Language Technology for Multilingual Europe

Georg Rehm

Network Manager META-NET
DFKI (German Research Center for Artificial Intelligence), Berlin, Germany

georg.rehm@dfki.de

Human Language Technology Day – Belgrade, Serbia
October 29, 2012

Co-funded by the 7th Framework Programme and the ICT Policy Support Programme of the European Commission through the contracts T4ME, CESAR, METANET4U, META-NORD (grant agreements no. 249119, 271022, 270893, 270899).

Outline

- Introduction
- META-SHARE
- Language White Paper Series
- Strategic Research Agenda
- Recent Developments – Next Steps

Multilingual Europe

META-NET

- **Challenge:** Providing each language community with the most advanced technologies for communication and information so that maintaining their mother tongue does not turn into a disadvantage.
- While research has made considerable progress in recent years, the pace of progress is not fast enough to meet the challenge within the next 10-20 years.
- All stakeholders – researchers, LT user and provider industries, language communities, funding programmes, policy makers – should **team up for a major dedicated push**.

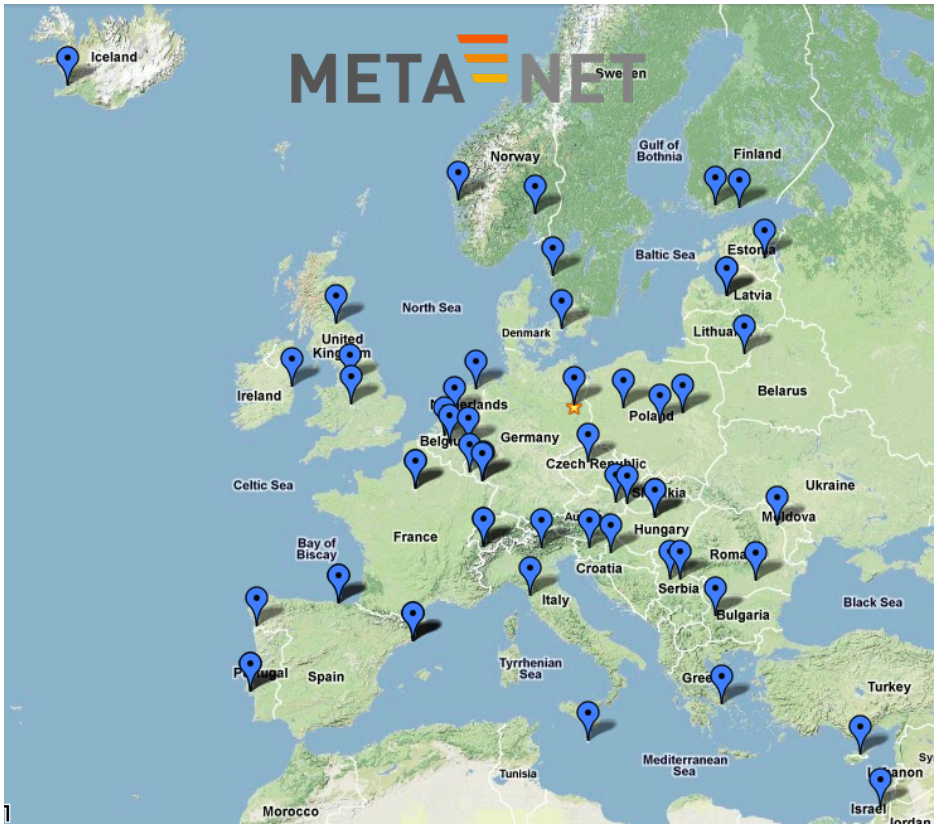
Objectives

META-NET

META-NET is a network of excellence dedicated to fostering the technological foundations of the European multilingual information society.

Four EU-Funded Projects

- Initial project: T4ME (FP7; 13 partners, 10 countries)
- Three ICT-PSP consortia since Feb. 2011: CESAR, METANET4U, META-NORD
- All EU member states and several non-member states covered.
- META-NET in Oct. 2012: **60** members in **34** countries.



http://www.meta-net.eu/members

META-NET

META-SHARE

META-SHARE at a Glance

- Open exchange infrastructure for language resources and tools.
- Language resources and tools are documented, uploaded, stored in repositories, catalogued, can be downloaded, shared, discussed.
- Improve their visibility, documentation, identification, availability, preservation, interoperability.
- Long-term goal: boost research, technology and innovation through wide availability, pooling, openness and sharing of resources.
- Repositories store and maintain inventories of resources and tools.
- Metadata inventories are exported and harvested in the network.
- Currently 13 repositories up and running; ~1.300 LRs available.

META-SHARE portal

Registration - Authentication - Authorisation

Search / Browse | License | Download | User support

Mappings | Reporting/Statistics | Recommendations | Billing/Payment

External repos

META-SHARE inventory

META-SHARE inventory

META-SHARE inventory

Metadata harvesting

Inventory LR repo

Inventory LR repo

Inventory LR repo

Inventory LR repo

META SHARE

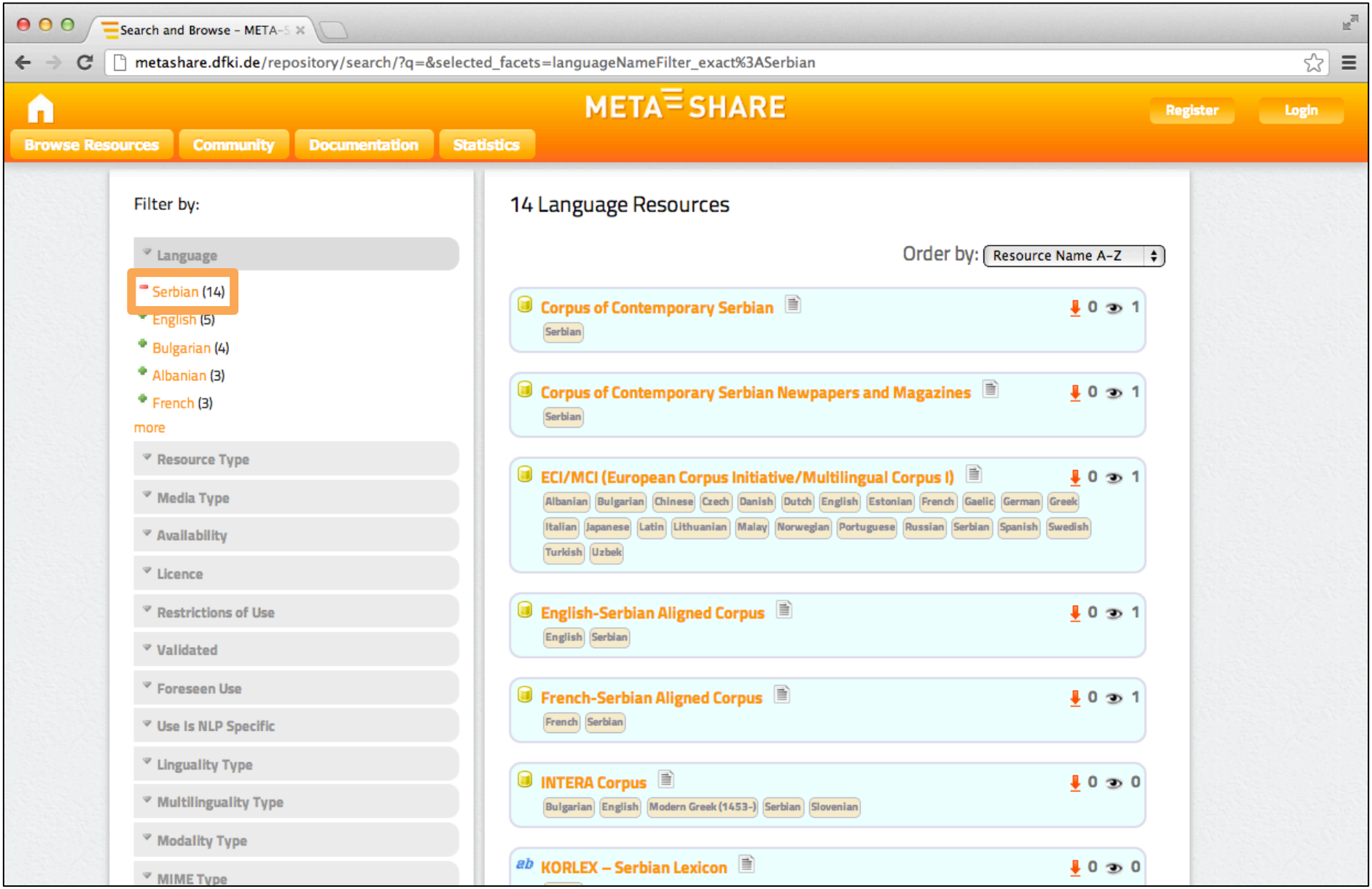
1,248 language resources at your disposal...

