



The language of Europe is translation.

« La langue de l'Europe, c'est la traduction. »

Assises de la traduction littéraire à Arles (France) le 14 novembre 1993,

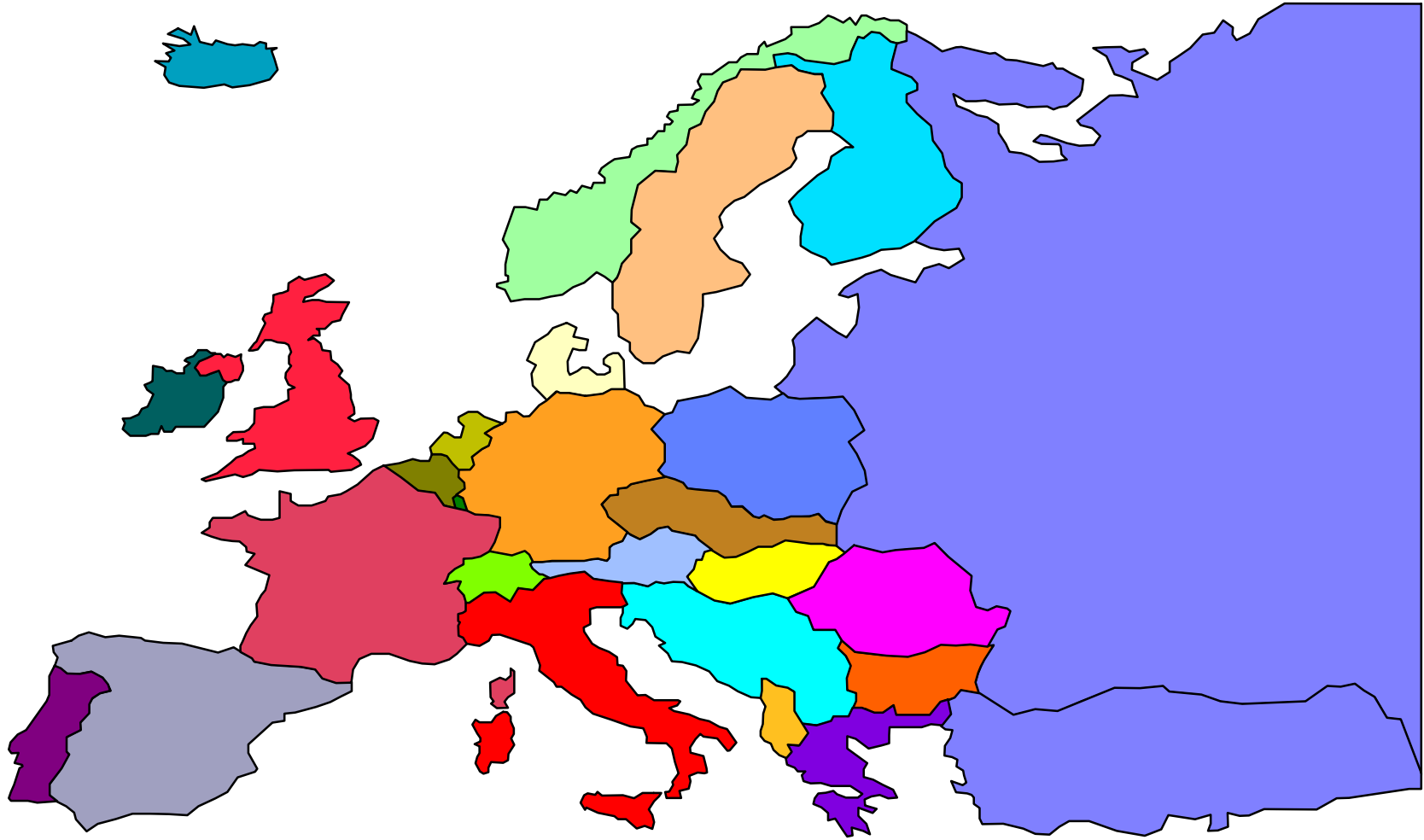
Human (Multilingual) Language Technologies for a Multilingual Europe

Khalid CHOUKRI

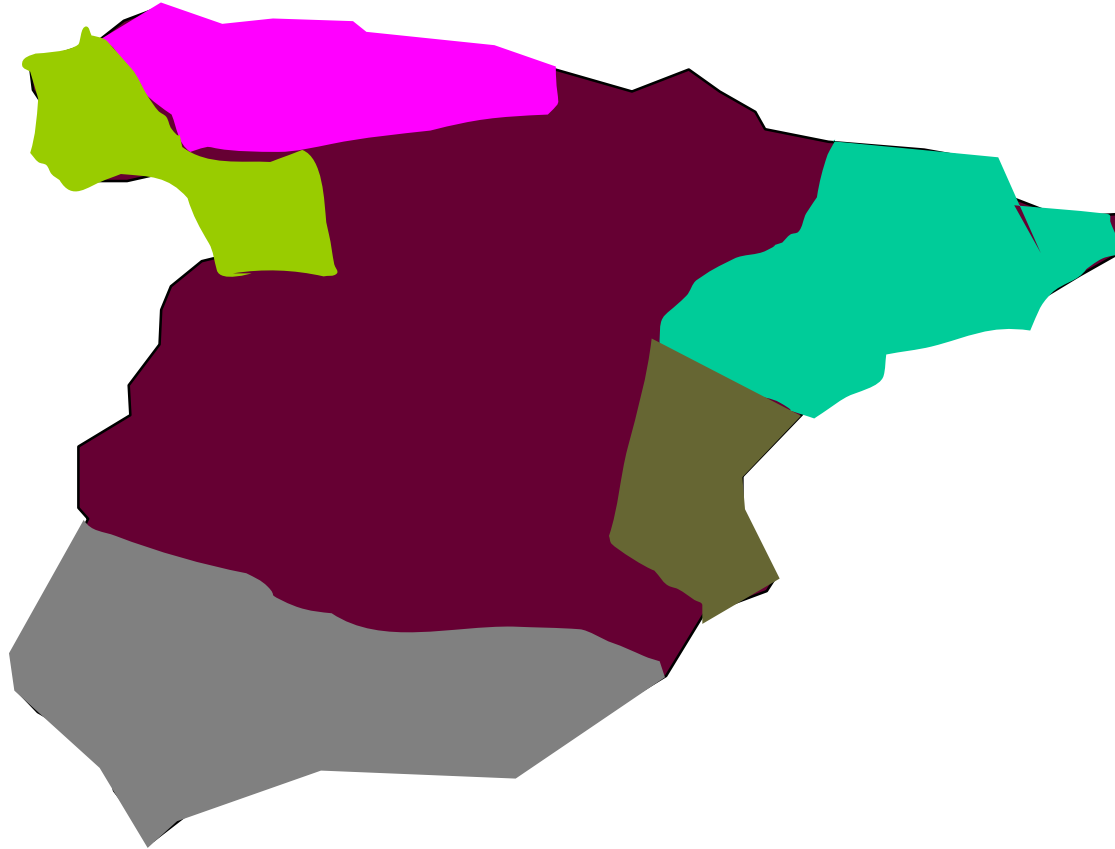
ELRA/ELDA

choukri@elda.org

Europe and the Multilingual Issue



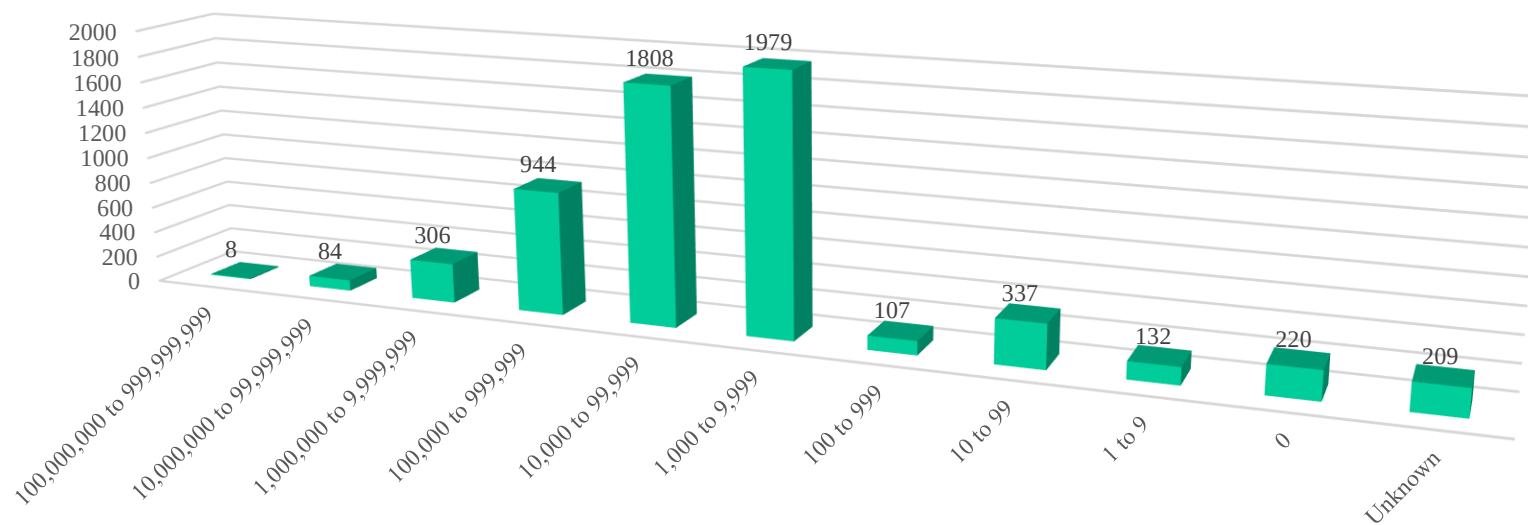
Even More Languages !

