European Language Resource Coordination
Connecting Europe Facility

European Commission

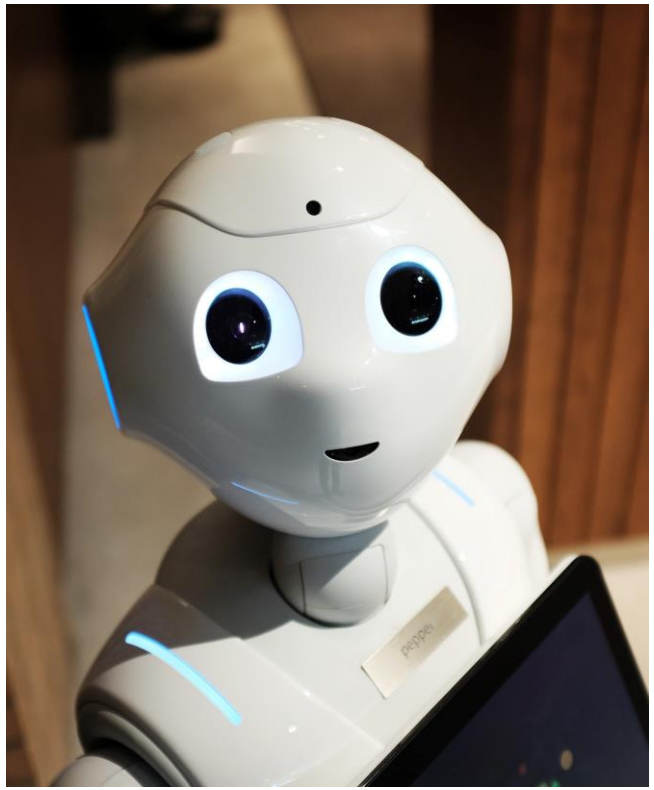
Ο ΕΥΡΩΠΑΪΚΟΣ ΣΥΝΤΟΝΙΣΜΟΣ ΓΛΩΣΣΙΚΩΝ ΠΟΡΩΝ ΓΛΩΣΣΙΚΕΣ ΤΕΧΝΟΛΟΓΙΕΣ ΚΑΙ ΔΕΔΟΜΕΝΑ ΓΙΑ ΤΗΝ ΕΛΛΗΝΙΚΗ

Στέλιος Πιπερίδης
spip@athenarc.gr

3η Ημερίδα του Ευρωπαϊκού Συντονισμού Γλωσσικών Πόρων στην Κύπρο

1 Δεκεμβρίου 2021

Η ΤΕΧΝΗΤΗ ΝΟΗΜΟΣΥΝΗ ΣΤΗΝ ΚΑΘΗΜΕΡΙΝΟΤΗΤΑ ΜΑΣ



- Ψηφιακοί προσωπικοί βοηθοί - digital personal assistants (Siri, Alexa, …)
- Chatbots (π.χ. σε τράπεζες, υποστήριξη και διαχείριση πελατών, ….)
- Αυτόματοι «συγγραφείς»
- Αλληλεπίδραση σε/με υπηρεσίες υγείας
- Marketing (π.χ. Netflix, Amazon etc.)
- Χρηματοοικονομικές και επενδυτικές υπηρεσίες (π.χ. αγοραπωλησίες μετοχών)
- ‘Εξυπνα αυτοκίνητα
- ……

Τι καιρό θα κάνει σήμερα;

Γλωσσικές Τεχνολογίες στο παρασκήνιο…

ΟΙ ΠΡΟΚΛΗΣΕΙΣ ΤΗΣ ΦΥΣΙΚΗΣ ΓΛΩΣΣΑΣ

- Μία λέξη/πρόταση μπορεί να έχει πολλές διαφορετικές σημασίες (αμφισημία)
- Πολλοί διαφορετικοί τρόποι για να εκφραστεί το ίδιο νόημα (ποικιλία)
- Η σημασία εξαρτάται από το περικείμενο (context)
- Κυριολεκτική και μεταφορική χρήση (μεταφορά)