Type in your keywords, please...

Search Browse Statistics

What is it? - About the project

META-NET aims at creating META-SHARE, a sustainable network of repositories of language data, tools and related web services documented with high-quality metadata, aggregated in central inventories allowing for uniform search and access to resources. Data and tools can be both open and with restricted access rights, free


Search and Browse – META-S
metashare.dfki.de/repository/search/?q=&selected_facets=languageNameFilter_exact%3ASerbian
META SHARE
Register
Login
Browse Resources
Community
Documentation
Statistics
Filter by:
Language
Serbian (14)
English (5)
Bulgarian (4)
Albanian (3)
French (3)
more
Resource Type
Media Type
Availability
Licence
Restrictions of Use
Validated
Foreseen Use
Use Is NLP Specific
Linguality Type
Multilinguality Type
Modality Type
MIME Type
14 Language Resources
Order by: Resource Name A-Z
Corpus of Contemporary Serbian
Serbian
0 1
Corpus of Contemporary Serbian Newpapers and Magazines
Serbian
0 1
ECI/MCI (European Corpus Initiative/Multilingual Corpus I)
Albanian Bulgarian Chinese Czech Danish Dutch English Estonian French Gaelic German Greek Italian Japanese Latin Lithuanian Malay Norwegian Portuguese Russian Serbian Spanish Swedish Turkish Uzbek
0 1
English-Serbian Aligned Corpus
English Serbian
0 1
French-Serbian Aligned Corpus
French Serbian
0 1
INTERA Corpus
Bulgarian English Modern Greek (1453-) Serbian Slovenian
0 0
KORLEX – Serbian Lexicon
0 0

META-VISION

Language White Paper Series

Language White Paper Series

- Reports on the state of our languages in the digital age and the level of support through language technology.
- Series covers 30 languages.
- Key communication instruments to address decision makers and journalists.
- Inform about societal and technological problems and challenges as well as economic opportunities.
- >2 years in the making.
- >200 national experts as contributors.
- >8.000 copies printed and distributed to politicians and journalists.

30 Languages Covered

- Basque
- Bulgarian*
- Catalan
- Czech*
- Danish*
- Dutch*
- English*
- Estonian*
- Finnish*
- French*
- Galician
- German*
- Greek*
- Hungarian*
- Icelandic
- Irish*
- Italian*
- Latvian*
- Lithuanian*
- Maltese*
- Norwegian
- Polish*
- Portuguese*
- Romanian*
- Serbian
- Slovak*
- Slovene*
- Spanish*
- Swedish*
- Croatian

* = Official EU language

Cross-Lingual Ranking

- In four application areas, each language is assigned to one of five clusters, ranging from *excellent LT support* to *weak/no support*:
  1. Machine Translation
  2. Speech Processing
  3. Text Analysis
  4. Resources
- Results finalised at a meeting in Berlin with representatives of all 30 languages (October 21/22, 2011).

MT

excellent	good	moderate	fragmentary	weak or no support
	English	French, Spanish	Catalan, Dutch, German, Hungarian, Italian, Polish, Romanian	Basque, Bulgarian, Croatian, Czech, Danish, Estonian, Finnish, Galician, Greek, Icelandic, Irish, Latvian, Lithuanian, Maltese, Norwegian, Portuguese, **Serbian**, Slovak, Slovene, Swedish

Text Analysis

excellent	good	moderate	fragmentary	weak or no support
	English	Dutch, French, German, Italian, Spanish	Basque, Bulgarian, Catalan, Czech, Danish, Finnish, Galician, Greek, Hungarian, Norwegian, Polish, Portuguese, Romanian, Slovak, Slovene, Swedish	Croatian, Estonian, Icelandic, Irish, Latvian, Lithuanian, Maltese, **Serbian**