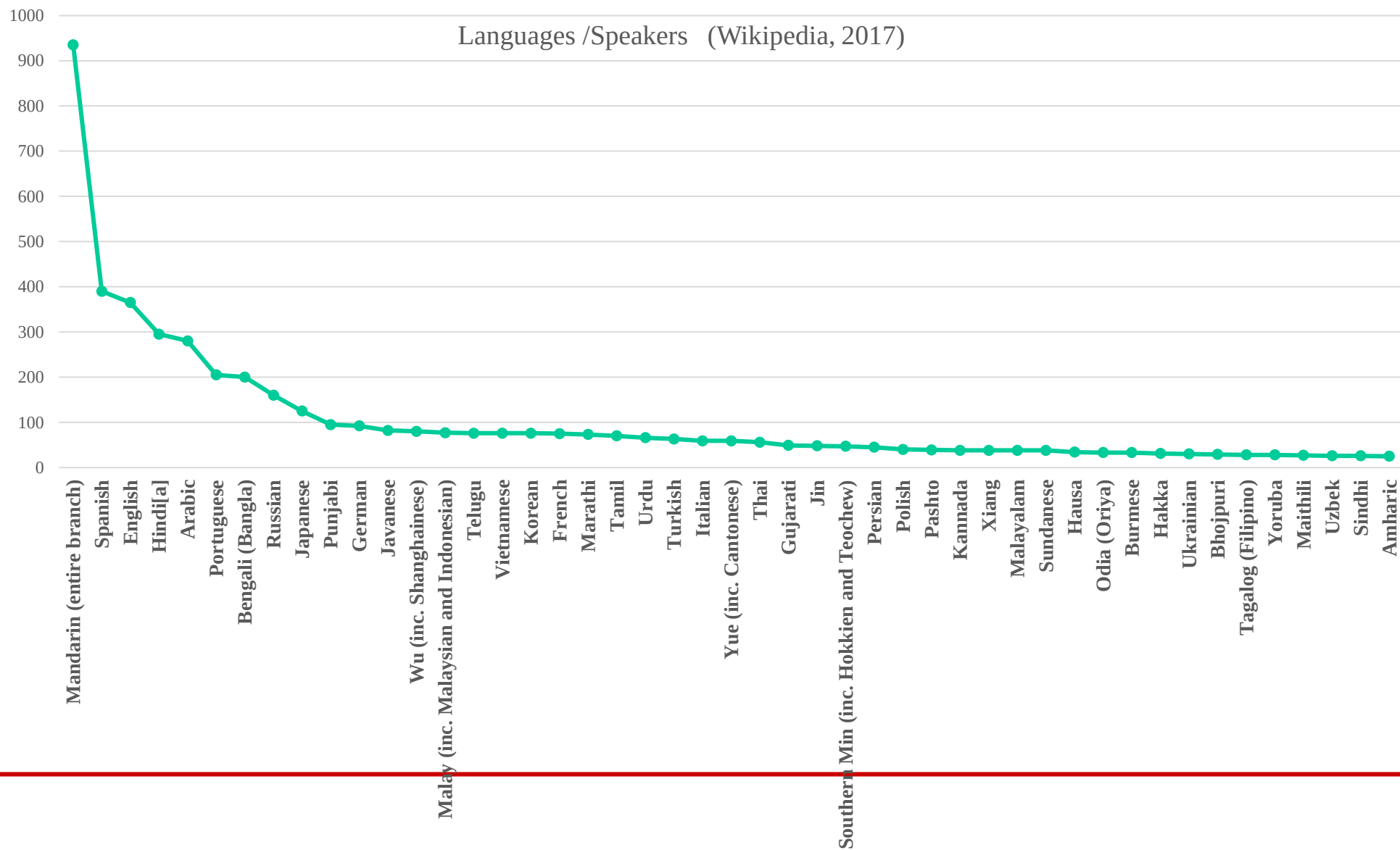


Languages: instruments of culture, identity and business

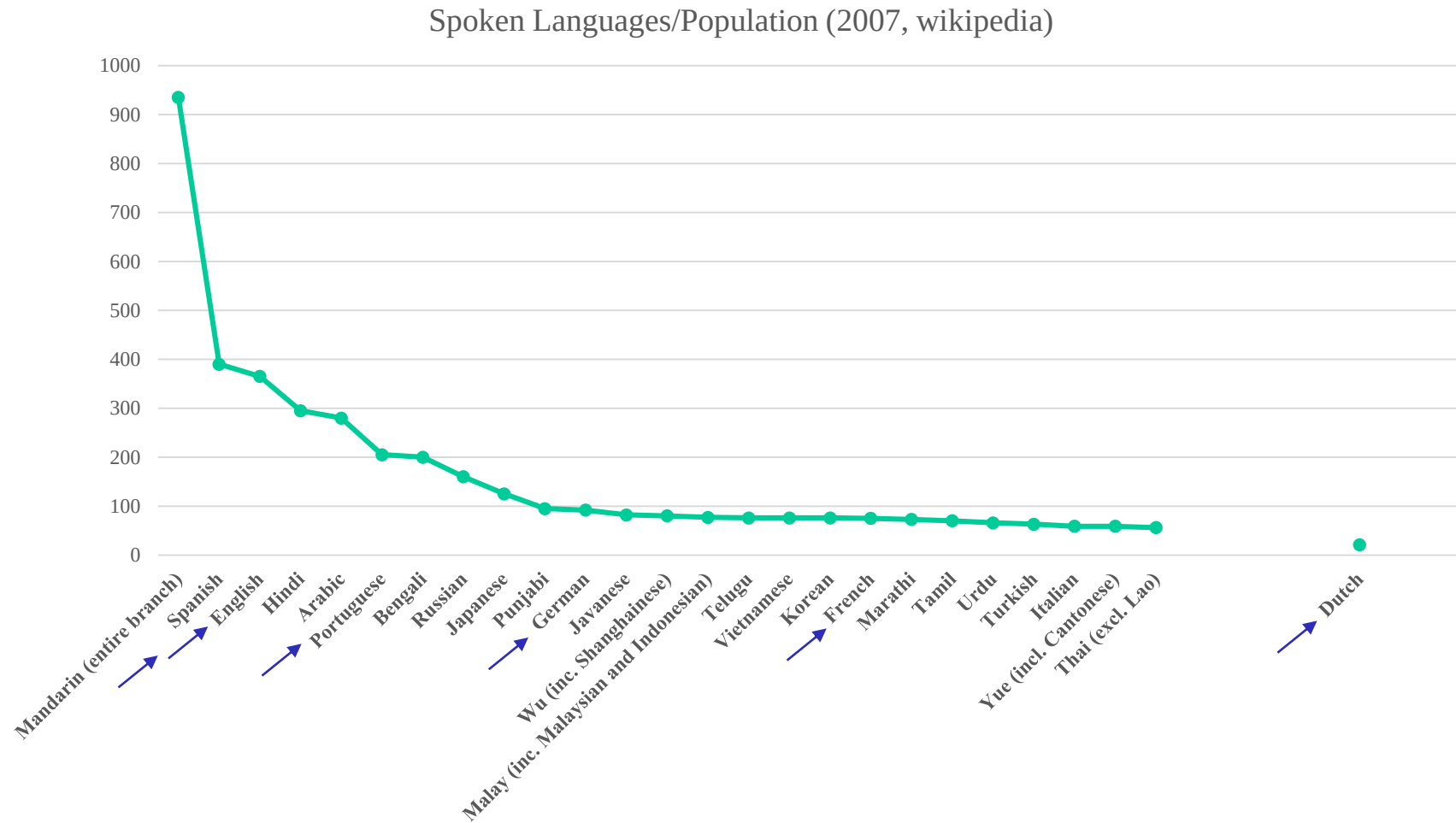
1. Over 7000 Languages
2. Languages have multiple modalities: **spoken**, written, signed
3. Only 200-300 have writing systems (about 50 different systems)

nb of languages vs nb of speakers

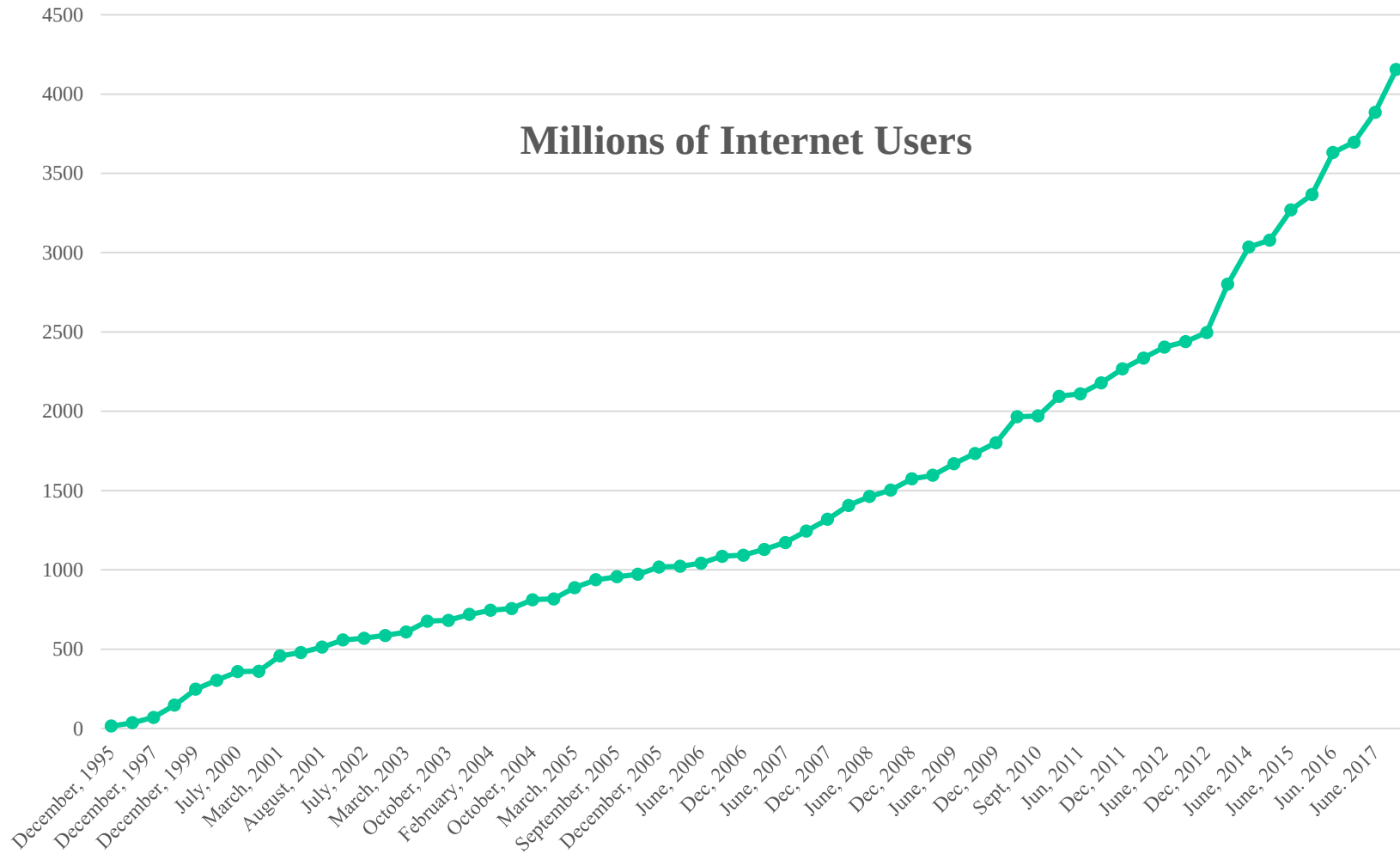




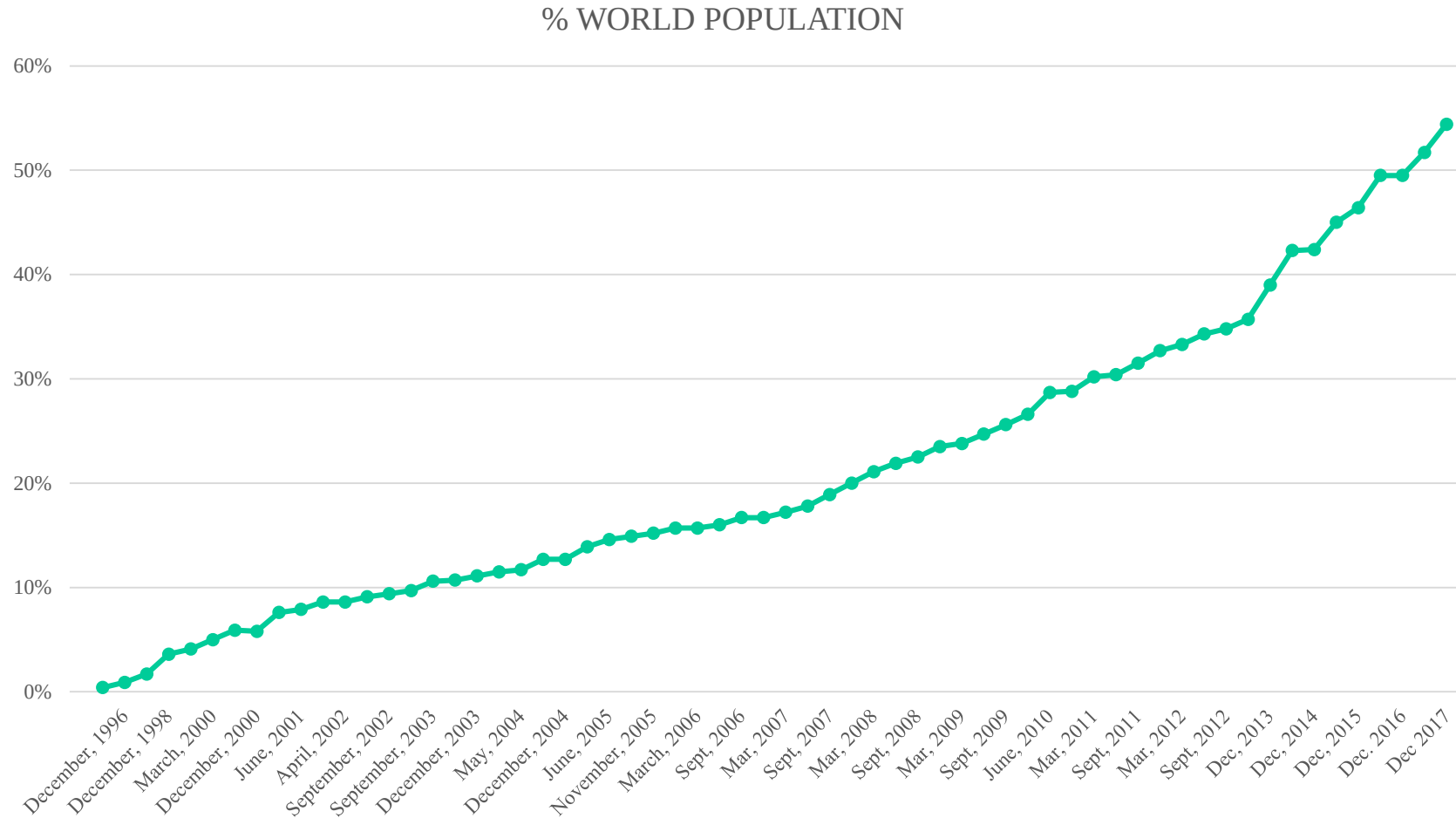
Languages: instruments of culture, identity and business



Internet Users over time

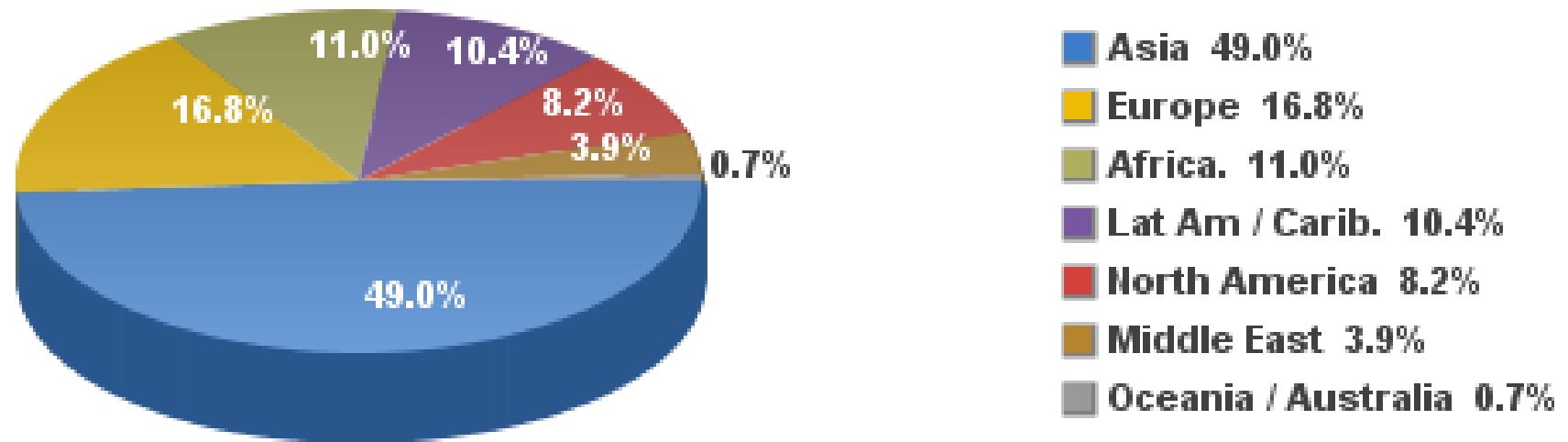


Ratio Users of Internet



Cyberspace Evolution (mostly web)

