



**European Language
Resource Coordination**
Connecting Europe Faculty



Ημερίδα εργασίας ELRC ■ 1 Δεκεμβρίου 2015 ■ Σπίτι της Ευρώπης, Λευκωσία

ΓΛΩΣΣΑ & ΓΛΩΣΣΙΚΗ ΤΕΧΝΟΛΟΓΙΑ ΣΤΗΝ ΚΥΠΡΟ



Πανεπιστήμιο Κύπρου
University of Cyprus

ΓΕΩΡΓΙΟΣ ΦΛΩΡΟΣ

Πανεπιστήμιο Κύπρου

Κοινοπραξία «Ευρωπαϊκός Συντονισμός Γλωσσικών Πόρων» (ELRC) –
Εθνικό Σημείο Επαφής για την Κύπρο

gfloros@ucy.ac.cy

Εισαγωγικά

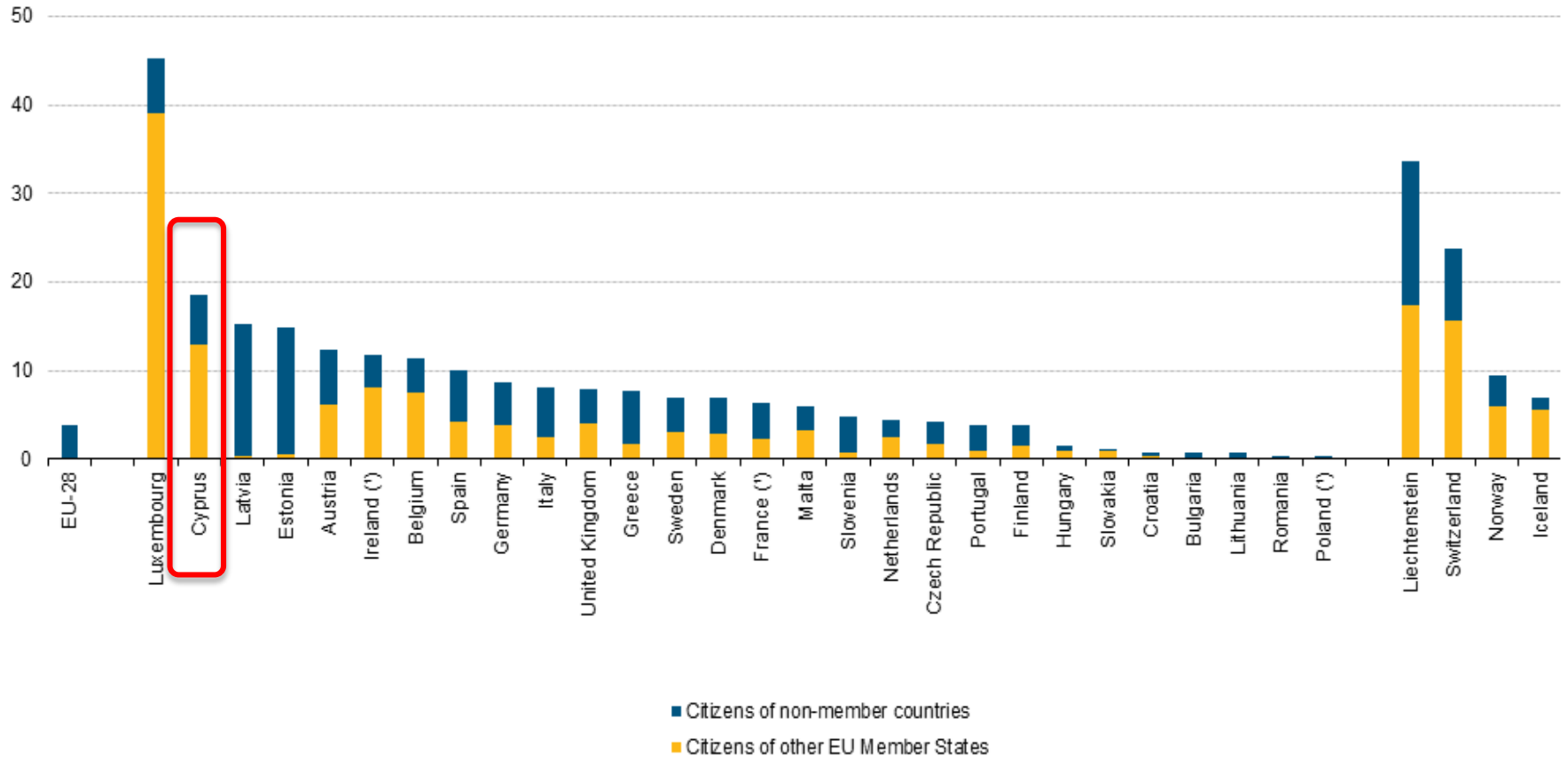
- Η Γλωσσική Τεχνολογία ως κρίσιμος παράγοντας για την επίτευξη ισότητας μεταξύ της Ελληνικής και των ευρύτερα χρησιμοποιούμενων, ομιλούμενων και διδασκόμενων γλωσσών
- Δύο επίσημες γλώσσες στην Κύπρο (Ελληνικά/Τουρκικά), **ακόμα μία ευρέως χρησιμοποιούμενη (Αγγλικά), διαλεκτικές διαφοροποιήσεις**
- Η δημόσια διοίκηση, όπως και οι επιχειρήσεις, έχουν εν πολλοίς πολυγλωσσικές ανάγκες, 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