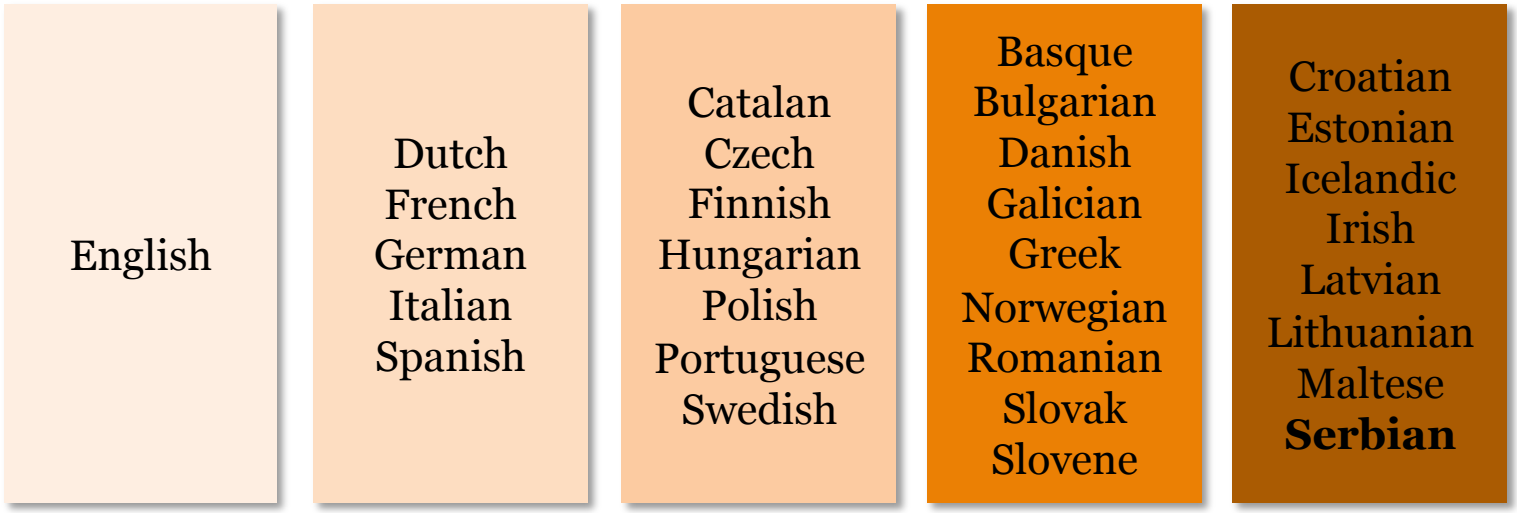
Speech

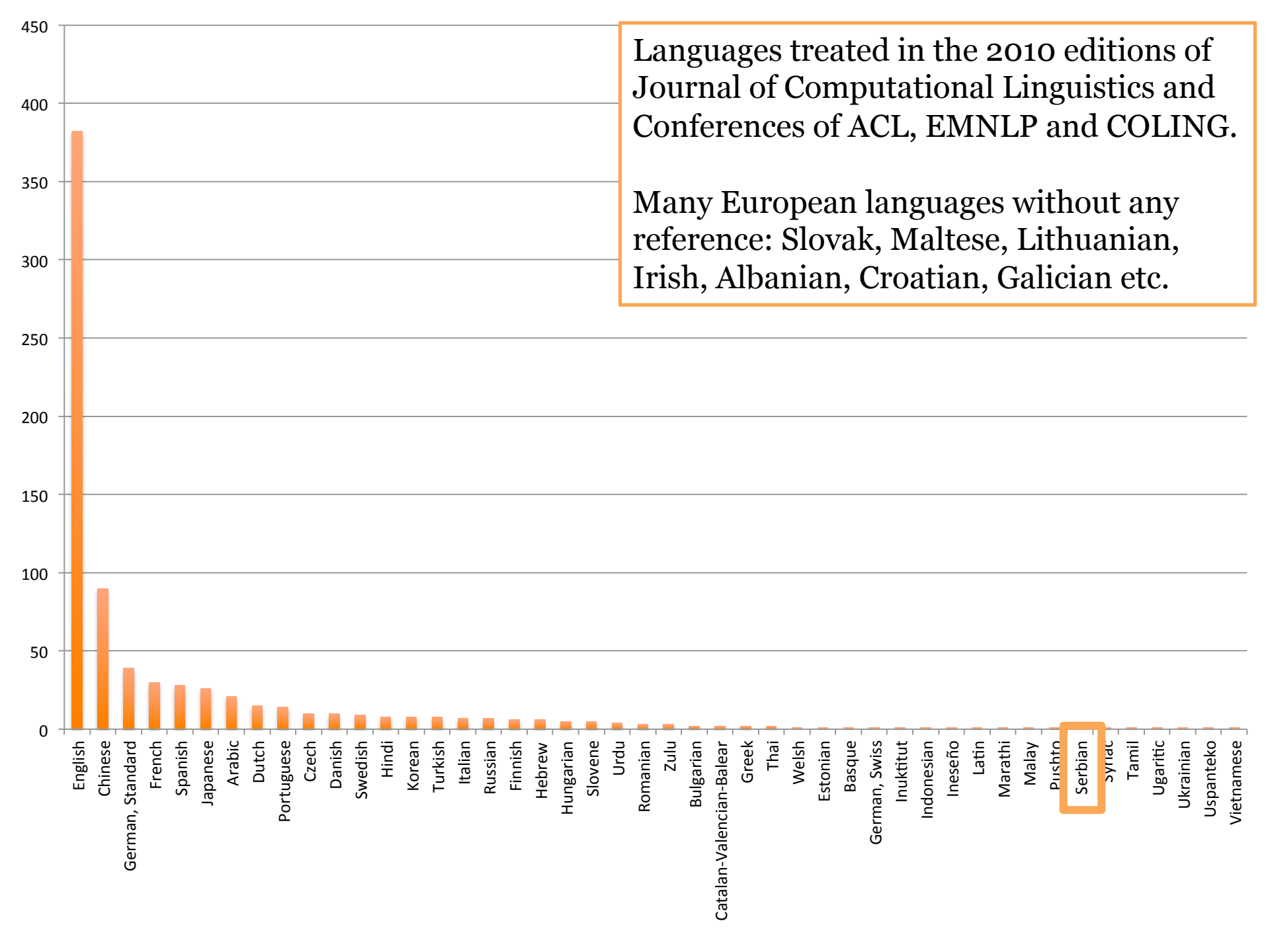
excellent	good	moderate	fragmentary	weak or no support
	English	Czech, Dutch, Finnish, French, German, Italian, Portuguese, Spanish	Basque, Bulgarian, Catalan, Danish, Estonian, Galician, Greek, Hungarian, Irish, Norwegian, Polish, **Serbian**, Slovak, Slovene, Swedish	Croatian, Icelandic, Latvian, Lithuanian, Maltese, Romanian

Resources

excellent	good	moderate	fragmentary	weak/no support
	English	Czech, Dutch, French, German, Hungarian, Italian, Polish, Spanish, Swedish	Basque, Bulgarian, Catalan, Croatian, Danish, Estonian, Finnish, Galician, Greek, Norwegian, Portuguese, Romanian, **Serbian**, Slovak, Slovene	Icelandic, Irish, Latvian, Lithuanian, Maltese

Europe's Languages and LT




Languages treated in the 2010 editions of Journal of Computational Linguistics and Conferences of ACL, EMNLP and COLING.
Many European languages without any reference: Slovak, Maltese, Lithuanian, Irish, Albanian, Croatian, Galician etc.
450
400
350
300
250
200
150
100
50
0
English
Chinese
German, Standard
French
Spanish
Japanese
Arabic
Dutch
Portuguese
Czech
Danish
Swedish
Hindi
Korean
Turkish
Italian
Russian
Finnish
Hebrew
Hungarian
Slovene
Urdu
Romanian
Zulu
Bulgarian
Catalan-Valencian-Balear
Greek
Thai
Welsh
Estonian
Basque
German, Swiss
Inuktitut
Indonesian
Ineseño
Latin
Marathi
Malay
Serbian
Tamil
Ugaritic
Ukrainian
Uspanteko
Vietnamese

White Paper Website

White Paper Press Campaign

- Headline of press release:

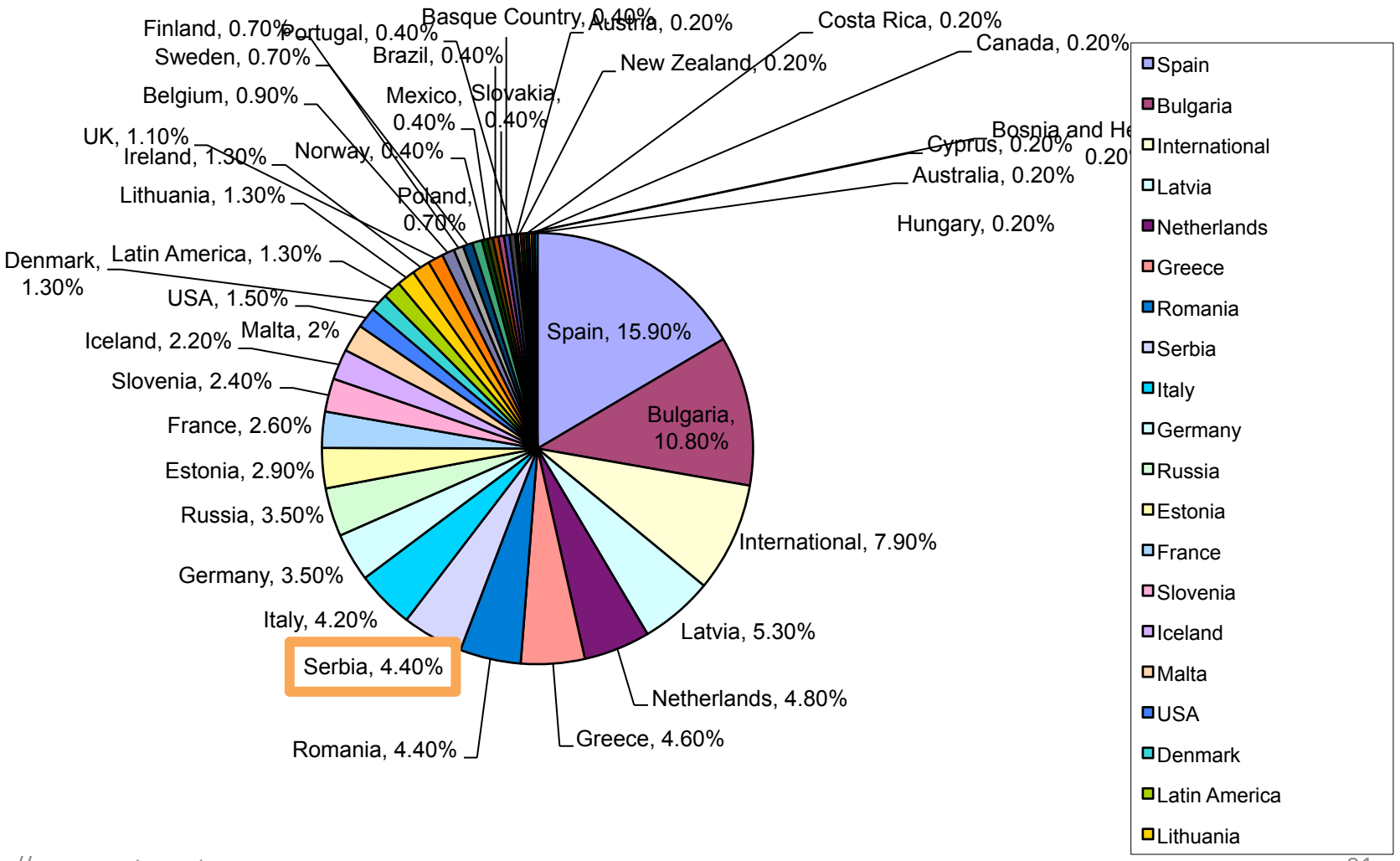
  At Least 21 European Languages in Danger of Digital Extinction. Good News and Bad News on the European Day of Languages.