Internet Users in the World by Regions - June 30, 2018

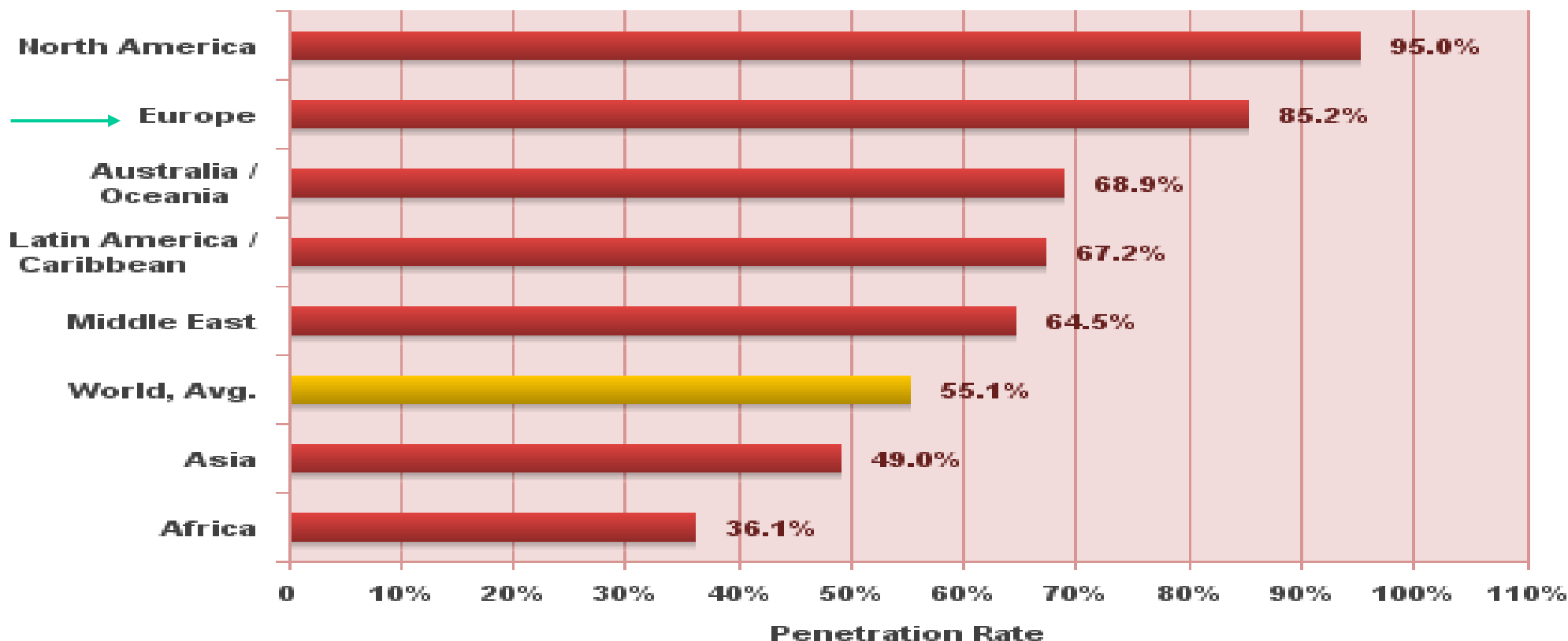


Source: Internet World Stats - www.internetworldstats.com/stats.htm

Basis: 4,208,571,287 Internet users in June 30, 2018

Copyright © 2018, Miniwatts Marketing Group

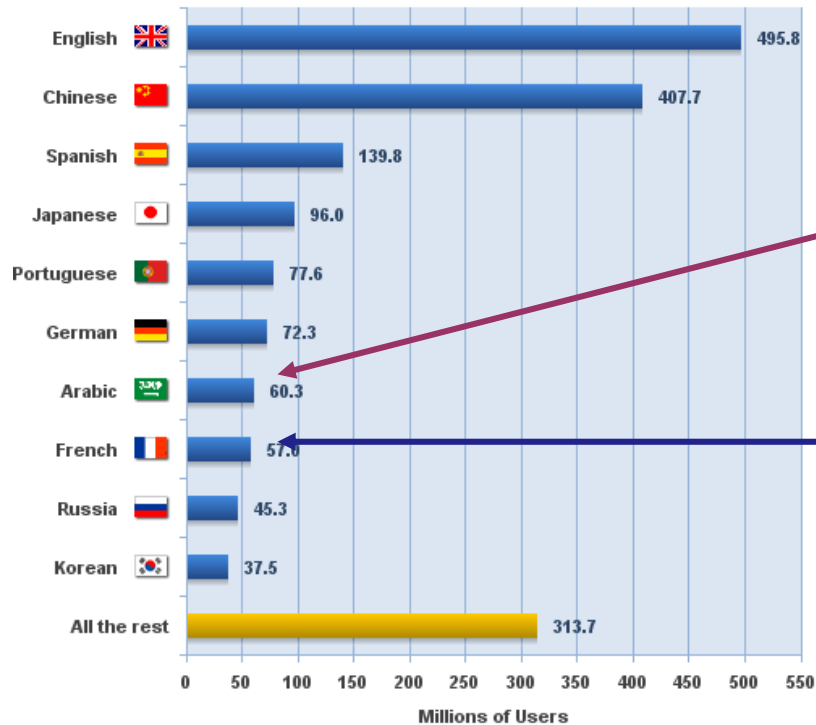
Internet World Penetration Rates by Geographic Regions - June 30, 2018



Source: Internet World Stats - www.internetworldstats.com/stats.htm
Penetration Rates are based on a world population of 7,634,758,428
and 4,208,571,287 estimated Internet users in June 30, 2018.
Copyright © 2018, Miniwatts Marketing Group

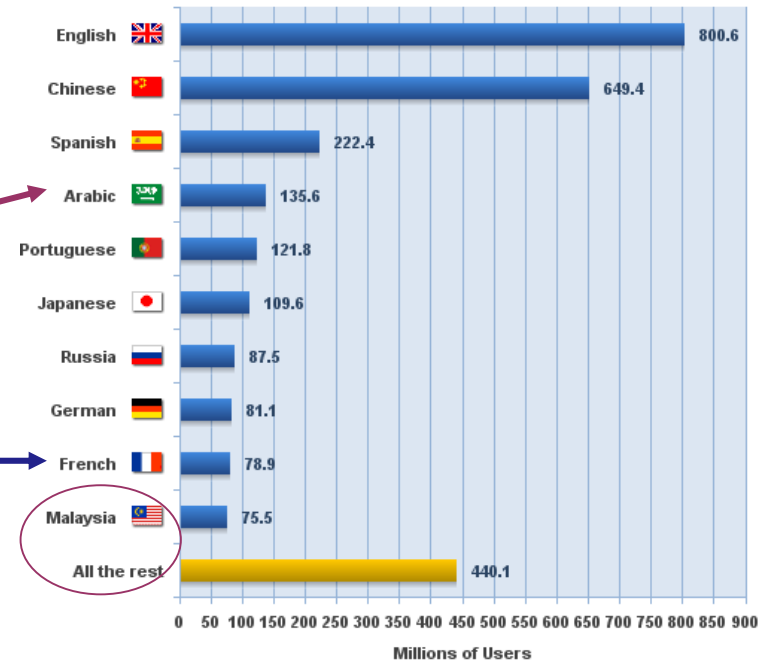
Language Cyberspace Evolution (mostly web)

**Top 10 Languages in the Internet
2009 in millions of users**



Source: Internet World Stats - www.internetworldstats.com/stats7.htm
Estimated Internet users are 1,802,330,457 for December 31, 2009
Copyright © 2000 - 2010, Miniwatts Marketing Group

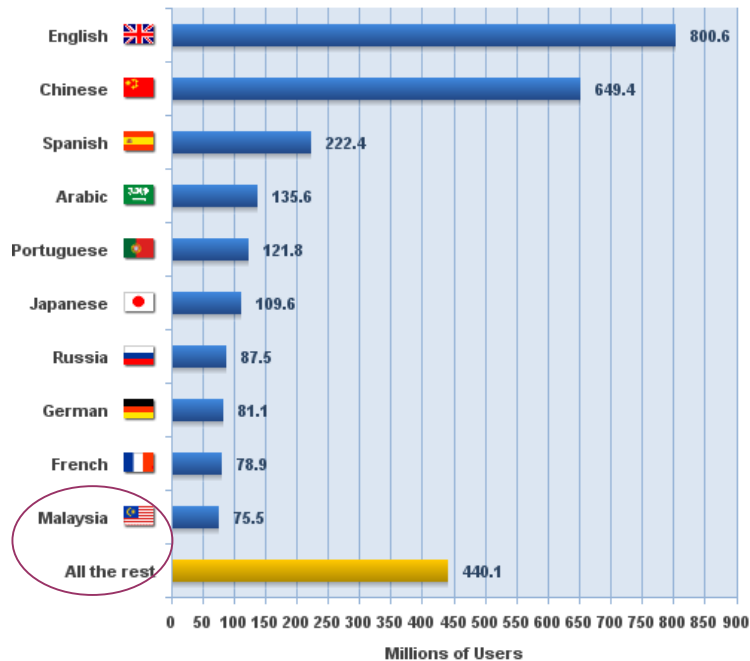
**Top Ten Languages in the Internet
2013 - in millions of users**



Source: Internet World Stats - www.internetworldstats.com/stats7.htm
Estimated Internet users are 2,802,478,934 on December 31, 2013
Copyright © 2014, Miniwatts Marketing Group

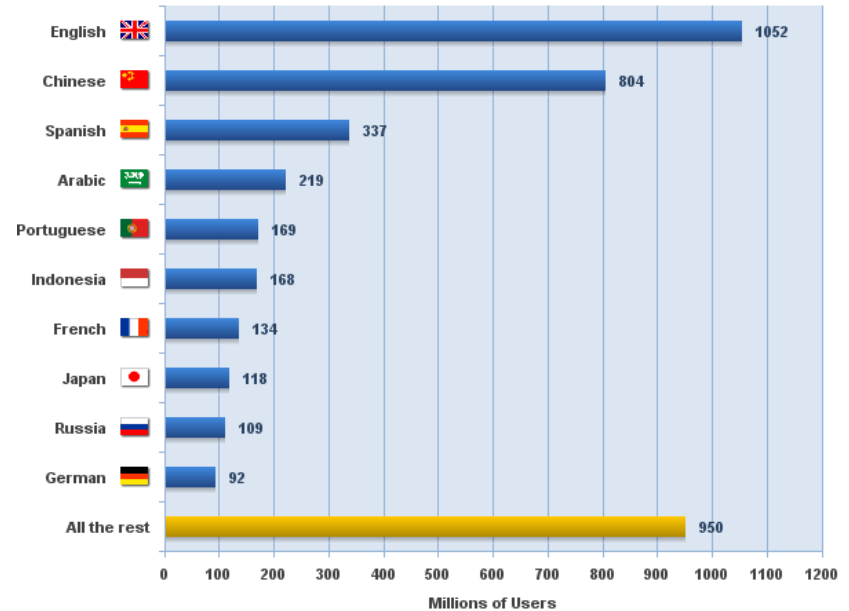
Cyberspace Evolution (mostly web)

Top Ten Languages in the Internet
2013 - in millions of users



Source: Internet World Stats - www.internetworldstats.com/stats7.htm
Estimated Internet users are 2,802,478,934 on December 31, 2013
Copyright © 2014, Miniwatts Marketing Group

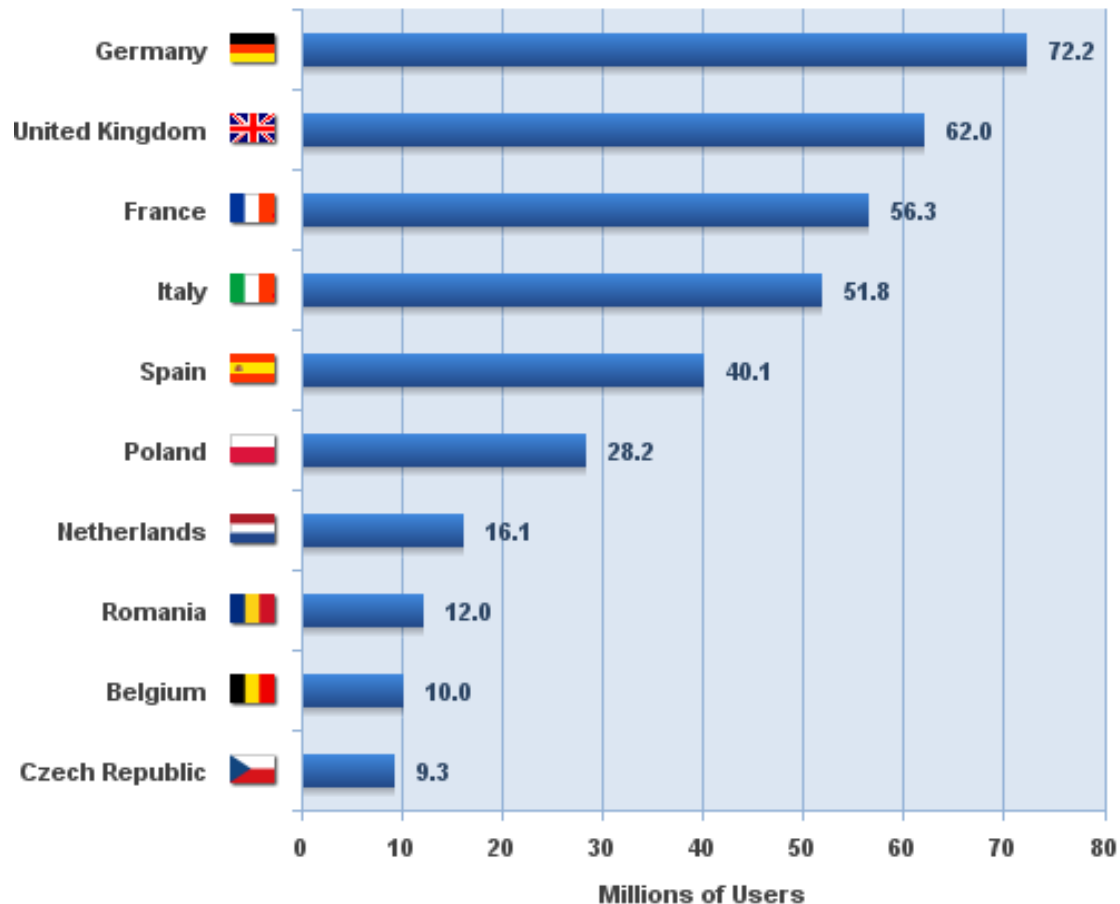
Top Ten Languages in the Internet
in Millions of users - December 2017