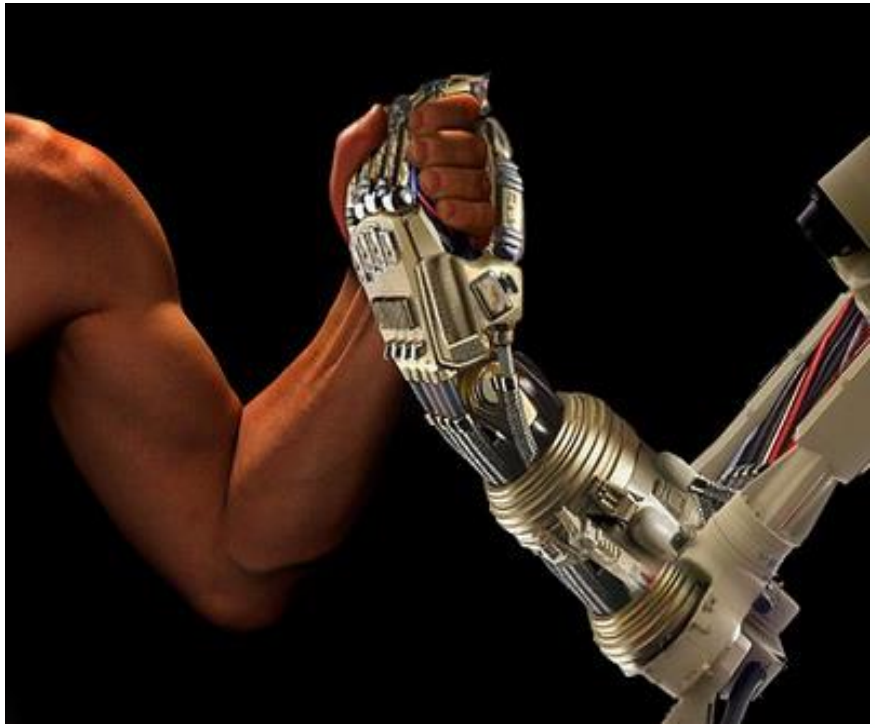


Image: http://workingtropes.lmc.gatech.edu/wiki/index.php/File:Man-vs-machine.jpg
License: CC BY-NC-SA 3.0

ΓΛΩΣΣΙΚΑ ΔΕΔΟΜΕΝΑ – ΓΙΑΤΙ;

- Το καύσιμο της Γλωσσικής Τεχνολογίας και της Τεχνητής Νοημοσύνης είναι τα **γλωσσικά δεδομένα** (σύνολα δεδομένων λόγου σε κάθε μορφή: κειμενικά, ηχητικά, λεξικά/εννοιολογικά δεδομένα, βίντεο…)
- **Γλωσσικά** δεδομένα: ειδική κατηγορία δεδομένων
- Υπάρχουν παντού, παράγονται διαρκώς, για πολλούς λόγους
  - αλλά όχι με στόχο τη γλωσσική επεξεργασία
- Τα συστήματα γλωσσικής τεχνολογίας χρειάζονται επεξεργάσιμα δεδομένα σε ψηφιακή μορφή
  - Σε μεγάλες ποσότητες – και κατά το δυνατόν καλύτερης ποιότητας
  - Σε διάφορες γλώσσες, θεματικές, κειμενικά είδη…

European Language Resource Coordination
Connecting Europe Facility

European Commission

Το Πλαισιο

Ευρωπαϊκός Συντονισμός Γλωσσικών Πόρων (ELRC)

Connecting Europe Facility (CEF)

Χρηματοδοτικό πλαίσιο και εργαλείο της ΕΕ

Διευρωπαϊκές υποδομές

CEF Telecom

Πανευρωπαϊκές ψηφιακές υπηρεσίες

Αυτόματη Μετάφραση (CEF.AT) / eTranslation

Ευρωπαϊκός Συντονισμός Γλωσσικών Πόρων (ELRC)

Τι κανει το ELRC

Προωθεί και Υποστηρίζει	• eTranslation
Παράγει και διαθέτει	• μεταφραστικούς πόρους με χρήση αυτόματων μεθόδων τεχνητής νοημοσύνης και γλωσσικής τεχνολογίας
Παρέχει και υποστηρίζει	• την πλατφόρμα ELRC-SHARE https://www.elrc-share.eu/
Παράγει προδιαγραφές	• για συναφείς τεχνολογίες π.χ ανωνυμοποίηση
Αξιολογεί	• τεχνολογίες πολυγλωσσικής ανάκτησης και εξαγωγής πληροφορίας, μηχανικής μετάφρασης

European Language Resource Coordination
Connecting Europe Facility

European Commission

Το ELRC στην Κυπρο

Εθνικοί εκπρόσωποι στο Συμβούλιο Γλωσσικών Πόρων

Νατάσσα Χαράτση - Αβραμίδη
Γραφείο Τύπου και Πληροφοριών

Δώρα Λοϊζίδου
Πανεπιστήμιο Κύπρου

European Language Resource Coordination
Connecting Europe Facility

European Commission

DIGITAL EUROPE PROGRAMME: A PROPOSED €9.2 BILLION OF FUNDING FOR 2021-2027

ΔΕΔΟΜΕΝΑ – ΓΙΑΤΙ;