- Sent out to journalists, politicians and other stakeholder groups before the European Day of Languages (September 26).
- Overwhelmed by the huge interest in the topic and our key findings!
- 470+ mentions in the online and traditional press.
- 40+ interviews with META-NET representatives (television, radio).
- News came in from 41 countries in 35 different languages.

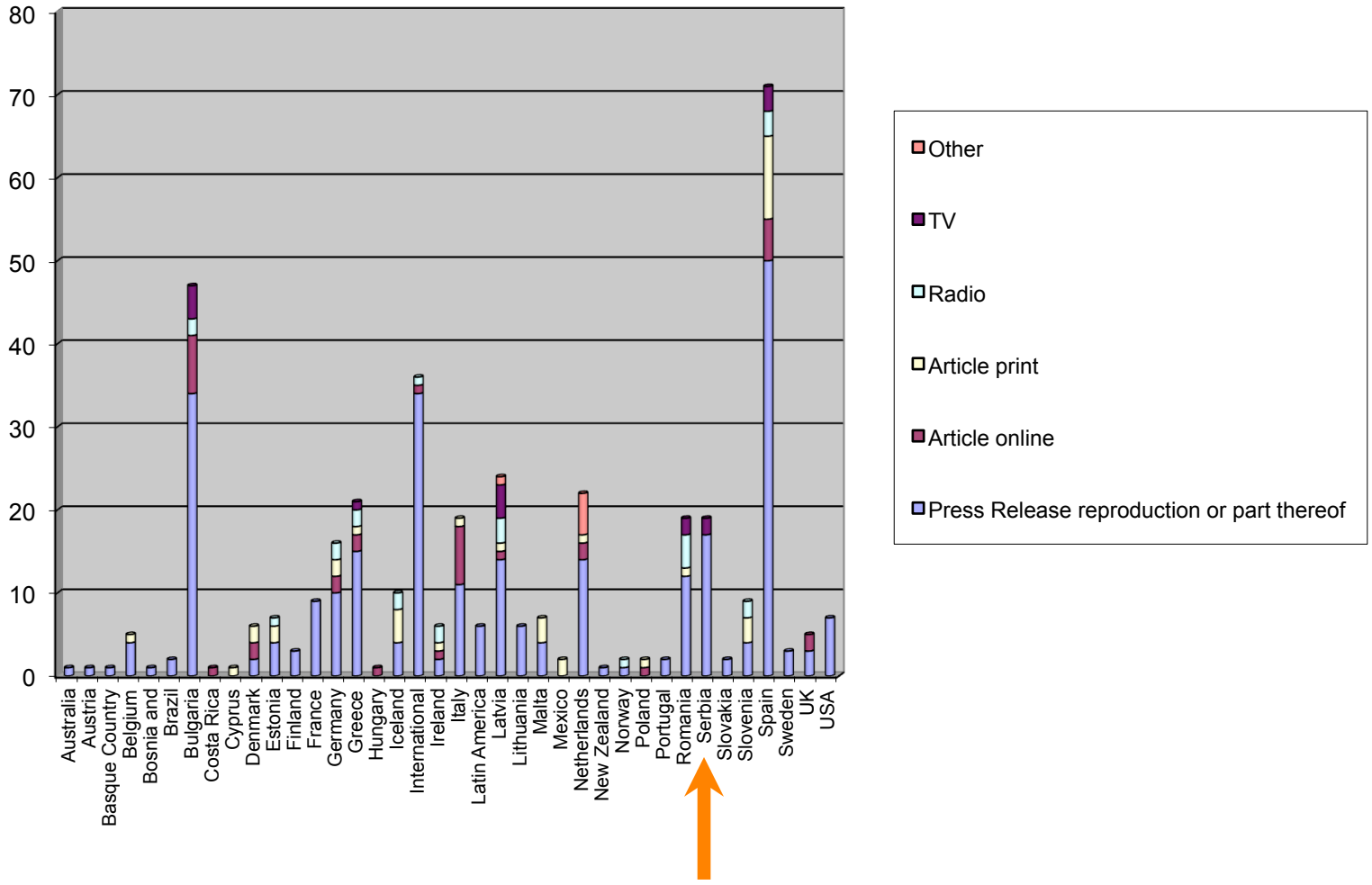
Response: Examples

- **Austria:** Der Standard.
- **Denmark:** Politiken, Berlingske Tidende.
- **Finland:** Tiede.
- **Germany:** Heise Newsticker, Süddeutsche Zeitung.
- **Greece:** in.gr, Πρώτο Θέμα, Prosilipsis.
- **Iceland:** Fréttablaðið, Morgunblaðið.
- **Italy:** Wired.
- **Norway:** Computerworld.
- **Slovenia:** Delo, Dnevnik, Demokracija.
- **Serbia:** Politika, PTC1.
- **Spain:** El Mundo.
- **UK:** Huffington Post.
- **USA:** Mashable, NBC News, Reddit.

Coverage by Country



Coverage by Category



Press Campaign: Highlights

META NET

Ord. Forskere arbejder på at forbedre danske oversættelser på internettet.

Dårlig sprogteknologi truer dansk på nettet

Af Jens Ejsing
// ejs@berlingske.dk

Det danske sprog har det svært i den digitale verden.

Det konstaterer danske sprogforskere- og eksperter i forbindelse med den nye internationale undersøgelse META-NET, der ser nærmere på, hvordan en lang række mindre, europæiske sprog som dansk klarer sig i den digitale verden.

Forskerne fra bl.a. Københavns Universitet og Dansk Sprognævn når frem til, at dansk i fremtiden kan få det endnu sværere i den digitale verden, fordi Google Translate, GPSer, applikationer til smartphones og andre sprogteknologiske programmer ikke i tilstrækkelig grad formår at behandle de mange nuancer i det danske sprog.

Professor i sprogteknologi på Københavns Universitet, Bolette Sandford Pedersen, mener, at der er brug for en slags digital dansk sprogbank fyldt med data, så bl.a. oversættelser bliver så præcise og gode som muligt. Med hjælp fra sprogbanken kan forskere ifølge professoren hjælpe virksomheder med at forbedre programmer, der skal håndtere sproglig viden om bl.a. maskinoversættelse, talegenkendelse og informationssøgning.

Dermed vil der blive længere mellem fejlagtige oversættelser, som når »hæld olie på panden« med Google Translate bliver til »pour oil on the forehead« på engelsk. Oversættelser, der er i værste fald er så upræcise, at danskere ender med at fravælge deres eget sprog i den digitale verden.

Sproghjælp til virksomheder

Hun anerkender dog, at »teknologien til automatiske oversættelser på mange måder er fantastisk«.