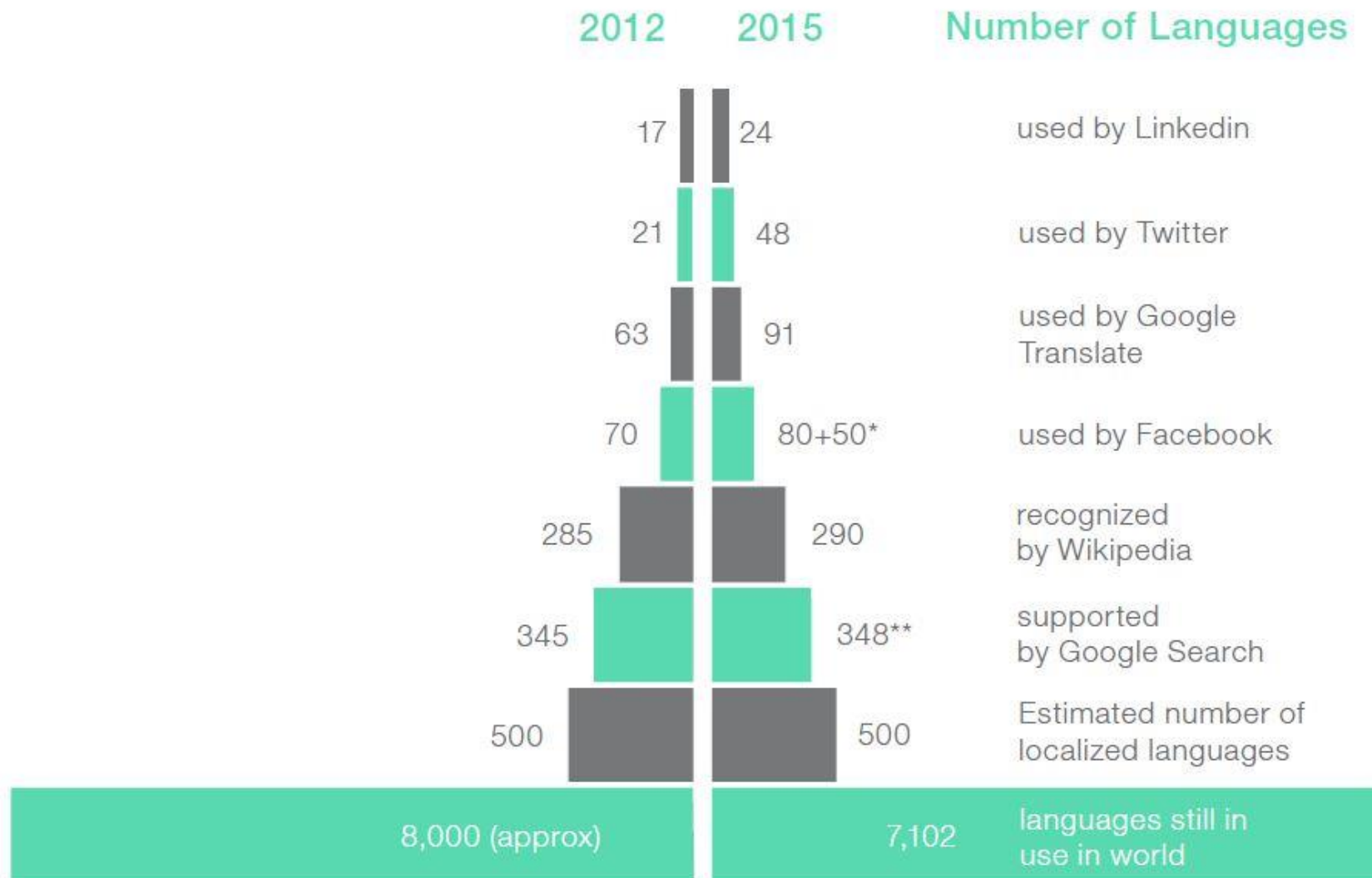
Source: Internet World Stats - www.internetworldstats.com/stats7.htm
Estimated total Internet users are 4,156,932,140 in December 31, 2017
Copyright © 2018, Miniwatts Marketing Group

At the EU Level

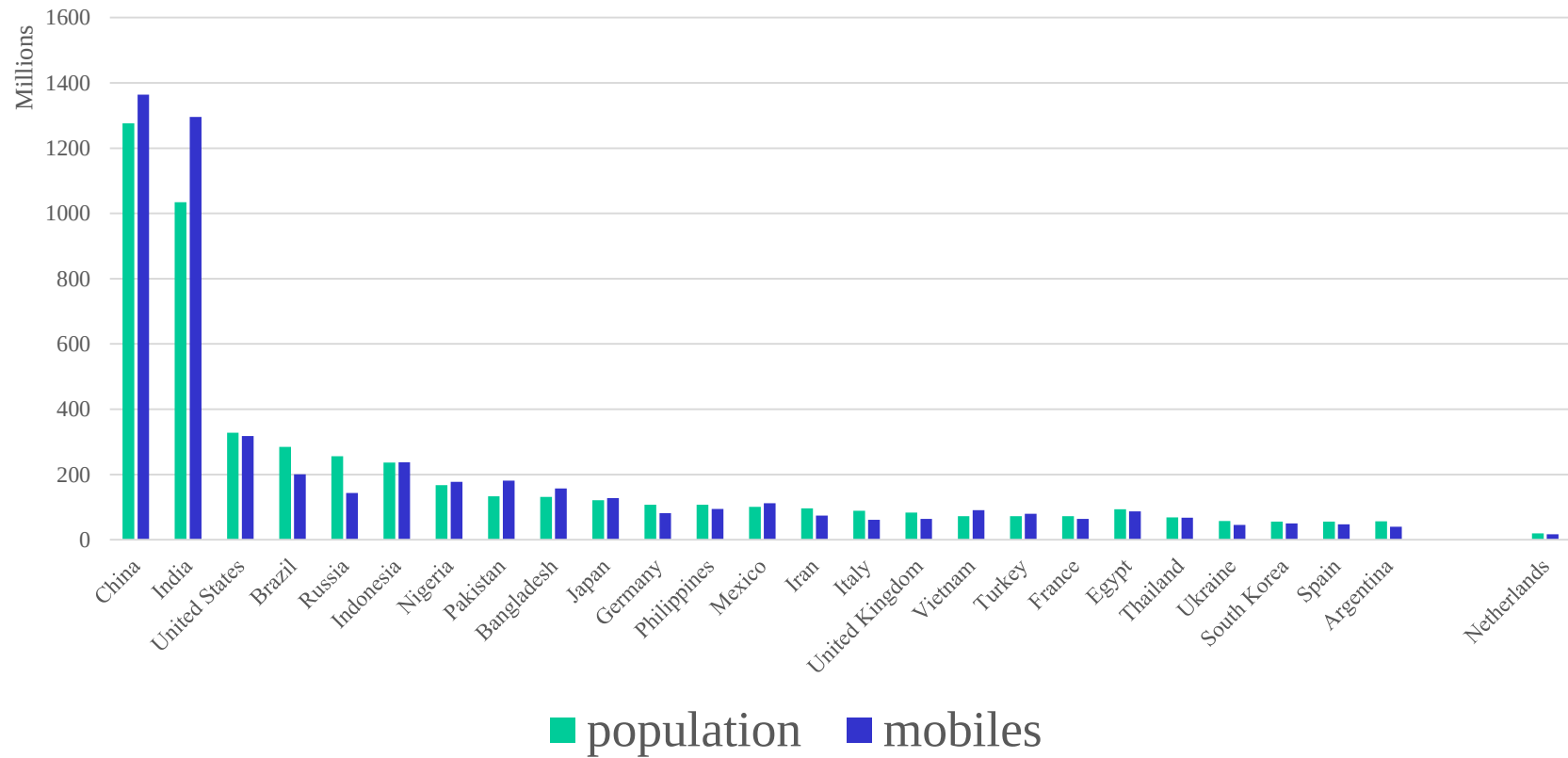
**European Union - EU28
Top 10 Internet Countries - June 2017**



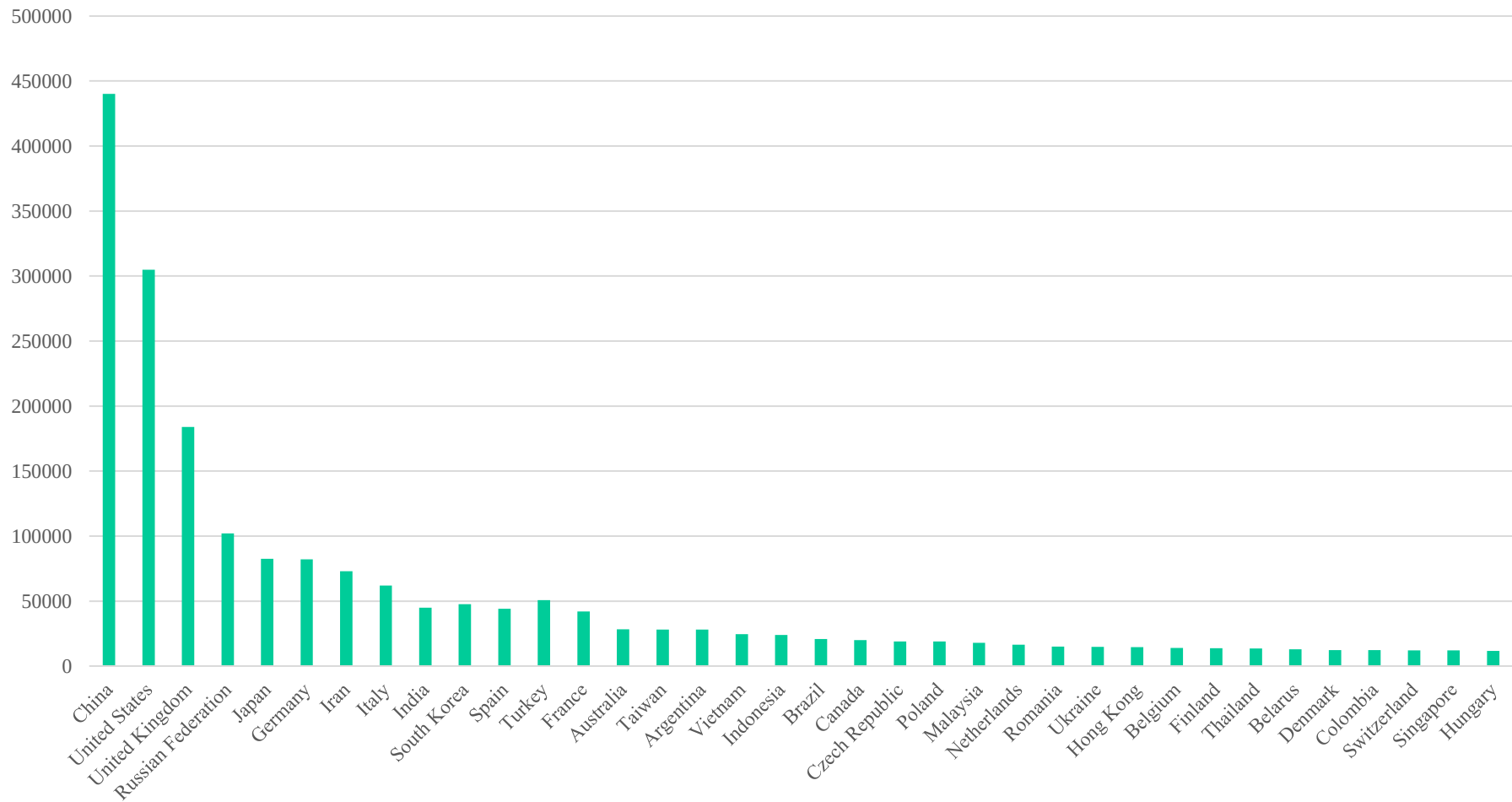
UNESCO Endangered Languages



Communication ... Population & Mobile phones



Books published /year



The Vision of a Digital Single Market



“Consumers need to be able to buy the best products at the best prices, wherever they are in Europe.”

Vice-President Ansip, Dec 2014

Accelerating growth through a connected Europe:

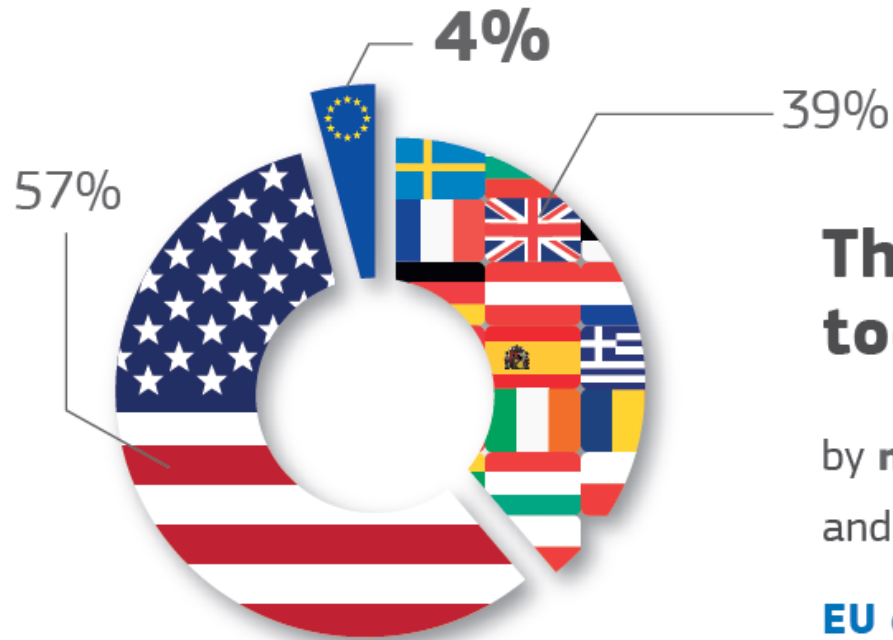
Speech at GSMA Mobile 360 conference in Brussels

http://europa.eu/rapid/press-release_SPEECH-14-2420_en.htm

Cortesy @JochenHummel

Infrastructures for a Multilingual Europe (Mli Project)

...and the Reality



The Digital Market today is made up

by **national** online services (39%)
and **US-based** online services (57%)

