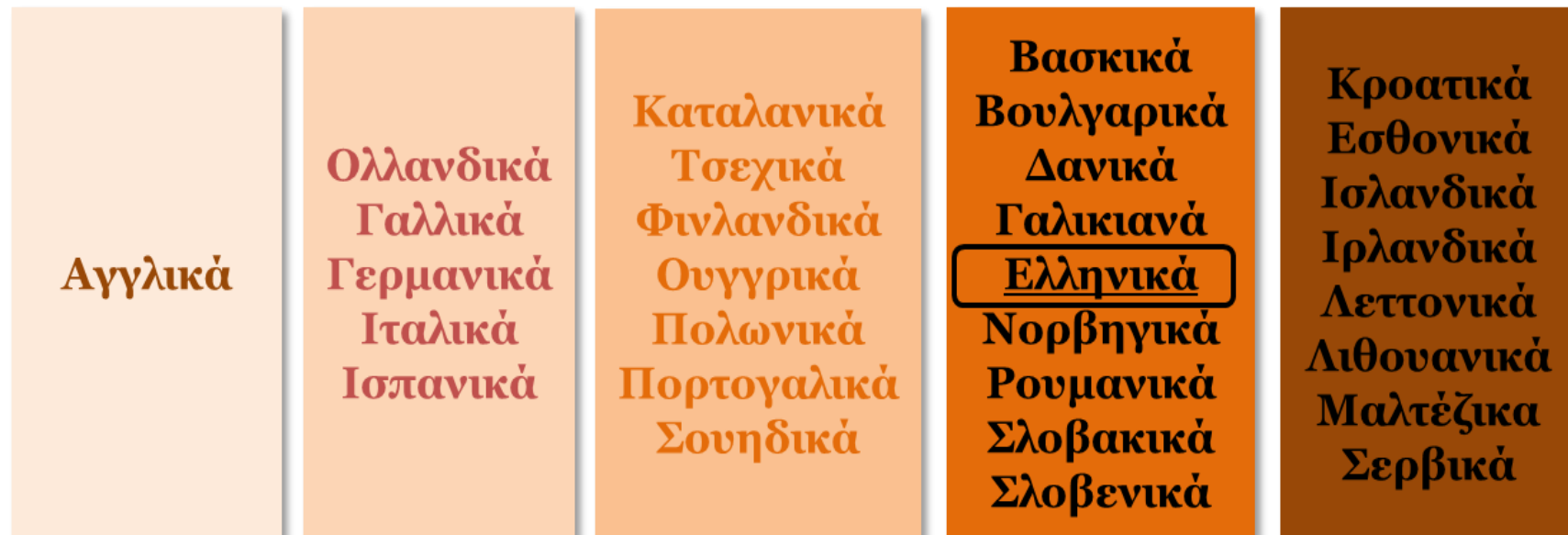
- Υπάρχουν οι βασικές τεχνολογίες τουλάχιστον σε δύο άξονες ανάπτυξης:
 - Τεχνολογίες επεξεργασίας, διαχείρισης και παραγωγής γλώσσας, πληροφορίας και γνώσης
 - Πολυγλωσσική και διαγλωσσική επεξεργασία πληροφορίας, κυρίως αυτόματη μετάφραση και μετάφραση υποβοηθούμενη από υπολογιστή
- Ινστιτούτο Επεξεργασίας του Λόγου (ΙΕΛ)/ΕΚ Αθηνά
- Εργαστήριο Τεχνολογίας Γνώσεων και Λογισμικού/ΕΚΕΦΕ Δημόκριτος
- Εργαστήρια στα Πανεπιστήμια Αθήνας, Πειραιά, Πάτρας, Θεσσαλονίκης, Αιγαίου, Ιονίου, Θράκης
- Κέντρο Ελληνικής Γλώσσας σε θέματα γλωσσικής εκπαίδευσης
- Ιδιωτικές εταιρίες