- Τα δεδομένα είναι το καύσιμο της οικονομικής ανάπτυξης:
  - για νέα προϊόντα και υπηρεσίες
  - για πιο στοχευμένα και προσωποποιημένα προϊόντα και υπηρεσίες
  - για την αύξηση της παραγωγικότητας σε όλους τους τομείς της οικονομίας
  - για τη διαμόρφωση πολιτικών και την αναβάθμιση των υπηρεσιών προς πολίτες και επιχειρήσεις
- Το όραμα της Ευρώπης:
  - “**Unleash the potential of data with European common data spaces** built on innovative secure and energy efficient cloud to edge technology” (Digital Europe Work Programme 2021-2022)

Αναδιαμόρφωση της πολύγλωσσης Ευρώπης: Γλωσσοκεντρική τεχνητή νοημοσύνη

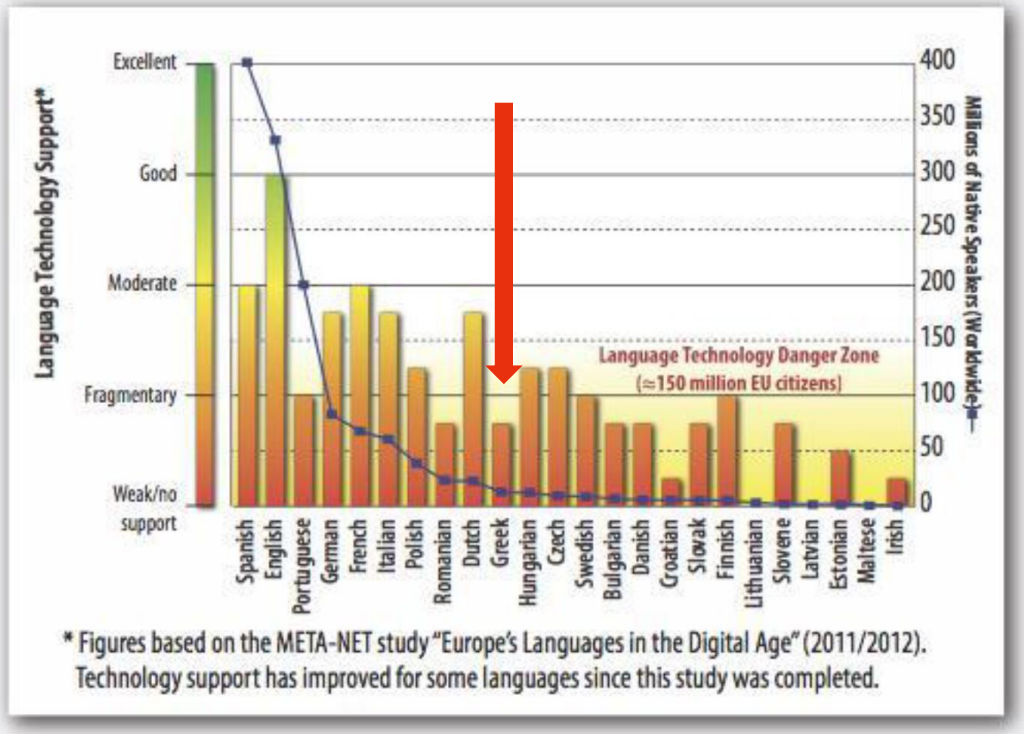
Οι μεγάλες γλώσσες ...

και οι μικρές γλώσσες

The native languages of approximately **140 million EU citizens are in the Language Technology Danger Zone**, where language technology is inadequate to support the DSM.

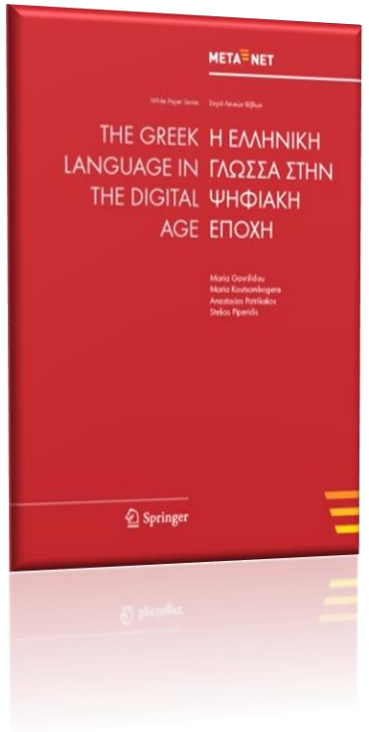
Online Automatic Translation Quality

Current online automatic translation provided by US tech giants does not solve the "language problem": **less than 30% of automatically translated content is truly useful** for online commerce.



* Figures based on the META-NET study "Europe's Languages in the Digital Age" (2011/2012). Technology support has improved for some languages since this study was completed.