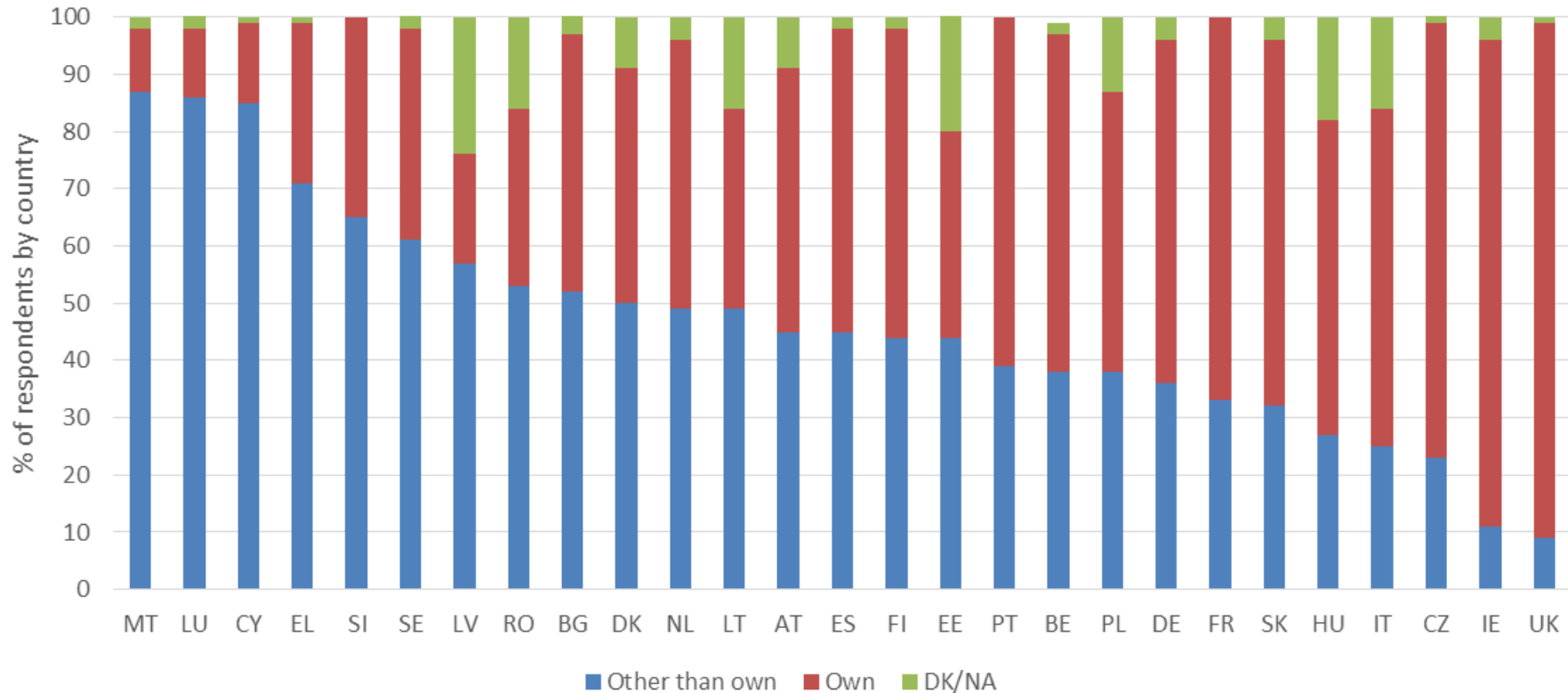
EU cross-border online services represent only 4%

http://europa.eu/rapid/attachment/IP-15-4653/en/Digital_Single_Market_Factsheet_20150325.pdf

Broken already by a Simple Search

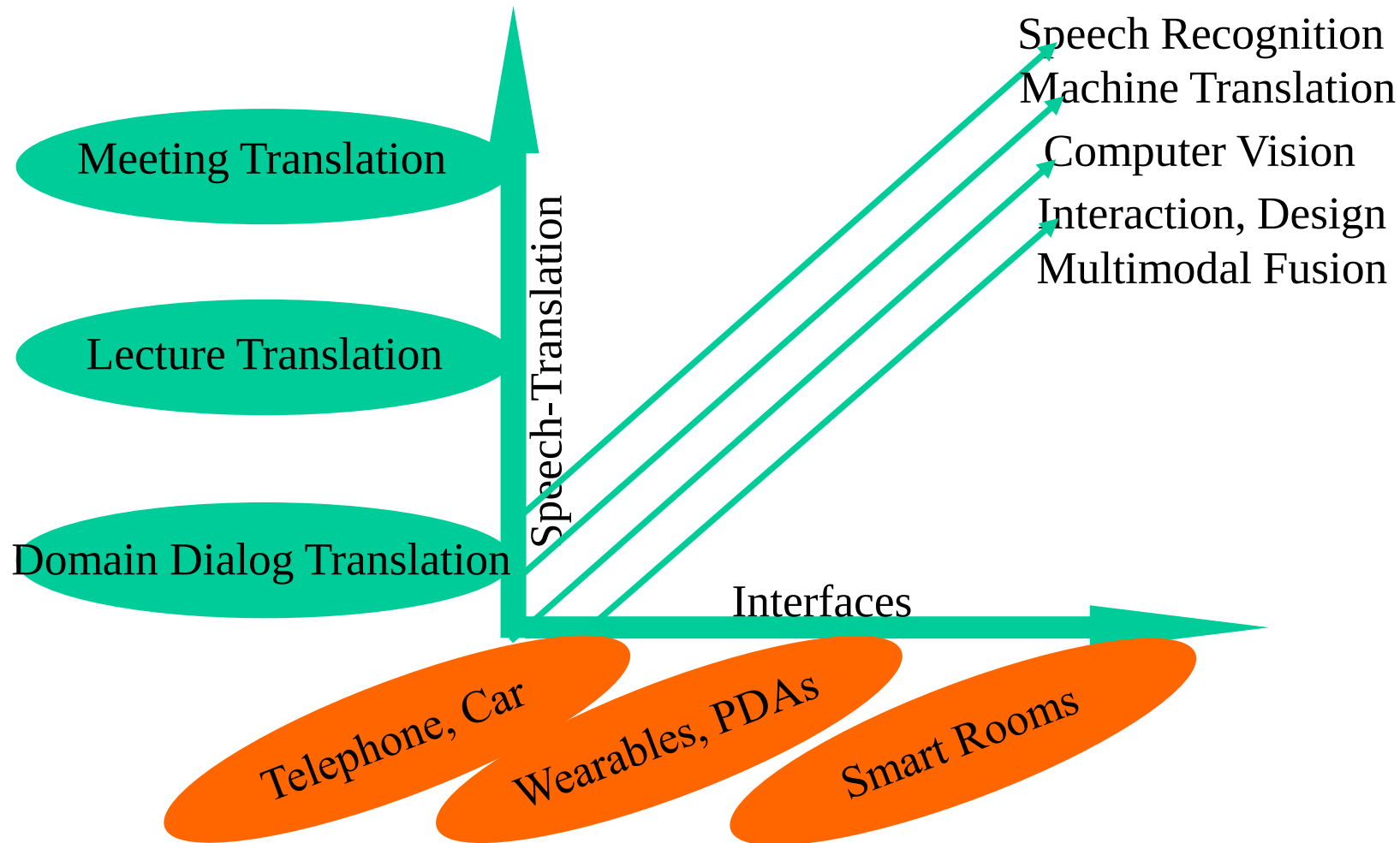


Figure 8. % of respondents by EU country using a language other than own language when writing on the Internet. Source: EC (2011: 10). Refers to writing e-mails, comments and messages posted on a website.



!42!

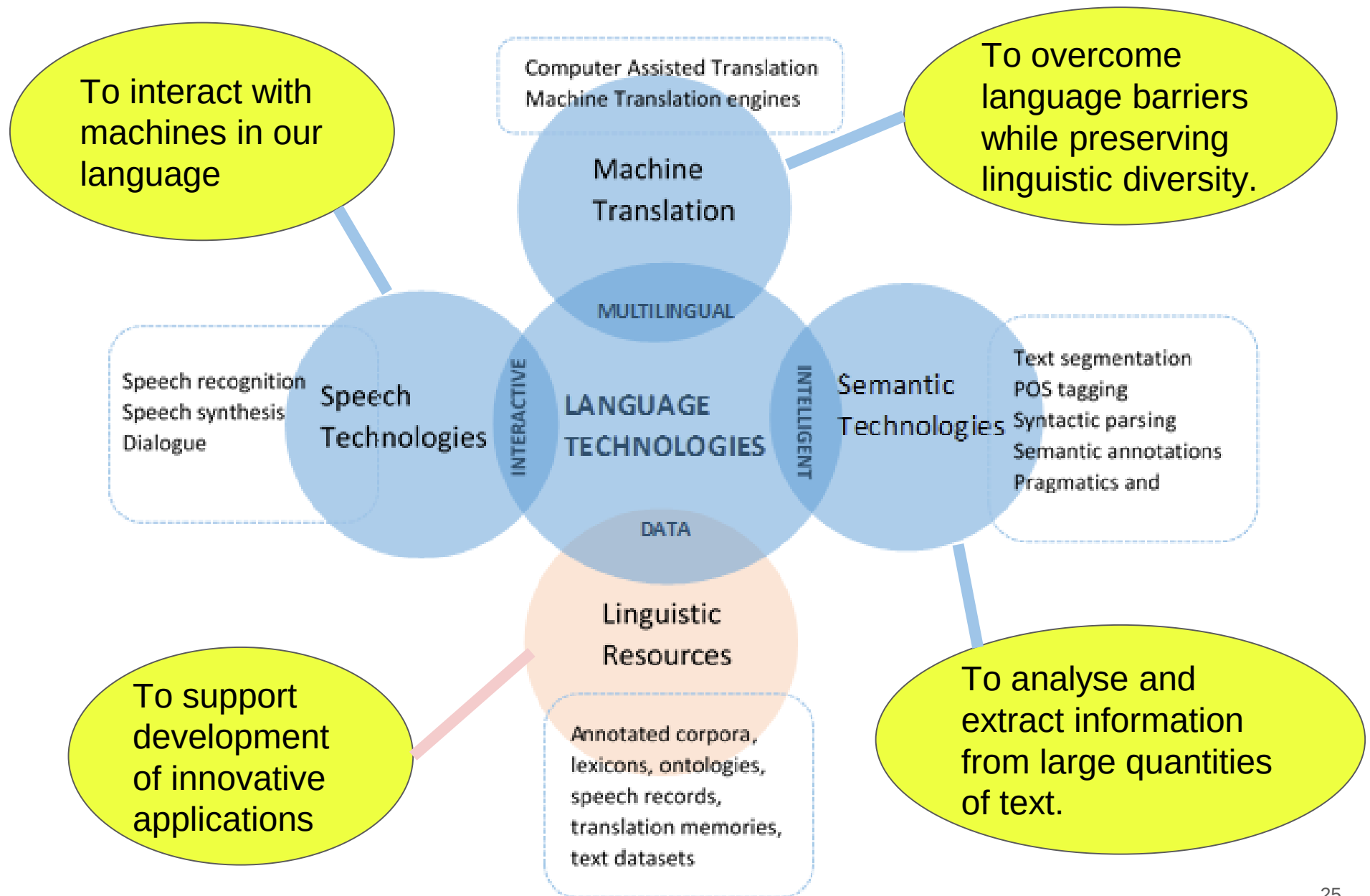
Grand Challenges // Problems



Dimensions of the Human Languages

- Speech
- Text inc. documents management (structure)
- Signs
- Hand writing
- Gestures ... pointing
- Images
- Biometrics
- Multimodal & Multimedia
-

Multilinguality



Human-Machine Interactions / Mediated by Computers

- **Person localization and tracking**
 - **Person identification: Face recognition, speaker identification (and fusion)**
 - **Gesture recognition**
 - **"attention" tracking**
 - **Conversational speech recognition & understanding**
 - **Acoustic scene analysis**
 - **Emotion identification (facial expression, emotional features ...)**
 - **Topic, emotion, sentiment, analytics,**
 - **Speech Recognition and Understanding for dictation**
 - **Speech Output (Synthesis & generation)**
 - **Document classification, Text categorization**
 - **... (Speech2Speech) Machine Translation**
-

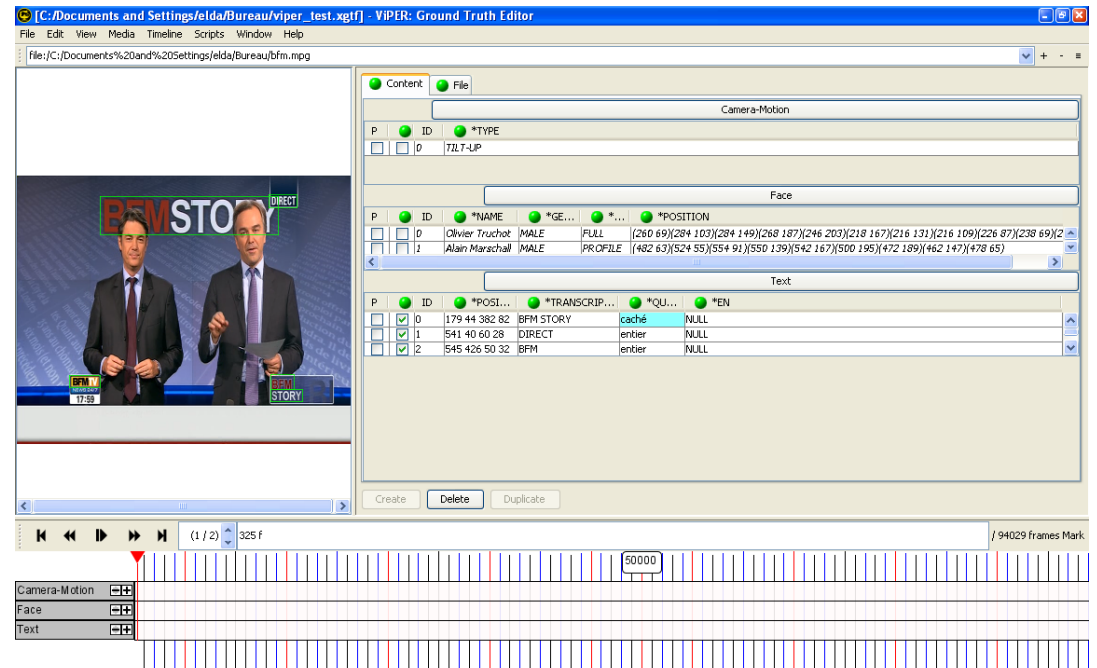
Automatic Transcription of Audio data

- Conversation (meetings)
- Broadcast News

[Demo courtesy of Vocapia Research / LIMSI-CNRS](#)