- Πλήθος πόρων, εργαλείων και τεχνολογιών που υποστηρίζουν τη χρήση της Ελληνικής σε Πληροφοριακά Συστήματα και διευκολύνουν την προσαρμοσμένη στις ιδιαιτερότητες των Ελληνικών επεξεργασία περιεχομένου
 - Ακολουθίες διαδικτυακών υπηρεσιών γλωσσικής επεξεργασίας
 - Επισημειωμένοι πόροι, λεξικά/οντολογίες, με μορφολογική, συντακτική και σημασιολογική πληροφορία
 - Εφαρμογές, ερευνητικά πρωτότυπα και εμπορικές λύσεις, π.χ. Εφαρμογή Αναζήτησης και Ανάκτησης στα ΦΕΚ, γλωσσικά ενήμερες υπηρεσίες διαχείρισης & παρακολούθησης σε αρκετές Πύλες Κυβερνητικών Οργανισμών

- META-SHARE (<http://www.meta-share.org/>)
 - Το πρώτο βήμα προς ένα πανευρωπαϊκό δίκτυο καταμεμημένων αποθετηρίων για γλωσσικά δεδομένα και τεχνολογίες
- αποθετήριο QT21 συμβατό με το META-SHARE (<http://qt21.metashare.ilsp.gr/>)
 - Επέκταση του META-SHARE σε διαδικτυακές υπηρεσίες (υπηρεσίες νέφους)
- CLARIN-EL (www.clarin.gr)
 - Υποδομή Γλωσσικών πόρων στην υπηρεσία της ερευνητικής κοινότητας (*ΚΑΕ)
 - Εθνικός σκέλος της ευρωπαϊκής υποδομής CLARIN (www.clarin.eu)
- Εθνικός Θησαυρός Ελληνικής Γλώσσας (<http://hnc.ilsp.gr/>)
- Σώμα Ελληνικών Κειμένων (<http://www.sek.edu.gr/>)
- Δίκτυο ορολογίας για την ελληνική γλώσσα και μετάφραση (http://ec.europa.eu/greece/news/2015/20150323_diktyo_oroologias_el.htm)
- ΕΛΕΤΟ (<http://www.eleto.gr/>)
- Ιστότοποι για τους μεταφραστές (π.χ. <http://www.translatum.gr/>)

- Ουσιαστική πρόοδος σε πολυγλωσσικές εφαρμογές, ως προς
 - Ποιότητα μετάφρασης
 - Ταχύτητα μετάφρασης
 - Προσαρμογή σε νέους θεματικούς τομείς
 - Μεταφερισιμότητα σε άλλη γλώσσα, ευχέρεια ανάπτυξης συστημάτων για νέα μεταφραστικά ζεύγη
- Προηγμένη πολυγλωσσική τεχνολογία
 - Για τον εντοπισμό και την αξιοποίηση μεγάλων πόρων παράλληλων κειμένων για πολλές γλώσσες, τομείς και είδη κειμένων
 - Με εξειδίκευση στα Ελληνικά ως γλώσσα-πηγή ή γλώσσα-στόχο



*Καλή υποστήριξη
γλωσσικής
τεχνολογίας*

*Μικρή έως καθόλου
υποστήριξη*

Πηγή: Η Ελληνική Γλώσσα στην Ψηφιακή Εποχή, META-NET White Paper Series: Greek,
<http://www.meta-net.eu/whitepapers/volumes/greek>