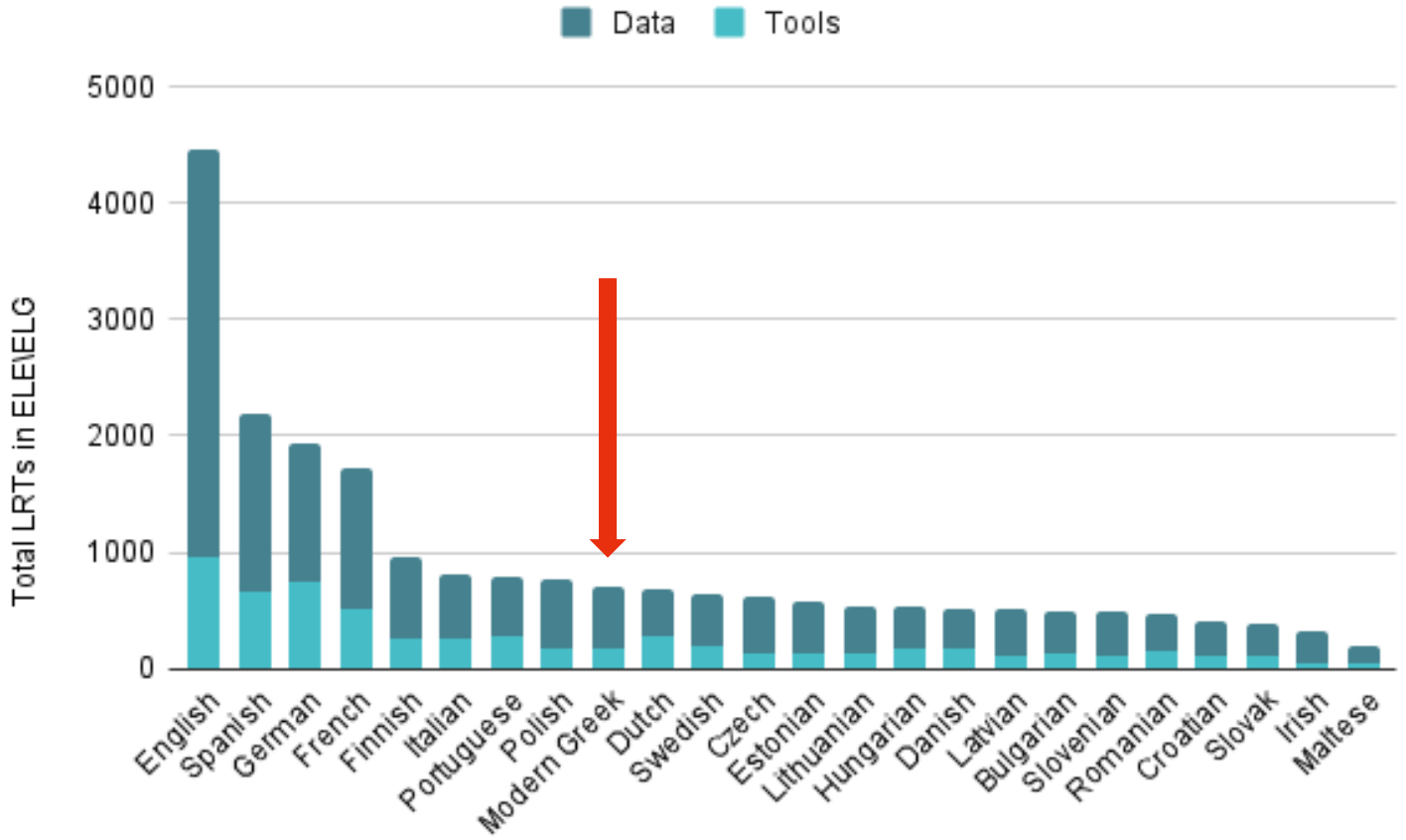
Only three European languages (Spanish, English, and French) meet at least the "moderate" level of language technology support.