Multimodal technologies

- TV Broadcast (REPERE project - ViPER)
 - Head localization & identification
 - Embedded text localization & transcription
 - Speech transcription & annotation



Machine Translation

1. Introduction
 2. Languages: instruments of cultures, identity and business
 3. Language Technologies: some examples, illustrations
 4. **Special focus on Automated Translation**
 - Automated/Machine Translation , need for Language Resources (Data sets)
 - The MT@EC and the next generation (CEF-AT)
 - How can we help to improve it
 5. What are the trends ... Challenges for the next « decade! »
 6. Quick conclusions
 7. Q/A Session
-

2009







Internet Delivery

- Students bring their own Devices
- Transcription/Translation Output is Delivered via Web Page
- Interpretation Done on Server
- User Can Select Languages
- Launched 2012 as KIT Student Service
- Data Collected in Use and Evaluate



Prof. Alex Waibel



[DEMO](#)

Skype Translator



What is common to all these technologies

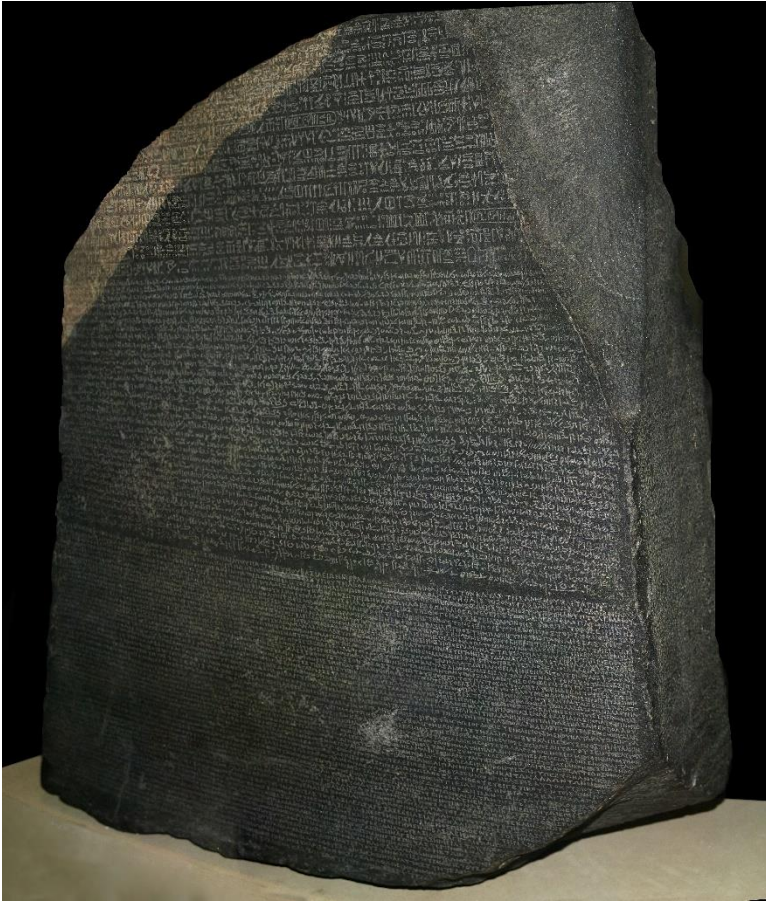
All are based on

MACHINE LEARNING FROM DATA
(The DATA driven Paradigm)

Some figures about the translation

- Youtube movies and Internet video growth
 - 13h of video every **MINUTE**
 - Human Transcriptions ... 3-50 times (1h audio = 3h to 50h of labor)
 - Translations & Interpretations
 - Over 400.000 translators (150.000 in Europe)
 - Need to translate 552 language pairs in EU, 110 in South Africa, 462 in India, (6000 languages all in all),
 - Not counting converting Sign language to/from Language A
 - Needs for translation grow by 30% **every year**
 - Consensus: 10% of data is translated
 - **Automation is essential**
-

Champollion & The Rosetta Stone



*Tableau des Signes Phonétiques
Des Écritures Hiéroglyphique et Démotique des anciens Égyptiens*