	Άριστη	Καλή	Μέτρια	Αποσπασματική	Μικρή ή καθόλου υποστήριξη
MM		Αγγλικά	Γαλλικά, Ισπανικά	Καταλανικά, Ολλανδικά, Γερμανικά, Ουγγρικά, Ιταλικά, Πολωνικά, Ρουμανικά	Βασκικά, Βουλγαρικά, Κροατικά, Τσεχικά, Δανικά, Εσθονικά, Φινλανδικά, Γαλικιανά, Ελληνικά , Ισλανδικά, Ιρλανδικά, Λεττονικά, Λιθουανικά, Μαλτέζικα, Νορβηγικά, Πορτογαλικά, Σερβικά, Σλοβακικά, Σλοβενικά, Σουηδικά
Ανάλυση Κειμένου	Άριστη	Καλή	Μέτρια	Αποσπασματική	Μικρή ή καθόλου υποστήριξη
		Αγγλικά	Ολλανδικά, Γαλλικά, Γερμανικά, Ιταλικά, Ισπανικά	Βασκικά, Βουλγαρικά, Καταλανικά, Τσεχικά, Δανικά, Φινλανδικά, Γαλικιανά, Ελληνικά , Ουγγρικά, Νορβηγικά, Πολωνικά, Πορτογαλικά, Ρουμανικά, Σλοβακικά, Σλοβενικά, Σουηδικά	Κροατικά, Εσθονικά, Ισλανδικά, Ιρλανδικά, Λεττονικά, Λιθουανικά, Μαλτέζικα, Σερβικά
Φωνή	Άριστη	Καλή	Μέτρια	Αποσπασματική	Μικρή ή καθόλου υποστήριξη
		Αγγλικά	Τσεχικά, Ολλανδικά, Φινλανδικά, Γαλλικά, Γερμανικά, Ιταλικά, Πορτογαλικά, Ισπανικά	Βασκικά, Βουλγαρικά, Καταλανικά, Δανικά, Εσθονικά, Γαλικιανά, Ελληνικά , Ουγγρικά, Ιρλανδικά, Νορβηγικά, Πολωνικά, Σερβικά, Σλοβακικά, Σλοβενικά, Σουηδικά	Κροατικά, Ισλανδικά, Λεττονικά, Λιθουανικά, Μαλτέζικα, Ρουμανικά
Γλ.Πόροι	Άριστη	Καλή	Μέτρια	Αποσπασματική	Μικρή ή καθόλου υποστήριξη
		Αγγλικά	Τσεχικά, Ολλανδικά, Γαλλικά, Γερμανικά, Ουγγρικά, Ιταλικά, Πολωνικά, Ισπανικά, Σουηδικά	Βασκικά, Βουλγαρικά, Καταλανικά, Κροατικά, Δανικά, Εσθονικά, Φινλανδικά, Γαλικιανά, Ελληνικά , Νορβηγικά, Πορτογαλικά, Ρουμανικά, Σερβικά, Σλοβακικά, Σλοβενικά	Ισλανδικά, Ιρλανδικά, Λεττονικά, Λιθουανικά, Μαλτέζικα



- Ορισμένα γλωσσικά χαρακτηριστικά της Ελληνικής αποτελούν πρόκληση για την υπολογιστική επεξεργασία:
 - Αλφάβητο, πλούσια μορφολογία, ελεύθερη σειρά όρων πρότασης, παράλειψη προσωπικών αντωνυμιών...
- Αντίστοιχες προκλήσεις και στην πολυγλωσσική επεξεργασία και ειδικά την μετάφραση
 - Αμφισημίες (σε λεξιλογικό και συντακτικό επίπεδο)
 - Προσωπικές αντωνυμίες
 - Το «δύσκολο» περιεχόμενο των κοινωνικών μέσων, όπως για όλες τις γλώσσες
 - Πολύπλοκη η ανάλυση και η σύνθεση της μορφολογίας και της σύνταξης



- Σχετικά περιορισμένη υποστήριξη από τη Γλωσσική Τεχνολογία σε σχέση με άλλες γλώσσες
 - Πολυπλοκότητα γλώσσας
 - Μικρή αγορά σε σύγκριση με αριθμό πόρων και τεχνολογιών που απαιτούνται συνήθως
 - Τα περισσότερα υπάρχοντα συστήματα είναι Αγγλο-κεντρικά και υποστηρίζουν λίγα ζεύγη Ελληνικών – Χ (άλλη γλώσσα εκτός Αγγλικής)
 - Ανάγκες σε «εξωτικά» ζεύγη γλωσσών
- Δυνατότητες βελτίωσης
 - Συντονισμένη, συνεργατική προσπάθεια συλλογής και κατάλληλης οργάνωσης δεδομένων απαραίτητων για σύγχρονη ΜΜ και ΓΤ
 - Προσαρμογή των ΓΠ και ΓΤ σε άλλους τομείς ή σε ανάγκες άλλων χρηστών
 - Ενσωμάτωση σε υπάρχουσες ροές εργασιών μαζί με ορολογικές βάσεις και μεταφραστικές μνήμες
 - Καινοτομικές εφαρμογές εξαγωγής γνώσης και πολύγλωσσης επεξεργασίας σε διάφορα περιβάλλοντα ιδιωτικού και δημόσιου τομέα σε εθνικό και ευρωπαϊκό επίπεδο