»Den er bare ikke god nok, når det gælder dansk,« siger hun:

»Det er som om, at vi i et vist omfang lægger det i hænderne på Google eller andre virksomheder at afgøre, om dansk skal behandles godt nok eller ej. Men det danske marked er ikke stort for dem. Spørgsmålet er derfor, om vi ikke i højere grad selv skal gøre noget for at sikre, at det fornødne datamateriale er til rådighed, så vi får gode oversættelser og anden god sprogteknologi. Det kunne f.eks. være ved, at vi gjorde en indsats for at få oprettet en sprogbank med en masse beriget materiale om dansk.«

»Hvis vi hele tiden oplever, at oversættelser er behæftede med fejl, tør vi ikke stole på dem,« siger hun og understreger, at »fejlagtige oversættelser kan føre til store misforståelser«.

Ifølge Dansk Sprognævns direktør, Sabine Kirchmeier-Andersen, kan dårlig sprogteknologi have konsekvenser for mange danskere, der ikke er så gode til engelsk.

»Hvis vi har ambitioner om at bruge det danske sprog i fremtidens teknologiske univers, skal der gøres en indsats nu for at fastholde ekspertise og udbygge den viden, vi har,« mener hun:

»Ellers risikerer vi, at kun folk, der taler flydende engelsk, vil få glæde af de nye generationer af web-, tele- og robotteknologi, der er på vej.« B

fakta

Sprog i Europa

- Der er omkring 80 sprog i EU. For 21 af dem – også dansk – gælder det, at der er store sprogteknologiske mangler, når det gælder bl.a. maskinoversættelse, talegenkendelse og informationssøgning.
- Ifølge en EU-undersøgelse køber et stigende antal europæiske internetbrugere varer eller tjenester på nettet, hvor det sprog, der bliver anvendt, ikke er deres eget. Det gælder over halvdelen af brugerne.
- Over hver tredje anvender et fremmedsprog til at skrive mail eller indlæg på nettet.

Press Campaign: Highlights

ΠΟΛΛΕΣ ΕΥΡΩΠΑΪΚΕΣ ΓΛΩΣΣΕΣ ΘΕΩΡΟΥΝΤΑΙ ΤΕΧΝΟΛΟΓΙΚΑ... ΞΕΠΕΡΑΣΜΕΝΕΣ

Με ψηφιακή εξαφάνιση κινδυνεύουν τα ελληνικά

Στην ψηφιακή εποχή δεν... μιλούν ελληνικά, όπως και αρκετές άλλες ευρωπαϊκές γλώσσες, σύμφωνα με πανευρωπαϊκή έκθεση με την υπογραφή 200 και πλέον ειδικών. Η συγκεκριμένη μελέτη δημοσιεύτηκε από το επιστημονικό δίκτυο META-NET με αφορμή τη χτεσινή Ευρωπαϊκή Ημέρα Γλωσσών.

Για τις ανάγκες της έρευνάς τους, γλωσσολόγοι από 34 χώρες της Ηπείρου βαθμολόγησαν τις διαθέσιμες γλωσσικές υπηρεσίες και δημιούργησαν ένα «Λευκό Βιβλίο» για κάθε ευρωπαϊκή γλώσσα. Στη μελέτη τους, οι ειδικοί αναζήτησαν μεταξύ άλλων τέσσερα βασικά πλεκτρονικά εργαλεία, δηλαδή την ύπαρξη αυτόματης μετάφρασης, τη δυνατότητα φωνητικής αλληλεπίδρασης και ψηφιακής ανάλυσης κειμένου, ενώ ταυτόχρονα διερευνήθηκε και η διαθεσιμότητα γλωσσικών πόρων ή πηγών.

Σε πρώτη φάση εξέτασαν τις ιστοσελίδες που επιτρέπουν στους χρήστες να κάνουν μεταφράσεις online, όπως, για παράδειγμα, η υπηρεσία του κολοσσού πληροφορικής Google Translate. Την ίδια ώρα, εξετάστηκε και η «επικοινωνία» των ελληνόφωνων χρηστών με τις... συσκευές τους, όπως για παράδειγμα η δυνατότητα

ΕΛΕΝΗ ΒΕΡΓΟΥ
evergou@e-typos.com

να «μιλήσει» κάποιος στο GPS στη μητρική του γλώσσα. Οι ερευνητές κατέληξαν στο συμπέρασμα ότι υπάρχουν τέτοιες συσκευές, αλλά δεν είναι τόσο διαδεδομένες όσο οι αγγλόφωνες.

Το «χρυσό» μετάλλιο κατακτά, όπως είναι άλλωστε και λογικό, η αγγλική γλώσσα. Οι αγγλόφωνοι χρήστες έχουν την καλύτερη δυνατή τεχνολογική υποστήριξη, κάτι το οποίο ευνοεί την περαιτέρω εξάπλωση της γλώσσας. Από «τεχνολογικό αποκλεισμό» κινδυνεύουν περισσότερο η ολανδική, η λετονική, η λιθουανική και η μαλτέζικη γλώσσα, ενώ σε λίγο καλύτερη μοίρα βρίσκονται η ελληνική, η βουλγαρική, η ουγγρική και η πολωνική, που όπως αναφέρει η έρευνα έχουν «αποσπασματική» τεχνολογική υποστήριξη.

«Μέτρια» χαρακτηρίζεται η υποστήριξη χρηστών σε αλλανδική, γαλλική, γερμανική, ιταλική και ισπανική γλώσσα. Οι επικεφαλής της επιστημονικής ομάδας, **Χανς Ουζκοράιτ** και **Γκεόργκ Ρεμ**, αναφέρουν χαρακτηριστικά: «Υπάρχουν δραματικές διαφορές στην υποστήριξη της γλωσσικής τεχνολογίας ανάμεσα στις διάφορες ευρωπαϊκές γλώσσες. Το χάσμα μεταξύ "μικρών" και "μεγάλων" γλωσσών ολοένα και διευρύνεται. Πρέπει να εξασφαλίσουμε τον εφοδιασμό των μικρότερων και λιγότερο πλούσιων σε ψηφιακούς πόρους γλωσσών με τις απαραίτητες βασικές τεχνολογίες. Διαφορετικά, οι γλώσσες αυτές είναι καταδικασμένες σε ψηφιακή εξαφάνιση».

Μάλιστα, οι ειδικοί τονίζουν ότι χωρίς αποφασιστική δράση οι γλώσσες αυτές δύσκολα θα... επιβιώσουν στον ψηφιακό κόσμου του 21ου αιώνα. Η κ. **Μαρία Γαβριηλίδου**, μέλος της επιστημονικής ομάδας από το Ινστιτούτο Επεξεργασίας του Λόγου Ερευνητικό Κέντρο Αθηνά, λέει στον «Ε.Τ.»: «Η έρευνα αυτή δεν λέει ότι δεν θα ζήσει η ελληνική γλώσσα ή ότι κινδυνεύει με εξαφάνιση». Η ειδικός εξηγεί ότι όσο υπάρχουν άνθρωποι που μιλάνε, γράφουν και επικοινωνούν με μια γλώσσα, τόσο αυτή θα συνεχίσει να υπάρχει. Είναι σημαντικό, όμως, να έχουν όλοι οι χρήστες τη δυνατότητα «μιλήσουν» στις μπακιές, όπως τα GPS τους, στα ελληνικά και να έχουν στη διάθεσή τους γλωσσικά εργαλεία πλεκτρονικών υπολογιστών.

> Οι αγγλόφωνοι χρήστες έχουν την καλύτερη δυνατή τεχνολογική υποστήριξη, γεγονός που ευνοεί την περαιτέρω εξάπλωση της γλώσσας

Μεταξύ αυτών των «εργαλείων» είναι οι διορθωτές ορθογραφικών και συντακτικών λαθών, που χρησιμοποιούνται καθημερινά από εκατοντάδες Ελληνες χρήστες και βασίζονται στη γλωσσική τεχνολογία.