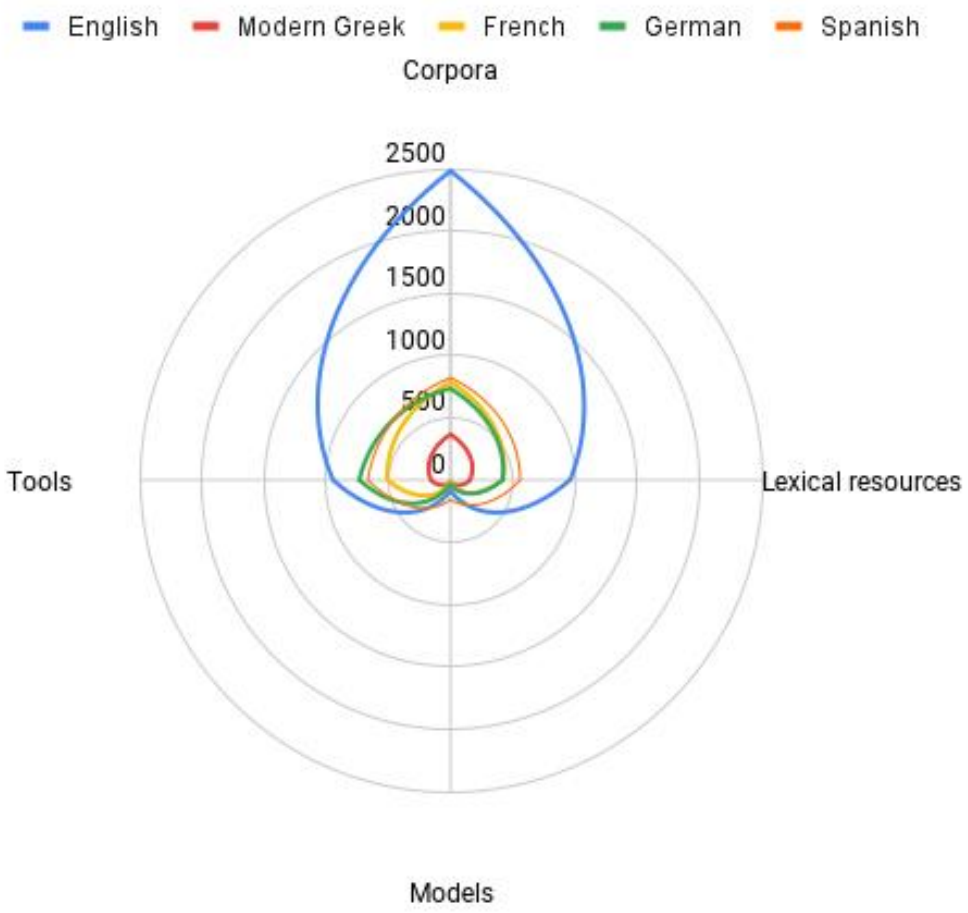
Πηγή: Strategic Research and Innovation Agenda. Language Technologies for a Multilingual Europe. Towards a Human Language Project. SRIA, 2017

EUROPEAN LANGUAGE EQUALITY

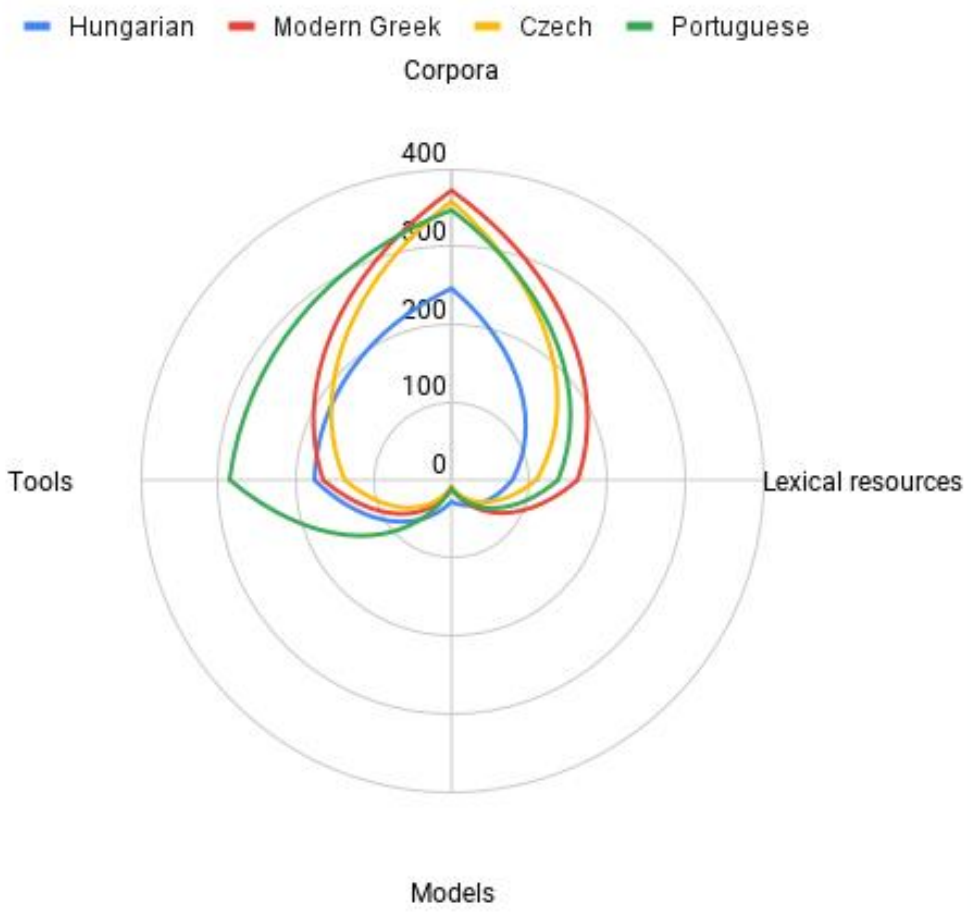
- **Κοινοπραξία:** 53 εταίροι από όλη την Ευρώπη
- **Στόχος:** Κατάρτιση στρατηγικής ερευνητικής ατζέντας και χάρτη πορείας (road map) για την επίτευξη πλήρους ισότητας των γλωσσών (επίσημων ή μη) που χρησιμοποιούνται στην Ευρωπαϊκή Ένωση μέσω της αποτελεσματικής χρήσης γλωσσικών τεχνολογιών μέχρι το 2030
- **Διερεύνηση της τρέχουσας κατάστασης** ως προς την τεχνολογική ετοιμότητα των γλωσσών