- Ακόμη και στην καθημερινή ζωή, η πολυγλωσσία είναι ζωτικής σημασίας στην Κύπρο, όπως και σε όλες της Ευρωπαϊκές χώρες:
 - Διεθνής συνεργασία δημόσιων υπηρεσιών σε σχέση με «διεθνείς πολίτες» (ασφάλεια, ιθαγένεια, γάμος, επιχειρήσεις, νομικές υποθέσεις κ.λπ.)
 - Εμπόριο και υπηρεσίες τουρισμού
 - Υπηρεσίες Εκπαίδευσης
 - Οικοσύστημα πολιτισμού
- Το πλεονέκτημα της Κύπρου έναντι της Ελλάδας είναι ότι μεγάλο μέρος του πληθυσμού (συμπεριλαμβανομένων των κρατικών λειτουργών) είναι αρκετά καλοί χρήστες της Αγγλικής, γλώσσα στην οποία διεξάγεται σχεδόν «αυτομάτως» το σύνολο της ξενόγλωσσης συνεννόησης
- Ωστόσο, η Κύπρος δεν διαφέρει από τις υπόλοιπες χώρες της ΕΕ ως προς την ανάγκη υποστήριξης των απαιτήσεων της πολυγλωσσικής επικοινωνίας, ιδίως όταν πρόκειται για έγγραφα.



Ημερίδα εργασίας ELRC ■ 1 Δεκεμβρίου 2015 ■ Σπίτι της Ευρώπης, Λευκωσία



(*) Provisional.

Source: Eurostat (online data code: migr_pop1ctz)



- Τα σημαντικότερα από τα στοιχεία που προαναφέρθηκαν είναι δύο:
 1. Η χρήση των Αγγλικών (έκταση, επίπεδο)
 2. Οι διαλεκτικές διαφοροποιήσεις της Κυπριακής
- Ως προς το 1., η πανεπιστημιακή έρευνα έχει δείξει ότι η Αγγλική χρησιμοποιείται με σχετική «άνεση» (αλλά όχι ως «δεύτερη γλώσσα»).
- Ως προς το 2., η πανεπιστημιακή έρευνα έχει δείξει ότι η χρήση της Κοινής Νέας Ελληνικής παρουσιάζει διαφοροποιήσεις στην Κύπρο. Οι διαφοροποιήσεις είναι άλλοτε αναμενόμενες, άλλοτε μη αναμενόμενες (ορολογία, αλλά και ύφος των κειμένων), πάντως σε κάθε περίπτωση φαίνεται να είναι σε μεγάλο βαθμό συνειδητές.
- Αποτέλεσμα:
 - Αποκλειστική χρήση της Αγγλικής για συνεννόηση και για μεταφραστικά εργαλεία, παρά τις ουσιαστικά πολυγλωσσικές ανάγκες.
 - Τα μεταφραστικά εργαλεία και βοηθήματα ΔΕΝ περιέχουν τις γλωσσικές διαφοροποιήσεις (προς την ΚΝΕ) που απαντώνται στην Κύπρο, με εξαίρεση τις πρόσφατες προσπάθειες της IATE.

Τι είναι η Γλωσσική Τεχνολογία (ΓΤ) και οι Γλωσσικοί Πόροι (ΓΠ);

A. Γλωσσική Τεχνολογία:

- Υπολογιστικά εργαλεία για την πολυεπίπεδη ανάλυση, επεξεργασία, επισημείωση, επαύξηση και μετασχηματισμό γλωσσικών δεδομένων
- Ανάπτυξη συστημάτων αναγνώρισης ομιλίας και γραφής, κατανόησης κειμένου, αυτόματης μετάφρασης και επικοινωνίας. Τα συστήματα είναι άλλοτε αρκετά οικεία, π.χ. ο ορθογραφικός διορθωτής στον επεξεργαστή κειμένου, άλλοτε μέσα σε πολύπλοκα δίκτυα, π.χ. μια μηχανή που διαβάζει αυτόματα τις ταχυδρομικές διευθύνσεις
- Δυνατότητα στους ανθρώπους να επικοινωνούν πιο φυσικά, πιο αποτελεσματικά και σε μεγαλύτερο εύρος με τους υπολογιστές τους
- Διατήρηση εθνικής και πολιτιστικής ταυτότητας και διαθεσιμότητα πολιτιστικών θησαυρών σε βιβλιοθήκες και μουσεία σε ολόκληρη την Ευρωπαϊκή Ένωση σε όλους και πέρα από οποιαδήποτε σύνορα
- Δεν εννοούμε τη χρήση τεχνολογικών μέσων π.χ. για εκμάθηση γλώσσας.