Παρ' όλα αυτά, τονίζει ότι η ψηφιακή εξάπλωση μιας γλώσσας είναι σημαντική. «Δεν είναι στα χέρια του μέσου χρήστη. Οι εκάστοτε κυβερνήσεις, η Ευρωπαϊκή Ενωση και ο ιδιωτικός τομέας πρέπει να χρηματοδοτήσουν την ανάπτυξη αυτής της τεχνολογίας για όλες τις γλώσσες», αναφέρει και συνεχίζει: «Οι χρήστες, όμως, πρέπει να απαιτούν να υπάρχουν και στη γλώσσα τους τα μέσα αυτά και να μην ικανοποιούνται με τα αγγλικά». ■


Γιώργος Μπαμπινιώτης

▸▸ GREEKLISH

Η γλώσσα της αποξένωσης...

ΜΕ GREEKLISH επικοινωνούν πλέον μέσω μηνυμάτων ή email οι περισσότεροι νέοι της χώρας μας. Παρά το γεγονός ότι τα τελευταία χρόνια υπάρχουν τα γλωσσικά εργαλεία, τα οποία επιτρέπουν τη χρήση της ελληνικής γραμματοσειράς, έφηβοι και νέοι ενήλικες φαίνεται ότι δεν έχουν «αγκαλιάσει» αυτές τις τεχνολογίες. Ο καθηγητής Γλωσσολογίας, κ. Γιώργος Μπαμπινιώτης, λέει στον «Ε.Τ.»: «Το greeklish είναι πρόβλημα για την ελληνική γλώσσα, ιδίως για ανθρώπους νέας ηλικίας για έναν καθαρά γλωσσικό λόγο. Με τη χρήση των greeklish αποξενώνονται από τη μορφή της λέξης ή όπως λέμε το ετυμολογικό ίνδαλμα που δηλώνεται με την ορθογραφία της λέξης και συνδέεται και με τη σημασία της λέξης και με την προέλευσή της». Ο κίνδυνος, με τον οποίο έρχονται αντιμέτωποι οι νέοι άνθρωποι, είναι η αποξένωση από τη γραπτή μορφή της γλώσσας. Αυτή η «οικειότητα», όμως, βοηθάει και στην κατανόηση της σημασίας αλλά και την προέλευση της λέξης. «Αυτή η αποξένωση δεν είναι άνευ σημασίας», αναφέρει ο ειδικός, ο οποίος εξηγεί ότι η διαδικασία της γραφής βοηθάει να εντυπωθεί η λέξη και να συνδεθεί με άλλες ομόρριζες λέξεις. «Οταν χρησιμοποιείται αυτή η μορφή επικοινωνίας, καταστρέφονται, ατονούν. Δεν είναι προς θάνατο, αλλά θα κάνει ζημιά», αναφέρει ο κ. Μπαμπινιώτης, ο οποίος συμβουλεύει τους χρήστες να επιλέγουν την ελληνική γραμματοσειρά.

THE SUNDAY TIMES
Date 30 September 2012
Page 16
eclips newspaper licensing agency

BRIEFING DIGITAL IRISH

Lost for words

The Irish language is one of dozens of European tongues in danger of electronic extinction due to lack of digitised content. How can Gaeilge jump the digital divide, asks **John Burns**

BACKGROUND
Some European languages lagging

Research has been carried out into how 30 of Europe's estimated 80 languages are coping in the digital age, and the conclusion is most are unlikely to survive it. Language technology (LT) researchers estimate that digital support for 21 of the tongues is "non-existent" or "weak". The study by Meta-Net, a European network of 60 research centres in 34 countries, concluded that most of the languages (such as Basque and Catalan) are under-resourced, while some are completely neglected. Irish is among those at risk of being left behind, but "with a strong pedigree in translation, localisation and LT research, Ireland is in a strong position to develop such technologies", according to the authors, including Dr John Judge of Dublin City University.

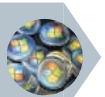
INTERACTION
Irish-language users do have digital tools

LT covers everything from translating web pages to operating a mobile phone with spoken commands. There is software support for the Irish language, for example Google, Microsoft Window and Mozilla Firefox all have localised interfaces. MS Office has a spell-checker, although it hasn't been updated recently, and there are online tools for spelling and grammar such as GaelSpell. "There are localised versions of word processors, internet browsers and other desktop tools as well as operating systems," said Judge. "These technologies are being used in the Irish-speaking community." However there is no automatic speech recognition in Irish, and "there is a long way to go" before such systems can be put in place, says the study.

TONGUE TWISTER
Irish has unique quirks

The Irish language has distinctive features which make speech- and text-based technologies tricky. Unlike most European languages, it has a verb-subject-object structure, and adjectives follow nouns. Irish has no words for "yes" or "no". To give a yes or no response, you must repeat the verb. There are two forms of "to be". Instead of a standard spoken form, there are three distinct dialects — Donegal, Connemara and Munster. As the standard spoken form has no native speakers — being a compromise of the three dialects — it "does not suit the building of a speech-synthesis system, which has to be based on a specific speaker". There is also an absence of resources which text-based technologies need, such as pronunciation dictionaries.

SPEAK UP
Urgent work needed to help Irish



Irish is largely out of the funding loop because it does not represent a market that investors are interested in, the report points out. Microsoft, Google and IBM all have centres for linguistics, and are based in Ireland, "but do very little work on Irish language data". Being an official EU language affords Irish some protection, according to Judge. "But we desperately need to start producing more core component tools like parsers, chunkers, speech recognisers and synthesisers," he said. "In order to do this we need the 'raw materials', ie linguistic data in a machine-usable form. There is a lack of continuity in the maintenance of what little we have, which means existing resources run the risk of becoming obsolete."

Press Campaign: Highlights

Latviešu valoda apdraudēta

Aisma ORUPE, tālr. 67886751

Eiropas Savienībā (ES) ir aptuveni 60 runāto valodu, taču – cik ilgi tās kopiena varēs lepoties ar tādu kultūras un valodu bagātību? Pētījumi liecina, ka digitalizācijas laikmetā vismaz 21 no tām draud izmiršana. Arī latviešu valodai izdzīvot nebūs viegli, un valodu tehnoloģiju jomā tai nav daudz, ar ko lepoties – vairākumā sadaļu tā ir pēdējās vietās.

Starptautiskajā konferencē *Valoda, tehnoloģijas un Eiropas nākotne*, kur tika vērtēti valodas tehnoloģiju sasniegumi un problēmas, vairākkārt tika uzsvērts, ka ES valstu dažādās valodas ir gan bagātība, gan arī barjera, kura traucē saziņu un sadarbību, kas sniedzas pāri valstu robežām, kavē preču un pakalpojumu eksportu, uzņēmumu spēju strādāt ārējos tirgos. «Lai Eiropa saglabātu savu vadošo lomu pasaulē, tai būs vajadzīgas visām Eiropas valodām pielāgotas valodu tehnoloģijas, kas būs viegli pieejamas un efektīvas,» norādīja Vācijas Mākslīgā intelekta pētījuma centra zinātniskais direktors Hanss Uškoreits, piebilstot, ka tas nozīmē, ka šajā jomā ES nāksies ieguldīt lielus līdzekļus. Ja 2008. gadā tulkošanai, programmatūru lokalizācijai un tīmekļa vietņu globalizācijas izmaksas Eiropas tirgū bija 8,4 miljardi eiro, tad tā ir maza daļa no tēriņiem, kas būs nepieciešami saziņas nodrošināšanai starp valodām. Un, lai arī valodu tehnoloģijas ir ievērojami progresējušas, jaunu produktu izstrādes temps tomēr ir pārāk gauss. Piemēram, tiešsaistes mašīntulkošana (piemēram, *Google*) piedāvā ātru aptuvenu tulkojumu, taču tas neder, ja vajadzīgs precīzs un kvalitatīvs tulkojums.