Σύγκριση με τις μεγάλες γλώσσες



Σύγκριση με γλώσσες με τον ίδιο περίπου αριθμό ομιλητών



European Language Resource Coordination
Connecting Europe Facility

European Commission

Αποθετηρια και Καταλογοι Γλωσσικων Πορων και Τεχνολογιων

ΕΜΦΑΣΗ ΣΤΗΝ ΕΛΛΗΝΙΚΗ ΓΛΩΣΣΑ

European Language Resource Coordination
Connecting Europe Facility

European Commission

Γλωσσικές Τεχνολογίες από την Ευρωπαϊκή Επιτροπή

https://language-tools.ec.europa.eu/

Connecting Europe
Language Tools

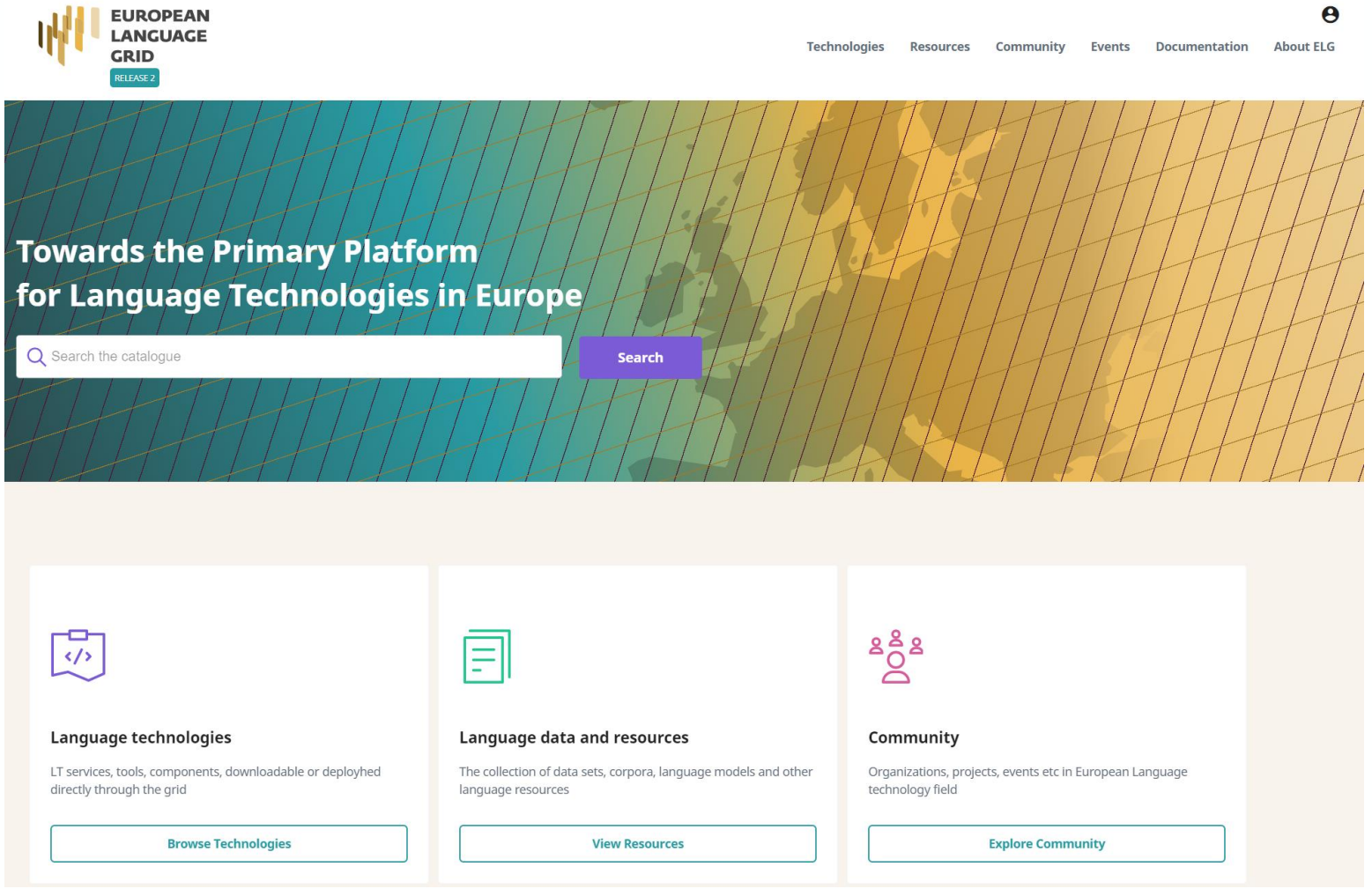
eTranslation	**Multilingual Tweet**	**Speech-to-Text**	**NLP Tools**
Interactive Terminology for Europe	**European Language Resource Coordination (ELRC)**	**Catalogue of services**	**CEF Building Block Information**