Πηγή: clarin:el, <http://www.clarin.gr/el> και EUROMAP/IEA, http://www.ilsp.gr/hope_files

Τι είναι η Γλωσσική Τεχνολογία (ΓΤ) και οι Γλωσσικοί Πόροι (ΓΠ);

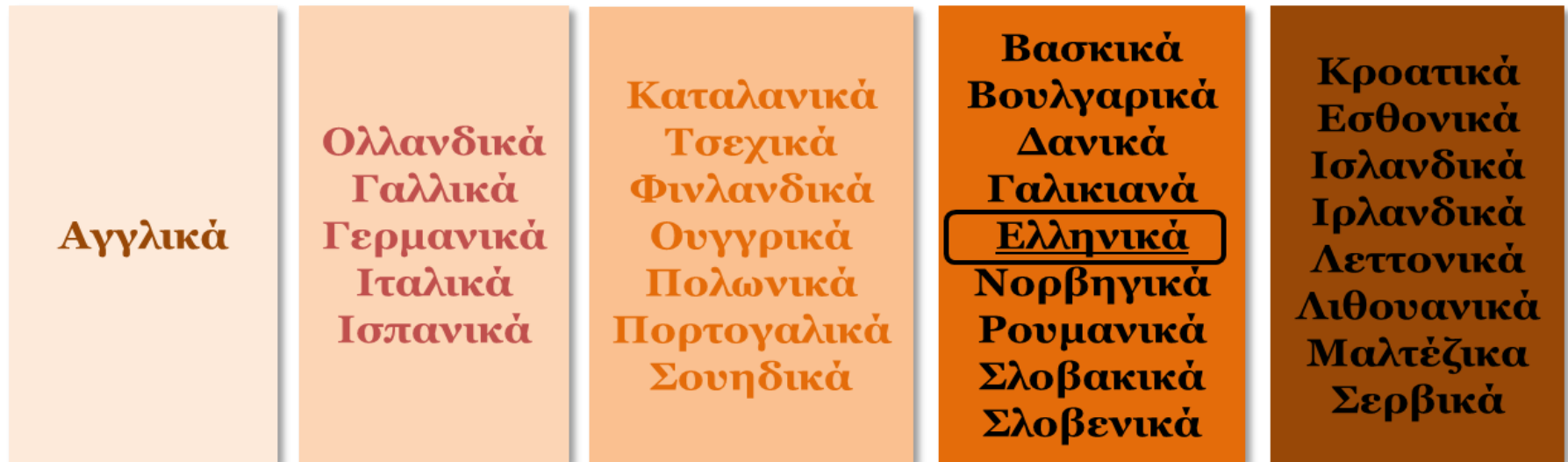
Β. Γλωσσικοί Πόροι:

- Σύνολα γλωσσικών δεδομένων (μονόγλωσσα, διγλωσσα ή πολύγλωσσα) που τροφοδοτούν τα συστήματα ΓΤ
- Πόροι πρωτογενούς περιεχομένου: πόροι ψηφιακού/ψηφιοποιημένου λόγου (π.χ. ψηφιοποιημένα βιβλία, κείμενα από διαδικτυακούς κόμβους, εφημερίδες, σώματα κειμένων κτλ.), ηχογραφήσεις προφορικού λόγου (π.χ. συνεντεύξεις, ραδιοφωνικές εκπομπές κτλ.), βιντεοσκοπήσεις, εικόνες (π.χ. ψηφιακές ή ψηφιοποιημένες φωτογραφίες με τις λεζάντες τους κτλ.)
- Πόροι επεξεργασμένου υλικού: Διαφόρων ειδών επισημειώσεις σε κείμενα, ήχο και πολυμεσικά δεδομένα που έχουν δημιουργηθεί αυτόματα ή με το χέρι (π.χ. μορφοσυντακτικά επισημειωμένα κείμενα, μεταγραφές ηχητικών αρχείων, επισημειωμένα αρχεία βίντεο κτλ.)
- Πόροι αναφοράς: Διάφοροι τύποι οργάνωσης της γνώσης για τη γλώσσα (π.χ. λίστες λέξεων, λεξικά, θησαυροί κτλ.), που χρησιμοποιούνται για την καλύτερη οργάνωση, επεξεργασία και μελέτη των πόρων πρωτογενούς περιεχομένου.