Lettres Grecques	Signes Démotiques	Signes Hiéroglyphiques
A	υ.υ.	𐀀 𐀁 𐀂 𐀃 𐀄 𐀅 𐀆 𐀇 𐀈 𐀉
B	4.2.	𐀊 𐀋 𐀌 𐀍 𐀎
Γ	κ.τ	𐀏 𐀐
Δ	κ.ς	𐀑 𐀒
E	ι.	𐀓 𐀔
Z		
H	III IIII. <III> IIII	𐀕 𐀖 𐀗 𐀘 𐀙 𐀚 𐀛 𐀜 𐀝 𐀞 𐀟 𐀠 𐀡 𐀢 𐀣 𐀤 𐀥 𐀦 𐀧 𐀨 𐀩 𐀪 𐀫 𐀬 𐀭 𐀮 𐀯 𐀰 𐀱 𐀲 𐀳 𐀴 𐀵 𐀶 𐀷 𐀸 𐀹 𐀺 𐀻 𐀼 𐀽 𐀾 𐀿 𐁀 𐁁 𐁂 𐁃 𐁄 𐁅 𐁆 𐁇 𐁈 𐁉 𐁊 𐁋 𐁌 𐁍 𐁎 𐁏 𐁐 𐁑 𐁒 𐁓 𐁔 𐁕 𐁖 𐁗 𐁘 𐁙 𐁚 𐁛 𐁜 𐁝 𐁞 𐁟 𐁠 𐁡 𐁢 𐁣 𐁤 𐁥 𐁦 𐁧 𐁨 𐁩 𐁪 𐁫 𐁬 𐁭 𐁮 𐁯 𐁰 𐁱 𐁲 𐁳 𐁴 𐁵 𐁶 𐁷 𐁸 𐁹 𐁺 𐁻 𐁼 𐁽 𐁾 𐁿 𐂀 𐂁 𐂂 𐂃 𐂄 𐂅 𐂆 𐂇 𐂈 𐂉 𐂊 𐂋 𐂌 𐂍 𐂎 𐂏 𐂐 𐂑 𐂒 𐂓 𐂔 𐂕 𐂖 𐂗 𐂘 𐂙 𐂚 𐂛 𐂜 𐂝 𐂞 𐂟 𐂠 𐂡 𐂢 𐂣 𐂤 𐂥 𐂦 𐂧 𐂨 𐂩 𐂪 𐂫 𐂬 𐂭 𐂮 𐂯 𐂰 𐂱 𐂲 𐂳 𐂴 𐂵 𐂶 𐂷 𐂸 𐂹 𐂺 𐂻 𐂼 𐂽 𐂾 𐂿 𐃀 𐃁 𐃂 𐃃 𐃄 𐃅 𐃆 𐃇 𐃈 𐃉 𐃊 𐃋 𐃌 𐃍 𐃎 𐃏 𐃐 𐃑 𐃒 𐃓 𐃔 𐃕 𐃖 𐃗 𐃘 𐃙 𐃚 𐃛 𐃜 𐃝 𐃞 𐃟 𐃠 𐃡 𐃢 𐃣 𐃤 𐃥 𐃦 𐃧 𐃨 𐃩 𐃪 𐃫 𐃬 𐃭 𐃮 𐃯 𐃰 𐃱 𐃲 𐃳 𐃴 𐃵 𐃶 𐃷 𐃸 𐃹 𐃺 𐃻 𐃼 𐃽 𐃾 𐃿 𐄀 𐄁 𐄂 𐄃 𐄄 𐄅 𐄆 𐄇 𐄈 𐄉 𐄊 𐄋 𐄌 𐄍 𐄎 𐄏 𐄐 𐄑 𐄒 𐄓 𐄔 𐄕 𐄖 𐄗 𐄘 𐄙 𐄚 𐄛 𐄜 𐄝 𐄞 𐄟 𐄠 𐄡 𐄢 𐄣 𐄤 𐄥 𐄦 𐄧 𐄨 𐄩 𐄪 𐄫 𐄬 𐄭 𐄮 𐄯 𐄰 𐄱 𐄲 𐄳 𐄴 𐄵 𐄶 𐄷 𐄸 𐄹 𐄺 𐄻 𐄼 𐄽 𐄾 𐄿 𐅀 𐅁 𐅂 𐅃 𐅄 𐅅 𐅆 𐅇 𐅈 𐅉 𐅊 𐅋 𐅌 𐅍 𐅎 𐅏 𐅐 𐅑 𐅒 𐅓 𐅔 𐅕 𐅖 𐅗 𐅘 𐅙 𐅚 𐅛 𐅜 𐅝 𐅞 𐅟 𐅠 𐅡 𐅢 𐅣 𐅤 𐅥 𐅦 𐅧 𐅨 𐅩 𐅪 𐅫 𐅬 𐅭 𐅮 𐅯 𐅰 𐅱 𐅲 𐅳 𐅴 𐅵 𐅶 𐅷 𐅸 𐅹 𐅺 𐅻 𐅼 𐅽 𐅾 𐅿 𐆀 𐆁 𐆂 𐆃 𐆄 𐆅 𐆆 𐆇 𐆈 𐆉 𐆊 𐆋 𐆌 𐆍 𐆎 𐆏 𐆐 𐆑 𐆒 𐆓 𐆔 𐆕 𐆖 𐆗 𐆘 𐆙 𐆚 𐆛 𐆜 𐆝 𐆞 𐆟 𐆠 𐆡 𐆢 𐆣 𐆤 𐆥 𐆦 𐆧 𐆨 𐆩 𐆪 𐆫 𐆬 𐆭 𐆮 𐆯 𐆰 𐆱 𐆲 𐆳 𐆴 𐆵 𐆶 𐆷 𐆸 𐆹 𐆺 𐆻 𐆼 𐆽 𐆾 𐆿 𐇀 𐇁 𐇂 𐇃 𐇄 𐇅 𐇆 𐇇 𐇈 𐇉 𐇊 𐇋 𐇌 𐇍 𐇎 𐇏 𐇐 𐇑 𐇒 𐇓 𐇔 𐇕 𐇖 𐇗 𐇘 𐇙 𐇚 𐇛 𐇜 𐇝 𐇞 𐇟 𐇠 𐇡 𐇢 𐇣 𐇤 𐇥 𐇦 𐇧 𐇨 𐇩 𐇪 𐇫 𐇬 𐇭 𐇮 𐇯 𐇰 𐇱 𐇲 𐇳 𐇴 𐇵 𐇶 𐇷 𐇸 𐇹 𐇺 𐇻 𐇼 𐇽 𐇾 𐇿 𐈀 𐈁 𐈂 𐈃 𐈄 𐈅 𐈆 𐈇 𐈈 𐈉 𐈊 𐈋 𐈌 𐈍 𐈎 𐈏 𐈐 𐈑 𐈒 𐈓 𐈔 𐈕 𐈖 𐈗 𐈘 𐈙 𐈚 𐈛 𐈜 𐈝 𐈞 𐈟 𐈠 𐈡 𐈢 𐈣 𐈤 𐈥 𐈦 𐈧 𐈨 𐈩 𐈪 𐈫 𐈬 𐈭 𐈮 𐈯 𐈰 𐈱 𐈲 𐈳 𐈴 𐈵 𐈶 𐈷 𐈸 𐈹 𐈺 𐈻 𐈼 𐈽 𐈾 𐈿 𐉀 𐉁 𐉂 𐉃 𐉄 𐉅 𐉆 𐉇 𐉈 𐉉 𐉊 𐉋 𐉌 𐉍 𐉎 𐉏 𐉐 𐉑 𐉒 𐉓 𐉔 𐉕 𐉖 𐉗 𐉘 𐉙 𐉚 𐉛 𐉜 𐉝 𐉞 𐉟 𐉠 𐉡 𐉢 𐉣 𐉤 𐉥 𐉦 𐉧 𐉨 𐉩 𐉪 𐉫 𐉬 𐉭 𐉮 𐉯 𐉰 𐉱 𐉲 𐉳 𐉴 𐉵 𐉶 𐉷 𐉸 𐉹 𐉺 𐉻 𐉼 𐉽 𐉾 𐉿 𐊀 𐊁 𐊂 𐊃 𐊄 𐊅 𐊆 𐊇 𐊈 𐊉 𐊊 𐊋 𐊌 𐊍 𐊎 𐊏 𐊐 𐊑 𐊒 𐊓 𐊔 𐊕 𐊖 𐊗 𐊘 𐊙 𐊚 𐊛 𐊜 𐊝 𐊞 𐊟 𐊠 𐊡 𐊢 𐊣 𐊤 𐊥 𐊦 𐊧 𐊨 𐊩 𐊪 𐊫 𐊬 𐊭 𐊮 𐊯 𐊰 𐊱 𐊲 𐊳 𐊴 𐊵 𐊶 𐊷 𐊸 𐊹 𐊺 𐊻 𐊼 𐊽 𐊾 𐊿 𐋀 𐋁 𐋂 𐋃 𐋄 𐋅 𐋆 𐋇 𐋈 𐋉 𐋊 𐋋 𐋌 𐋍 𐋎 𐋏 𐋐 𐋑 𐋒 𐋓 𐋔 𐋕 𐋖 𐋗 𐋘 𐋙 𐋚 𐋛 𐋜 𐋝 𐋞 𐋟 𐋠 𐋡 𐋢 𐋣 𐋤 𐋥 𐋦 𐋧 𐋨 𐋩 𐋪 𐋫 𐋬 𐋭 𐋮 𐋯 𐋰 𐋱 𐋲 𐋳 𐋴 𐋵 𐋶 𐋷 𐋸 𐋹 𐋺 𐋻 𐋼 𐋽 𐋾 𐋿 𐌀 𐌁 𐌂 𐌃 𐌄 𐌅 𐌆 𐌇 𐌈 𐌉 𐌊 𐌋 𐌌 𐌍 𐌎 𐌏 𐌐 𐌑 𐌒 𐌓 𐌔 𐌕 𐌖 𐌗 𐌘 𐌙 𐌚 𐌛 𐌜 𐌝 𐌞 𐌟 𐌠 𐌡 𐌢 𐌣 𐌤 𐌥 𐌦 𐌧 𐌨 𐌩 𐌪 𐌫 𐌬 𐌭 𐌮 𐌯 𐌰 𐌱 𐌲 𐌳 𐌴 𐌵 𐌶 𐌷 𐌸 𐌹 𐌺 𐌻 𐌼 𐌽 𐌾 𐌿 𐍀 𐍁 𐍂 𐍃 𐍄 𐍅 𐍆 𐍇 𐍈 𐍉 𐍊 𐍋 𐍌 𐍍 𐍎 𐍏 𐍐 𐍑 𐍒 𐍓 𐍔 𐍕 𐍖 𐍗 𐍘 𐍙 𐍚 𐍛 𐍜 𐍝 𐍞 𐍟 𐍠 𐍡 𐍢 𐍣 𐍤 𐍥 𐍦 𐍧 𐍨 𐍩 𐍪 𐍫 𐍬 𐍭 𐍮 𐍯 𐍰 𐍱 𐍲 𐍳 𐍴 𐍵 𐍶 𐍷 𐍸 𐍹 𐍺 𐍻 𐍼 𐍽 𐍾 𐍿 𐎀 𐎁 𐎂 𐎃 𐎄 𐎅 𐎆 𐎇 𐎈 𐎉 𐎊 𐎋 𐎌 𐎍 𐎎 𐎏 𐎐 𐎑 𐎒 𐎓 𐎔 𐎕 𐎖 𐎗 𐎘 𐎙 𐎚 𐎛 𐎜 𐎝 𐎞 𐎟 𐎠 𐎡 𐎢 𐎣 𐎤 𐎥 𐎦 𐎧 𐎨 𐎩 𐎪 𐎫 𐎬 𐎭 𐎮 𐎯 𐎰 𐎱 𐎲 𐎳 𐎴 𐎵 𐎶 𐎷 𐎸 𐎹 𐎺 𐎻 𐎼 𐎽 𐎾 𐎿 𐏀 𐏁 𐏂 𐏃 𐏄 𐏅 𐏆 𐏇 𐏈 𐏉 𐏊 𐏋 𐏌 𐏍 𐏎 𐏏 𐏐 𐏑 𐏒 𐏓 𐏔 𐏕 𐏖 𐏗 𐏘 𐏙 𐏚 𐏛 𐏜 𐏝 𐏞 𐏟 𐏠 𐏡 𐏢 𐏣 𐏤 𐏥 𐏦 𐏧 𐏨 𐏩 𐏪 𐏫 𐏬 𐏭 𐏮 𐏯 𐏰 𐏱 𐏲 𐏳 𐏴 𐏵 𐏶 𐏷 𐏸 𐏹 𐏺 𐏻 𐏼 𐏽 𐏾 𐏿 𐐀 𐐁 𐐂 𐐃 𐐄 𐐅 𐐆 𐐇 𐐈 𐐉 𐐊 𐐋 𐐌 𐐍 𐐎 𐐏 𐐐 𐐑 𐐒 𐐓 𐐔 𐐕 𐐖 𐐗 𐐘 𐐙 𐐚 𐐛 𐐜 𐐝 𐐞 𐐟 𐐠 𐐡 𐐢 𐐣 𐐤 𐐥 𐐦 𐐧 𐐨 𐐩 𐐪 𐐫 𐐬 𐐭 𐐮 𐐯 𐐰 𐐱 𐐲 𐐳 𐐴 𐐵 𐐶 𐐷 𐐸 𐐹 𐐺 𐐻 𐐼 𐐽 𐐾 𐐿 𐑀 𐑁 𐑂 𐑃 𐑄 𐑅 𐑆 𐑇 𐑈 𐑉 𐑊 𐑋 𐑌 𐑍 𐑎 𐑏 𐑐 𐑑 𐑒 𐑓 𐑔 𐑕 𐑖 𐑗 𐑘 𐑙 𐑚 𐑛 𐑜 𐑝 𐑞 𐑟 𐑠 𐑡 𐑢 𐑣 𐑤 𐑥 𐑦 𐑧 𐑨 𐑩 𐑪 𐑫 𐑬 𐑭 𐑮 𐑯 𐑰 𐑱 𐑲 𐑳 𐑴 𐑵 𐑶 𐑷 𐑸 𐑹 𐑺 𐑻 𐑼 𐑽 𐑾 𐑿 𐒀 𐒁 𐒂 𐒃 𐒄 𐒅 𐒆 𐒇 𐒈 𐒉 𐒊 𐒋 𐒌 𐒍 𐒎 𐒏 𐒐 𐒑 𐒒 𐒓 𐒔 𐒕 𐒖 𐒗 𐒘 𐒙 𐒚 𐒛 𐒜 𐒝 𐒞 𐒟 𐒠 𐒡 𐒢 𐒣 𐒤 𐒥 𐒦 𐒧 𐒨 𐒩 𐒪 𐒫 𐒬 𐒭 𐒮 𐒯 𐒰 𐒱 𐒲 𐒳 𐒴 𐒵 𐒶 𐒷 𐒸 𐒹 𐒺 𐒻 𐒼 𐒽 𐒾 𐒿 𐓀 𐓁 𐓂 𐓃 𐓄 𐓅 𐓆 𐓇 𐓈 𐓉 𐓊 𐓋 𐓌 𐓍 𐓎 𐓏 𐓐 𐓑 𐓒 𐓓 𐓔 𐓕 𐓖 𐓗 𐓘 𐓙 𐓚 𐓛 𐓜 𐓝 𐓞 𐓟 𐓠 𐓡 𐓢 𐓣 𐓤 𐓥 𐓦 𐓧 𐓨 𐓩 𐓪 𐓫 𐓬 𐓭 𐓮 𐓯 𐓰 𐓱 𐓲 𐓳 𐓴 𐓵 𐓶 𐓷 𐓸 𐓹 𐓺 𐓻 𐓼 𐓽 𐓾 𐓿 𐔀 𐔁 𐔂 𐔃 𐔄 𐔅 𐔆 𐔇 𐔈 𐔉 𐔊 𐔋 𐔌 𐔍 𐔎 𐔏 𐔐 𐔑 𐔒 𐔓 𐔔 𐔕 𐔖 𐔗 𐔘 𐔙 𐔚 𐔛 𐔜 𐔝 𐔞 𐔟 𐔠 𐔡 𐔢 𐔣 𐔤 𐔥 𐔦 𐔧 𐔨 𐔩 𐔪 𐔫 𐔬 𐔭 𐔮 𐔯 𐔰 𐔱 𐔲 𐔳 𐔴 𐔵 𐔶 𐔷 𐔸 𐔹 𐔺 𐔻 𐔼 𐔽 𐔾 𐔿 𐕀 𐕁 𐕂 𐕃 𐕄 𐕅 𐕆 𐕇 𐕈 𐕉 𐕊 𐕋 𐕌 𐕍 𐕎 𐕏 𐕐 𐕑 𐕒 𐕓 𐕔 𐕕 𐕖 𐕗 𐕘 𐕙 𐕚 𐕛 𐕜 𐕝 𐕞 𐕟 𐕠 𐕡 𐕢 𐕣 𐕤 𐕥 𐕦 𐕧 𐕨 𐕩 𐕪 𐕫 𐕬 𐕭 𐕮 𐕯 𐕰 𐕱 𐕲 𐕳 𐕴 𐕵 𐕶 𐕷 𐕸 𐕹 𐕺 𐕻 𐕼 𐕽 𐕾 𐕿 𐖀 𐖁 𐖂 𐖃 𐖄 𐖅 𐖆 𐖇 𐖈 𐖉 𐖊 𐖋 𐖌 𐖍 𐖎 𐖏 𐖐 𐖑 𐖒 𐖓 𐖔 𐖕 𐖖 𐖗 𐖘 𐖙 𐖚 𐖛 𐖜 𐖝 𐖞 𐖟 𐖠 𐖡 𐖢 𐖣 𐖤 𐖥 𐖦 𐖧 𐖨 𐖩 𐖪 𐖫 𐖬 𐖭 𐖮 𐖯 𐖰 𐖱 𐖲 𐖳 𐖴 𐖵 𐖶 𐖷 𐖸 𐖹 𐖺 𐖻 𐖼 𐖽 𐖾 𐖿 𐗀 𐗁 𐗂 𐗃 𐗄 𐗅 𐗆 𐗇 𐗈 𐗉 𐗊 𐗋 𐗌 𐗍 𐗎 𐗏 𐗐 𐗑 𐗒 𐗓 𐗔 𐗕 𐗖 𐗗 𐗘 𐗙 𐗚 𐗛 𐗜 𐗝 𐗞 𐗟 𐗠 𐗡 𐗢 𐗣 𐗤 𐗥 𐗦 𐗧 𐗨 𐗩 𐗪 𐗫 𐗬 𐗭 𐗮 𐗯 𐗰 𐗱 𐗲 𐗳 𐗴 𐗵 𐗶 𐗷 𐗸 𐗹 𐗺 𐗻 𐗼 𐗽 𐗾 𐗿 𐘀 𐘁 𐘂 𐘃 𐘄 𐘅 𐘆 𐘇 𐘈 𐘉 𐘊 𐘋 𐘌 𐘍 𐘎 𐘏 𐘐 𐘑 𐘒 𐘓 𐘔 𐘕 𐘖 𐘗 𐘘 𐘙 𐘚 𐘛 𐘜 𐘝 𐘞 𐘟 𐘠 𐘡 𐘢 𐘣 𐘤 𐘥 𐘦 𐘧 𐘨 𐘩 𐘪 𐘫 𐘬 𐘭 𐘮 𐘯 𐘰 𐘱 𐘲 𐘳 𐘴 𐘵 𐘶 𐘷 𐘸 𐘹 𐘺 𐘻 𐘼 𐘽 𐘾 𐘿 𐙀 𐙁 𐙂 𐙃 𐙄 𐙅 𐙆 𐙇 𐙈 𐙉 𐙊 𐙋 𐙌 𐙍 𐙎 𐙏 𐙐 𐙑 𐙒 𐙓 𐙔 𐙕 𐙖 𐙗 𐙘 𐙙 𐙚 𐙛 𐙜 𐙝 𐙞 𐙟 𐙠 𐙡 𐙢 𐙣 𐙤 𐙥 𐙦 𐙧 𐙨 𐙩 𐙪 𐙫 𐙬 𐙭 𐙮 𐙯 𐙰 𐙱 𐙲 𐙳 𐙴 𐙵 𐙶 𐙷 𐙸 𐙹 𐙺 𐙻 𐙼 𐙽 𐙾 𐙿 𐚀 𐚁 𐚂 𐚃 𐚄 𐚅 𐚆 𐚇 𐚈 𐚉 𐚊 𐚋 𐚌 𐚍 𐚎 𐚏 𐚐 𐚑 𐚒 𐚓 𐚔 𐚕 𐚖 𐚗 𐚘 𐚙 𐚚 𐚛 𐚜 𐚝 𐚞 𐚟 𐚠 𐚡 𐚢 𐚣 𐚤 𐚥 𐚦 𐚧 𐚨 𐚩 𐚪 𐚫 𐚬 𐚭 𐚮 𐚯 𐚰 𐚱 𐚲 𐚳 𐚴 𐚵 𐚶 𐚷 𐚸 𐚹 𐚺 𐚻 𐚼 𐚽 𐚾 𐚿 𐛀 𐛁 𐛂 𐛃 𐛄 𐛅 𐛆 𐛇 𐛈 𐛉 𐛊 𐛋 𐛌 𐛍 𐛎 𐛏 𐛐 𐛑 𐛒 𐛓 𐛔 𐛕 𐛖 𐛗 𐛘 𐛙 𐛚 𐛛 𐛜 𐛝 𐛞 𐛟 𐛠 𐛡 𐛢 𐛣 𐛤 𐛥 𐛦 𐛧 𐛨 𐛩 𐛪 𐛫 𐛬 𐛭 𐛮 𐛯 𐛰 𐛱 𐛲 𐛳 𐛴 𐛵 𐛶 𐛷 𐛸 𐛹 𐛺 𐛻 𐛼 𐛽 𐛾 𐛿 𐜀 𐜁 𐜂 𐜃 𐜄 𐜅 𐜆 𐜇 𐜈 𐜉 𐜊 𐜋 𐜌 𐜍 𐜎 𐜏 𐜐 𐜑 𐜒 𐜓 𐜔 𐜕 𐜖 𐜗 𐜘 𐜙 𐜚 𐜛 𐜜 𐜝 𐜞 𐜟 𐜠 𐜡 𐜢 𐜣 𐜤 𐜥 𐜦 𐜧 𐜨 𐜩 𐜪 𐜫 𐜬 𐜭 𐜮 𐜯 𐜰 𐜱 𐜲 𐜳 𐜴 𐜵 𐜶 𐜷 𐜸 𐜹 𐜺 𐜻 𐜼 𐜽 𐜾 𐜿 𐝀 𐝁 𐝂 𐝃 𐝄 𐝅 𐝆 𐝇 𐝈 𐝉 𐝊 𐝋 𐝌 𐝍 𐝎 𐝏 𐝐 𐝑 𐝒 𐝓 𐝔 𐝕 𐝖 𐝗 𐝘 𐝙 𐝚 𐝛 𐝜 𐝝 𐝞 𐝟 𐝠 𐝡 𐝢 𐝣 𐝤 𐝥 𐝦 𐝧 𐝨 𐝩 𐝪 𐝫 𐝬 𐝭 𐝮 𐝯 𐝰 𐝱 𐝲 𐝳 𐝴 𐝵 𐝶 𐝷 𐝸 𐝹 𐝺 𐝻 𐝼 𐝽 𐝾 𐝿 𐞀 𐞁 𐞂 𐞃 𐞄 𐞅 𐞆 𐞇 𐞈 𐞉 𐞊 𐞋 𐞌 𐞍 𐞎 𐞏 𐞐 𐞑 𐞒 𐞓 𐞔 𐞕 𐞖 𐞗 𐞘 𐞙 𐞚 𐞛 𐞜 𐞝 𐞞 𐞟 𐞠 𐞡 𐞢 𐞣 𐞤 𐞥 𐞦 𐞧 𐞨 𐞩 𐞪 𐞫 𐞬 𐞭 𐞮 𐞯 𐞰 𐞱 𐞲 𐞳 𐞴 𐞵 𐞶 𐞷 𐞸 𐞹 𐞺 𐞻 𐞼 𐞽 𐞾 𐞿 𐟀 𐟁 𐟂 𐟃 𐟄 𐟅 𐟆 𐟇 𐟈 𐟉 𐟊 𐟋 𐟌 𐟍 𐟎 𐟏 𐟐 𐟑 𐟒 𐟓 𐟔 𐟕 𐟖 𐟗 𐟘 𐟙 𐟚 𐟛 𐟜 𐟝 𐟞 𐟟 𐟠 𐟡 𐟢 𐟣 𐟤 𐟥 𐟦 𐟧 𐟨 𐟩 𐟪 𐟫 𐟬 𐟭 𐟮 𐟯 𐟰 𐟱 𐟲 𐟳 𐟴 𐟵 𐟶 𐟷 𐟸 𐟹 𐟺 𐟻 𐟼 𐟽 𐟾 𐟿 𐠀 𐠁 𐠂 𐠃 𐠄 𐠅 𐠆 𐠇 𐠈 𐠉 𐠊 𐠋 𐠌 𐠍 𐠎 𐠏 𐠐 𐠑 𐠒 𐠓 𐠔 𐠕 𐠖 𐠗 𐠘 𐠙 𐠚 𐠛 𐠜 𐠝 𐠞 𐠟 𐠠 𐠡 𐠢 𐠣 𐠤 𐠥 𐠦 𐠧 𐠨 𐠩 𐠪 𐠫 𐠬 𐠭 𐠮 𐠯 𐠰 𐠱 𐠲 𐠳 𐠴 𐠵 𐠶 𐠷 𐠸 𐠹 𐠺 𐠻 𐠼 𐠽 𐠾 𐠿 𐡀 𐡁 𐡂 𐡃 𐡄 𐡅 𐡆 𐡇 𐡈 𐡉 𐡊 𐡋 𐡌 𐡍 𐡎 𐡏 𐡐 𐡑 𐡒 𐡓 𐡔 𐡕 𐡖 𐡗 𐡘 𐡙 𐡚 𐡛 𐡜 𐡝 𐡞 𐡟 𐡠 𐡡 𐡢 𐡣 𐡤 𐡥 𐡦 𐡧 𐡨 𐡩 𐡪 𐡫 𐡬 𐡭 𐡮 𐡯 𐡰 𐡱 𐡲 𐡳 𐡴 𐡵 𐡶 𐡷 𐡸 𐡹 𐡺 𐡻 𐡼 𐡽 𐡾 𐡿 𐢀 𐢁 𐢂 𐢃 𐢄 𐢅 𐢆 𐢇 𐢈 𐢉 𐢊 𐢋 𐢌 𐢍 𐢎 𐢏 𐢐 𐢑 𐢒 𐢓 𐢔 𐢕 𐢖 𐢗 𐢘 𐢙 𐢚 𐢛 𐢜 𐢝 𐢞 𐢟 𐢠 𐢡 𐢢 𐢣 𐢤 𐢥 𐢦 𐢧 𐢨 𐢩 𐢪 𐢫 𐢬 𐢭 𐢮 𐢯 𐢰 𐢱 𐢲 𐢳 𐢴 𐢵 𐢶 𐢷 𐢸 𐢹 𐢺 𐢻 𐢼 𐢽 𐢾 𐢿 𐣀 𐣁 𐣂 𐣃 𐣄 𐣅 𐣆 𐣇 𐣈 𐣉 𐣊 𐣋 𐣌 𐣍 𐣎 𐣏 𐣐 𐣑 𐣒 𐣓 𐣔 𐣕 𐣖 𐣗 𐣘 𐣙 𐣚 𐣛 𐣜 𐣝 𐣞 𐣟 𐣠 𐣡 𐣢 𐣣 𐣤 𐣥 𐣦 𐣧 𐣨 𐣩 𐣪 𐣫 𐣬 𐣭 𐣮 𐣯 𐣰 𐣱 𐣲 𐣳 𐣴 𐣵 𐣶 𐣷 𐣸 𐣹 𐣺 𐣻 𐣼 𐣽 𐣾 𐣿 𐤀 𐤁 𐤂 𐤃 𐤄 𐤅 𐤆 𐤇 𐤈 𐤉 𐤊 𐤋 𐤌 𐤍 𐤎 𐤏 𐤐 𐤑 𐤒 𐤓 𐤔 𐤕 𐤖 𐤗 𐤘 𐤙 𐤚 𐤛 𐤜 𐤝 𐤞 𐤟 𐤠 𐤡 𐤢 𐤣 𐤤 𐤥 𐤦 𐤧 𐤨 𐤩 𐤪 𐤫 𐤬 𐤭 𐤮 𐤯 𐤰 𐤱 𐤲 𐤳 𐤴 𐤵 𐤶 𐤷 𐤸 𐤹 𐤺 𐤻 𐤼 𐤽 𐤾 𐤿 𐥀 𐥁 𐥂 𐥃 𐥄 𐥅 𐥆 𐥇 𐥈 𐥉 𐥊 𐥋 𐥌 𐥍 𐥎 𐥏 𐥐 𐥑 𐥒 𐥓 𐥔