Konferencē prezentētajā pētījumā *Latviešu valoda digitālajā laikmetā*, kurā piedalījušies 200 ekspertu, vilktas paralēles starp diviem laikmetiem: kad parādījās grāmatu spiede un pašreizējo – digitālo. Tās valodas, kurās netika drukātas grāmatas, ar laiku izmira, un tas varot notikt arī ar valodām, kas netiek lietotas globālajā tīmeklī. Patlaban privileģētākā stāvoklī esot plašāk lietotās Eiropas valodas, jo īpaši – angļu valoda. Salīdzinošā analīze rāda, ka tikai tai sniegtais atbalsts vērtējams kā labs. Holandiešu, franču, vācu, itāliešu un spāņu valodai tas ir viduvējs, bet skaitliski nelielās Eiropas valodas, to skaitā lietuviešu un latviešu, vērtētas kā visapdraudētākās, liecina pētījums. Latviešu valoda četros sektoros (runas un teksta resursi, tehnoloģijas runas apstrādei, mašīntulkošanai un teksta analīzei) ierakstīta sadaļā: vājš atbalsts vai nav atbalsta. Jāteic, ka igauņi Latviju kārtējo reizi ir apsteiguši. Kadrija Videre no Igauņu valodas resursu centra skaidroja, ka tas noticis, jo Igaunijā valodu tehnoloģiju atbalstam ir ilgstošs un pastāvīgs valsts finansējums.

Latvijas pārstāve datorzinātņu doktore Inguna Skadiņa bija spiesta atzīt, ka kopumā neizskatāmies pārāk labi. «Līdzšinējais valsts atbalsts bijis nepietiekams, un tas novedis pie diezgan lielas atpalicības valodas ilgtspējai nepieciešamo resursu un rīku jomā. Arī pašreizējais pētniecības darbs bijis fragmentārs. Tāpat problēmas rada datorlingvistikas mācību trūkums Latvijas augstskolās, un šobrīd tādu kursu var apgūt tikai Liepājas Universitātē un Rēzeknes augstskolā,» rezumēja I. Skadiņa, norādot, ka tikai ilgtspējīga politika un valsts atbalsts var glābt latviešu valodu. ■

Atbalsts valodu tehnoloģijām (30 Eiropas valodās)

Sektors	labs	viduvējs	fragmentārs	vājš vai nav
Runas apstrāde	1 (angļu)	9 (vācu, spāņu u. c.)	14 (igauņu, poļu u. c.)	6 (latviešu, lietuviešu u. c.)
Mašīntulkošana	1 (angļu)	2 (franču, spāņu)	7 (itāliešu, poļu)	19 (latviešu, lietuviešu u. c.)
Teksta analīze	1 (angļu)	5 (franču, itāliešu u. c.)	14 (somu, basku u. c.)	8 (latviešu, lietuviešu u. c.)
Runas un teksta resursi	1 (angļu)	9 (čehu, franču u. c.)	15 (igauņu, dāņu u. c.)	5 (latviešu, lietuviešu u. c.)

LIETOTĀJI

Latviešu valodas lietotāju skaits internetā 1,3 miljoni

Pēdējo 6 mēnešu laikā internetu lietojuši 64% iedzīvotāju

64%

Vecumā no 15 līdz 74 gadiem

Foto: Dmitrijs SALĀNČS, *F64 Photo Agency*

LATVIEŠU VALODAS ILGTSPĒJAS

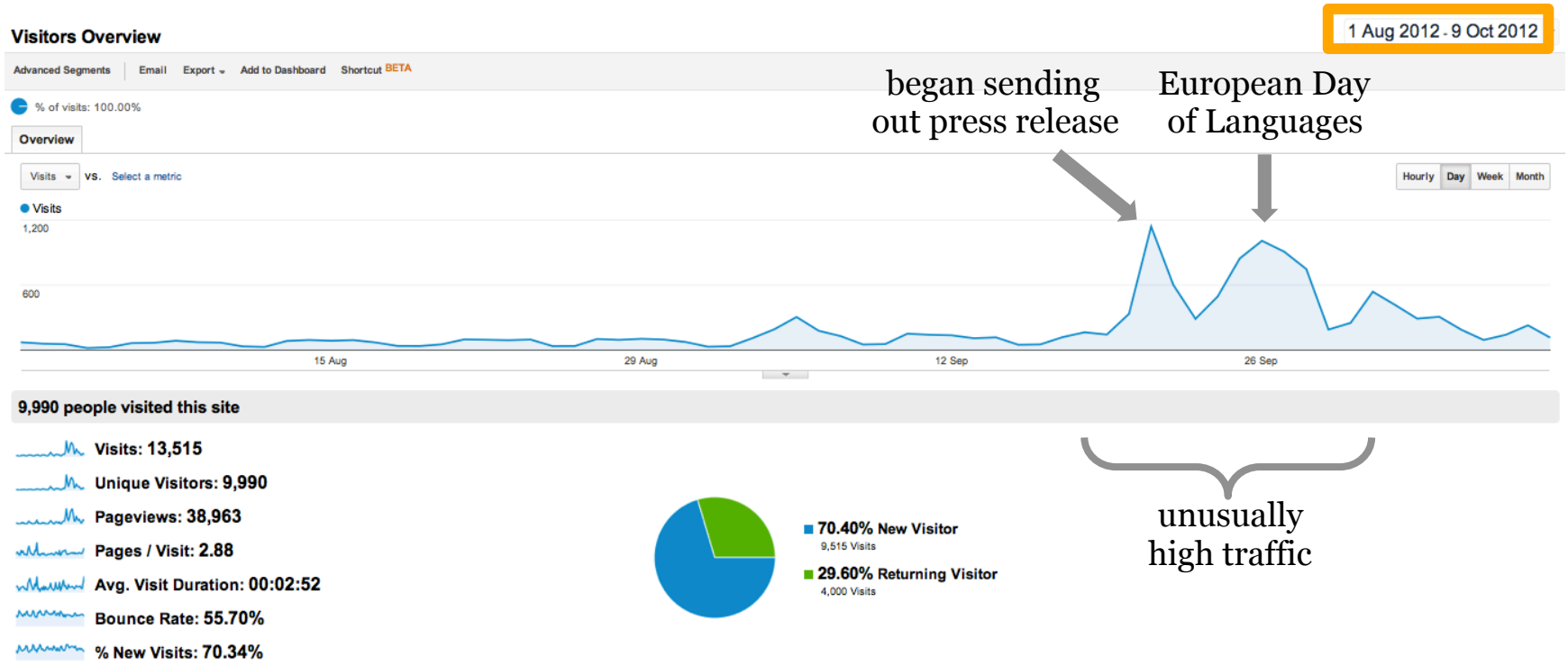
POZITĪVIE ASPEKTI:
- ES oficiālā valoda
- Diezgan liels lietotāju skaits
- Valodai ir liela lingvistiskā kvalitāte
- ES fondu pieejamība projektos

RISKI:
- Mazs tirgus
- Latviešiem saglabājies minoritātes komplekss
- Demogrāfiskā situācija
- Ekonomiskā atpalicība
- Emigrācija
- Minoritāšu pašpietiekamība
- Krievijas spiediens