Access to some of these tools requires registration. EU staff are pre-registered.
Please visit the registration page: https://webgate.ec.europa.eu/etranslation/public/welcome.html.
For any other issues, please contact help@cefat-tools-services.eu.

European Language Resource Coordination
Connecting Europe Facility

European Commission

European Language Grid
https://live.european-language-grid.eu/



European Language Resource Coordination
Connecting Europe Facility

European Commission

Clarin:el
https://inventory.clarin.gr/

CLARIN:EL portal ›

Help Sign in

clarin:el

Central inventory of language resources and services

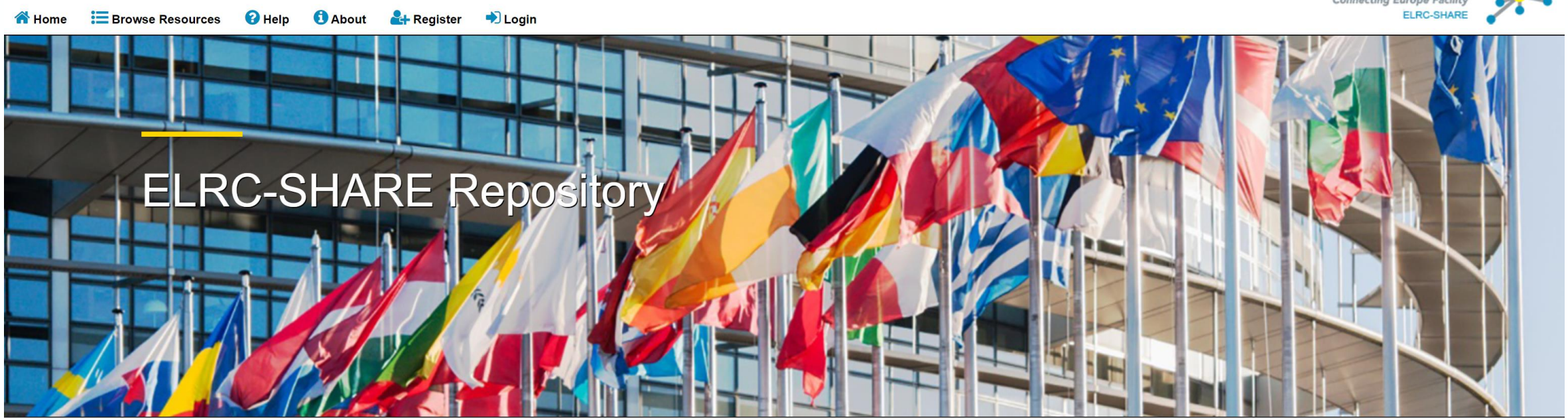
Browse or Search for services, tools, datasets... Search

European Language Resource Coordination
Connecting Europe Facility

European Commission

ELRC-SHARE
https://www.elrc-share.eu/

European Language Resource Coordination
Connecting Europe Facility
ELRC-SHARE



ELRC response to COVID-19 crisis

ELRC is participating in a collaborative COVID-19 initiative, which gathers language resources supporting the development of urgently needed applications and services in relation to the pandemic. Check out the resources collected so far.

Search

Filter by:

- Language
  - Modern Greek (1453-) (188)
  - English (132)
  - French (48)
  - German (37)
  - Spanish; Castilian (35)

more

- Resource Type
- Media Type
- Licence
- Conditions of Use
- Linguality Type
- Multilinguality Type
- Data Format
- Domain

188 Language Resources (Page 1 of 10)

« Previous | Next »

Order by: Resource Name A-Z

Anonymised ParaCrawl release 7 Greek-English
English | Modern Greek (1453-)
0 18
CC0-1.0

Anonymised ParaCrawl release 8 Greek-English
English | Modern Greek (1453-)
0 1
CC0-1.0

Anonymised ParaCrawl release 9 English-Modern Greek (1453-)
English | Modern Greek (1453-)
0 1
CC0-1.0