Πηγή: clarin:el, <http://www.clarin.gr/el>



Που βρίσκονται σήμερα τα Ελληνικά σε σχέση με ΓΤ και ΓΠ;



*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

Τι ΓΤ υπάρχει σήμερα για τα Ελληνικά (στην Ελλάδα);

- 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

Τι ΓΤ υπάρχει σήμερα για τα Ελληνικά (στην Ελλάδα); Παραδείγματα:

- META-SHARE (<http://www.meta-share.org/>): Το πρώτο βήμα προς ένα πανευρωπαϊκό δίκτυο κατανεμημένων 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_orologias_el.htm)
- ΕΛΕΤΟ (<http://www.eleto.gr/>)
- Ιστότοποι για τους μεταφραστές (π.χ. <http://www.translatum.gr/>)

Τι ΓΤ υπάρχει σήμερα για τα Ελληνικά (στην Κύπρο); Παραδείγματα:

- (Με κάθε επιφύλαξη) Η Κύπρος δεν διαθέτει εκτεταμένη γλωσσική τεχνολογία, υπό την έννοια της πλούσιας δημιουργίας πόρων ή συστημάτων «αμιγώς» κυπριακών, την έννοια της χρήσης γλωσσικών πόρων από την κυπριακή δημόσια υπηρεσία, ή την έννοια της ποικιλίας πόρων. Ωστόσο:
- Αρχείο προφορικής παράδοσης του Κέντρου Επιστημονικών Ερευνών Κύπρου (ψηφιοποιημένες κασέτες με συνεντεύξεις από την περίοδο 1960 – 2002)
- Λεξιλογική Βάση Δεδομένων της Κυπριακής Διαλέκτου, Πανεπιστήμιο Κύπρου (<http://lexcy.library.ucy.ac.cy/>)
- Λεξικό Κυπριακής αργκό slang.cy, Πανεπιστήμιο Κύπρου (υπό κατάσκευή, Δρ Μαριάννα Κατσογιάννου)

Τι ΓΤ υπάρχει σήμερα για τα Ελληνικά (στην Ελλάδα); Παραδείγματα:

- Ουσιαστική πρόοδος σε πολυγλωσσικές εφαρμογές, ως προς
 - ποιότητα μετάφρασης
 - ταχύτητα μετάφρασης
 - προσαρμογή σε νέους θεματικούς τομείς
 - μεταφερσιμότητα 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

Τι ΓΤ υπάρχει σήμερα για τα Ελληνικά (στην Ελλάδα); Προκλήσεις:

- 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 στη 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

Συμπεράσματα:

Γιατί είναι αναγκαία η συμμετοχή της Κύπρου στην ανάπτυξη Γλωσσικής Τεχνολογίας και Γλωσσικών Πόρων;

- Η Γλωσσική Τεχνολογία θα πρέπει να εμπλουτιστεί στο μέγιστο δυνατό βαθμό για καλύτερα αποτελέσματα, ιδίως μάλιστα σε σχέση με τις γλωσσικές ιδιαιτερότητες της Κύπρου.
- Η Γλωσσική Τεχνολογία θα παρέχει MM ανάμεσα σε όλες τις γλώσσες της ΕΕ, όχι μόνο μεταξύ Ελληνικών και Αγγλικών. Η Κύπρος έχει ανάγκη πολλών γλωσσικών συνδυασμών εκτός EN - EL (διαγλωσσική συνεννόηση).
- Θα καταγραφούν όροι και εκφράσεις με τους/τις οποίους/ες είναι εξοικειωμένη η Κυπριακή Δημόσια Διοίκηση.
- Θα αυξηθεί ο βαθμός κατανόησης ακόμα και ανάμεσα σε έγγραφα που συντάσσονται στην Ελλάδα και στην Κύπρο (ενδογλωσσική συνεννόηση).



**European Language
Resource Coordination**
Connecting Europe Faculty



Ημερίδα εργασίας ELRC ■ 1 Δεκεμβρίου 2015 ■ Σπίτι της Ευρώπης, Λευκωσία

Ευχαριστώ πολύ για την προσοχή σας!



Πανεπιστήμιο Κύπρου
University of Cyprus

ΓΕΩΡΓΙΟΣ ΦΛΩΡΟΣ

Αναπληρωτής Καθηγητής
Πανεπιστήμιο Κύπρου
Σχολή Ανθρωπιστικών Επιστημών, Τμήμα Αγγλικών Σπουδών

gfloros@ucy.ac.cy