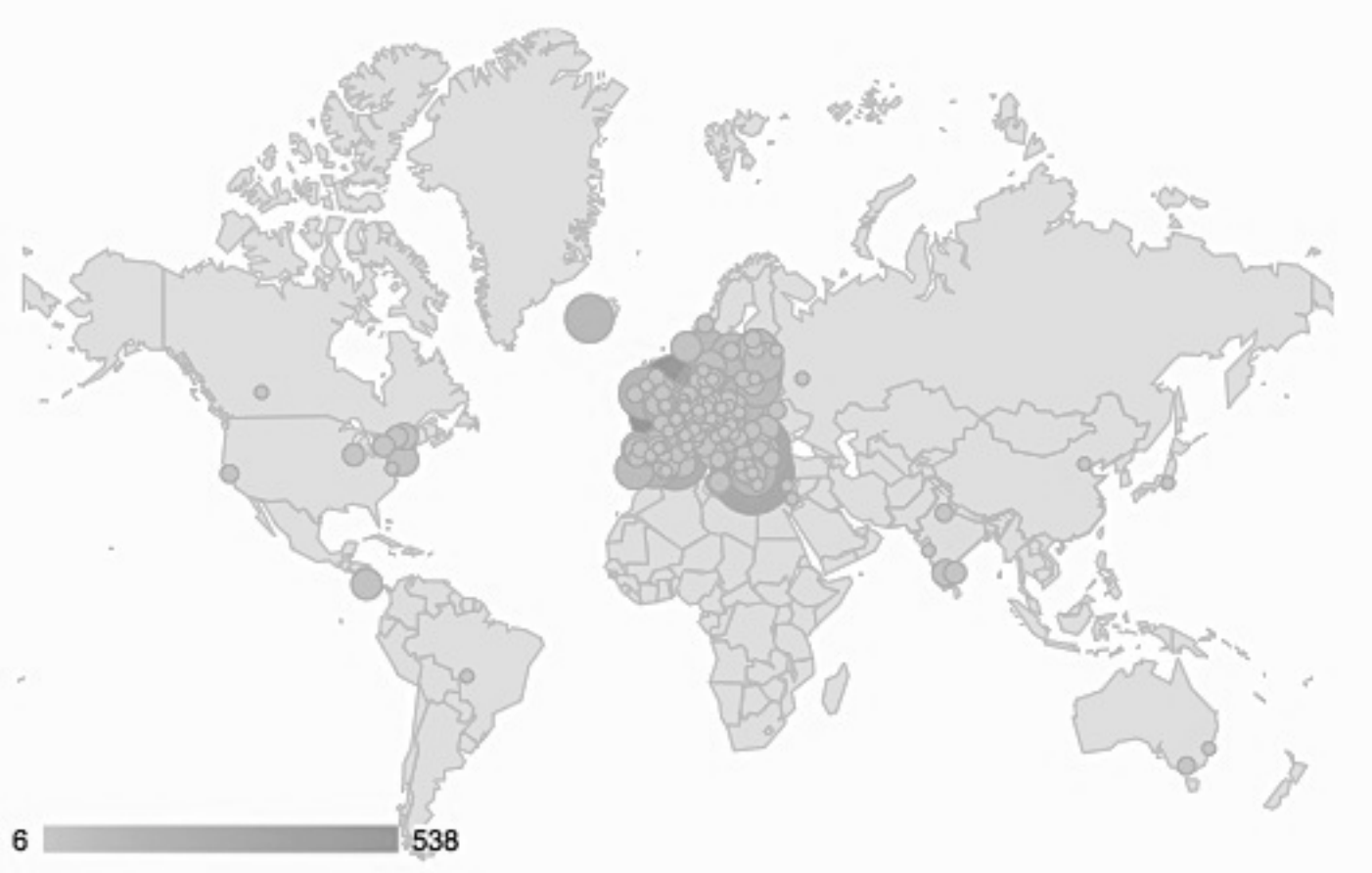
PAREDZAMI LIELI TĒRIŅI. Vācijas Mākslīgā intelekta pētījuma centra zinātniskais direktors Hanss Uškoreits norādīja, ka Eiropai, lai tā saglabātu savu vadošo lomu pasaulē, būs vajadzīgas visām Eiropas valodām pielāgotas valodu tehnoloģijas

Press Campaign: Highlights

Website: Visitors Overview



Website: Visitors' Cities



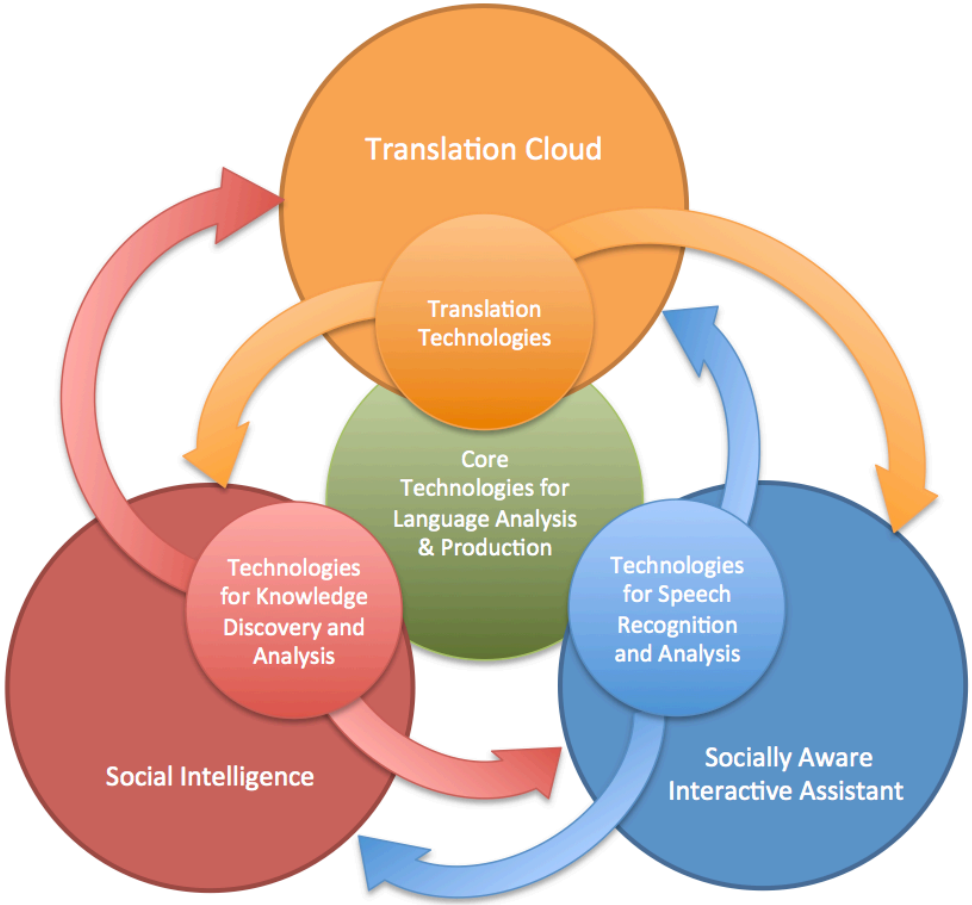
META-VISION

Strategic Research Agenda

Strategic Research Agenda

- *META-NET Strategic Research Agenda for Multilingual Europe 2020.*
- Addresses the problems we found when preparing the white papers.
- Three priority research themes and application/innovation scenarios.
- Can put Europe ahead of its competitors in this technology area.
- 180+ contributors.
- Final version to be ready in Nov. 2012.
- SRA will be presented to the EC and national bodies.

Strategic Research Agenda



META

Multilingual Europe Technology Alliance

META

- **META-NET** is a network of excellence.
- **META** is an open and growing strategic technology alliance: **Multilingual Europe Technology Alliance**.
  - 640+ members, including W3C, Google, Microsoft, GALA, research centres, LT companies etc.
  - Main goal: to support our Strategic Research Agenda.
  - **Join us!** **http://www.meta-net.eu/join**

META-NET

Recent News – Next Steps

Recent News – Next Steps

- **META-SHARE** Version 3.0 released in September 2012.
- Incoming resources from many EU-funded projects (40+ CAs), especially CESAR, METANET4U and META-NORD.
- META-SHARE launch event in January 2013.
- **META-TRUST AISBL** is an international non-profit organisation, founded on September 12, 2012: http://www.meta-trust.eu.
- Legal person of META-NET.
- **Strategic Research Agenda** press campaign, focus on social media (videos, infographics etc.).
- Meet with national research planners, funders, policy makers and inform them about the Strategic Research Agenda.
- **Data Liberation campaign.**

Q/A

Upcoming opportunities can provide sufficient resources to make our visions for Europe's citizens and economy, as described in the SRA, a reality (Horizon 2020; Connecting Europe Facility, CEF).

Thank you very much!

office@meta-net.eu

http://www.meta-net.eu
http://www.facebook.com/META.Alliance