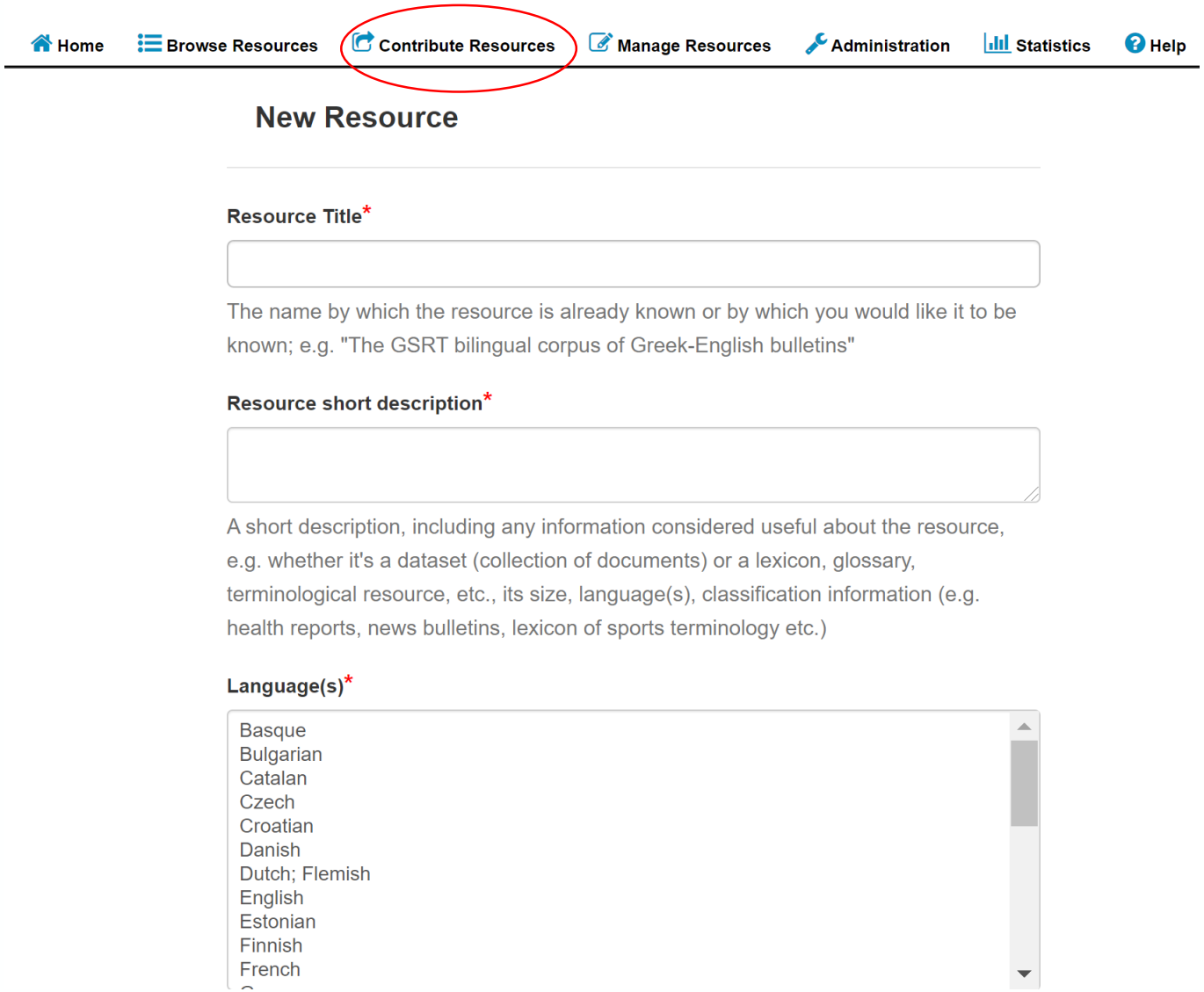
Apache Tika - a content analysis toolkit
Danish | Dutch; Flemish | English | Estonian | Finnish | French | German | Hungarian | Icelandic | Italian | Modern Greek (1453-) | Norwegian Bokmål | Polish | Portuguese | Spanish; Castilian | Swedish
0 74
Apache-2.0

European Language Resource Coordination
Connecting Europe Facility

European Commission

Μοιραστείτε τα γλωσσικά σας δεδομένα

#LanguageDataMatters: η μικρή μας συνδρομή θα αποτρέψει τον ψηφιακό αφανισμό της ελληνικής γλώσσας



European Language Resource Coordination
Connecting Europe Facility

European Commission

ΕΥΧΑΡΙΣΤΩ ΓΙΑ ΤΗΝ ΠΡΟΣΟΧΗ ΣΑΣ!

Στέλιος Πιπερίδης
spip@athenarc.gr