Basis of MT



"Also knowing nothing official about, but having guessed and inferred considerable about, powerful new mechanized methods in cryptography - methods which I believe succeed even when one does not know what language has been coded - one naturally wonders if the problem of translation could conceivably be treated as a problem in cryptography. When I look at an article in Russian, I say "This is really written in English, but it has been coded in some strange symbols. I will now proceed to decode."

Warren Weaver (1894-1978)

Warren Weaver and Andrew Booth

"One naturally wonders if the problem of translation could conceivably be treated as a problem in cryptography. When I look at an article in Russian, I say: 'This is really written in English, but it has been coded in some strange symbols. I will now proceed to decode.'"

"Translation" (1955), in W.N. Locke and A.D. Booth (eds.),
Machine Translation of Languages (MIT Press, Cambridge, Mass.)."

Courtesy Marcelo Federico.

How does Machine Translation Work today?

Statistical MT learns from data

Two kinds of data:

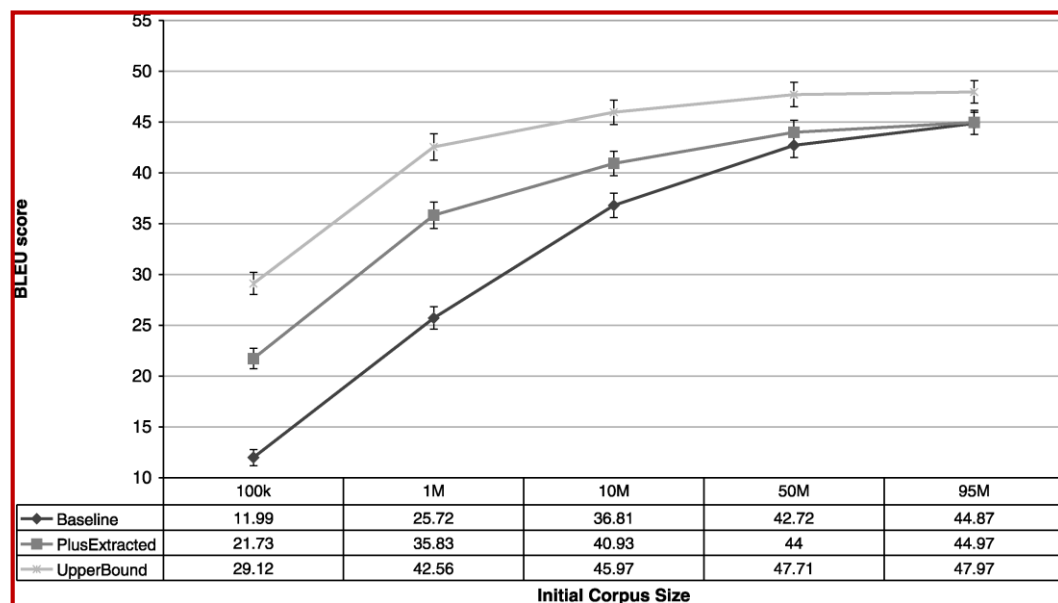
- Source documents and their human translations
- Target language collections
- The more data the better!
- Also: the right kind of data!

GERMAN	ENGLISH	FRENCH
Einleitung	Introduction	Introduction
<i>I. Von dem Unterschiede der reinen und empirischen Erkenntnis</i>	<i>I. Of the difference between Pure and Empirical Knowledge</i>	<i>I. De la différence de la connaissance pure et de la connaissance empirique.</i>
Daß alle unsere Erkenntnis mit der Erfahrung anfangt, daran ist gar kein Zweifel; denn wodurch sollte das Erkenntnisvermögen sonst zur Ausübung erweckt werden, geschähe es nicht durch Gegenstände, die unsere Sinne rühren und teils von selbst Vorstellungen bewirken, teils unsere Verstandstätigkeit in Bewegung bringen, diese zu vergleichen, sie zu verknüpfen oder zu trennen, und so den rohen Stoff sinnlicher Eindrücke zu einer Erkenntnis der Gegenstände zu verarbeiten, die Erfahrung heißt? Der Zeit nach geht also keine Erkenntnis in uns vor der Erfahrung vorher, und mit dieser fängt alle an.	That all our knowledge begins with experience there can be no doubt. For how is it possible that the faculty of cognition should be awakened into exercise otherwise than by means of objects which affect our senses, and partly of themselves produce representations, partly rouse our powers of understanding into activity, to compare to connect, or to separate these, and so to convert the raw material of our sensuous impressions into a knowledge of objects, which is called experience? In respect of time, therefore, no knowledge of ours is antecedent to experience, but begins with it.	Que toute notre connaissance commence avec l'expérience, cela ne soulève aucun doute. En effet, par quoi notre pouvoir de connaître pourrait-il être éveillé et mis en action, si ce n'est par des objets qui frappent nos sens et qui, d'une part, produisent par eux-mêmes des représentations et, d'autre part, mettent en mouvement notre faculté intellectuelle, afin qu'elle compare, lie ou sépare ces représentations, et travaille ainsi la matière brute des impressions sensibles pour en tirer une connaissance des objets, celle qu'on nomme l'expérience? Ainsi, chronologiquement, aucune connaissance ne précède en nous l'expérience et c'est avec elle que toutes commencent.

Importance of data and Re-usability

- ✓ **Almost all technologies are data driven and based on statistical paradigms ...
(modeling based on huge amounts of data)**

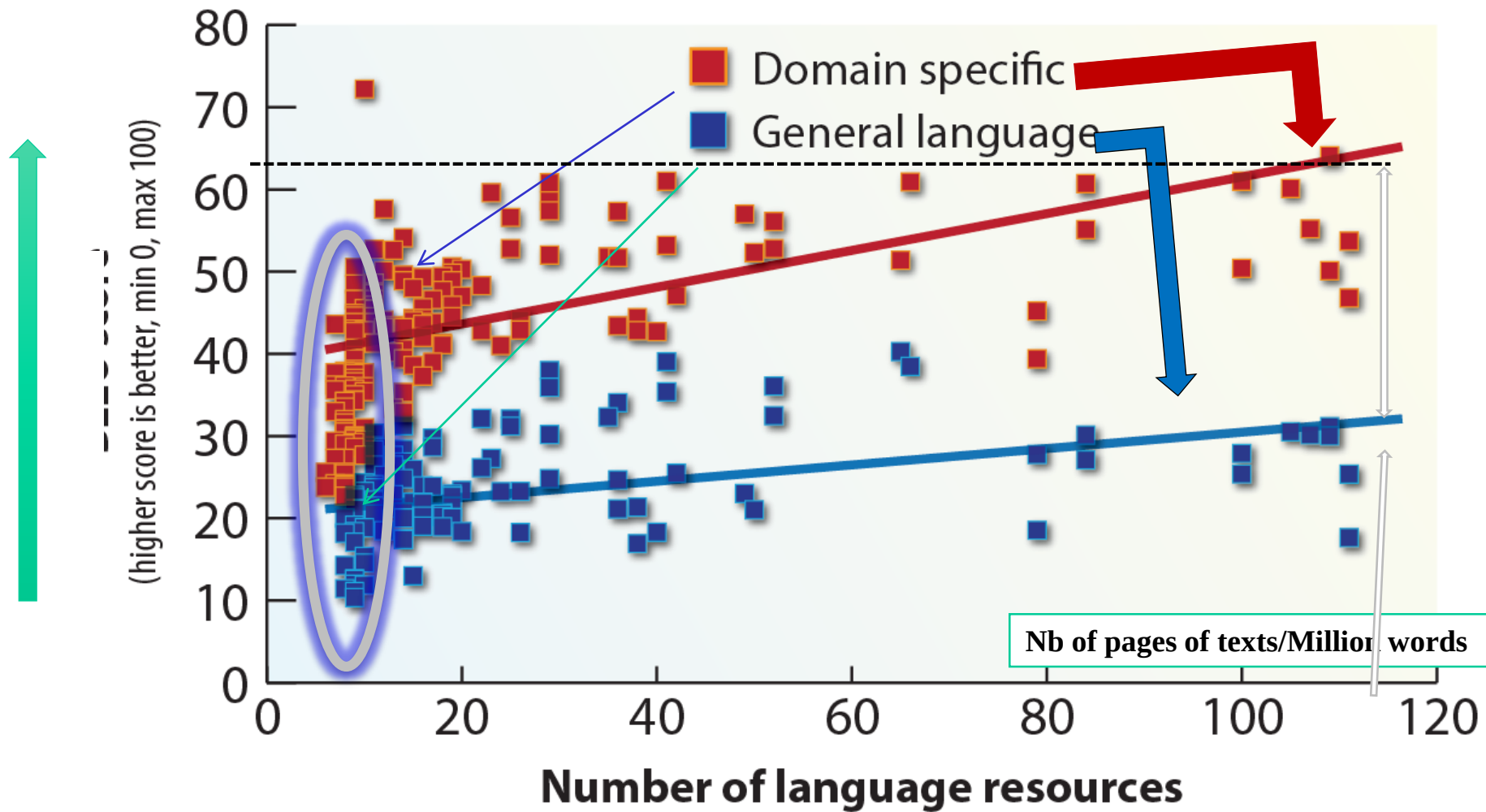
Let us look at MT performance when "simply" adding data



MT performance improvements for Arabic-English
(Courtesy Dragos Stefan Munteanu and Daniel Marcu)

Impact of number of language resources on Statistical MT quality

A Commonly used measure
(quality assessment)



The role of Machine Translation

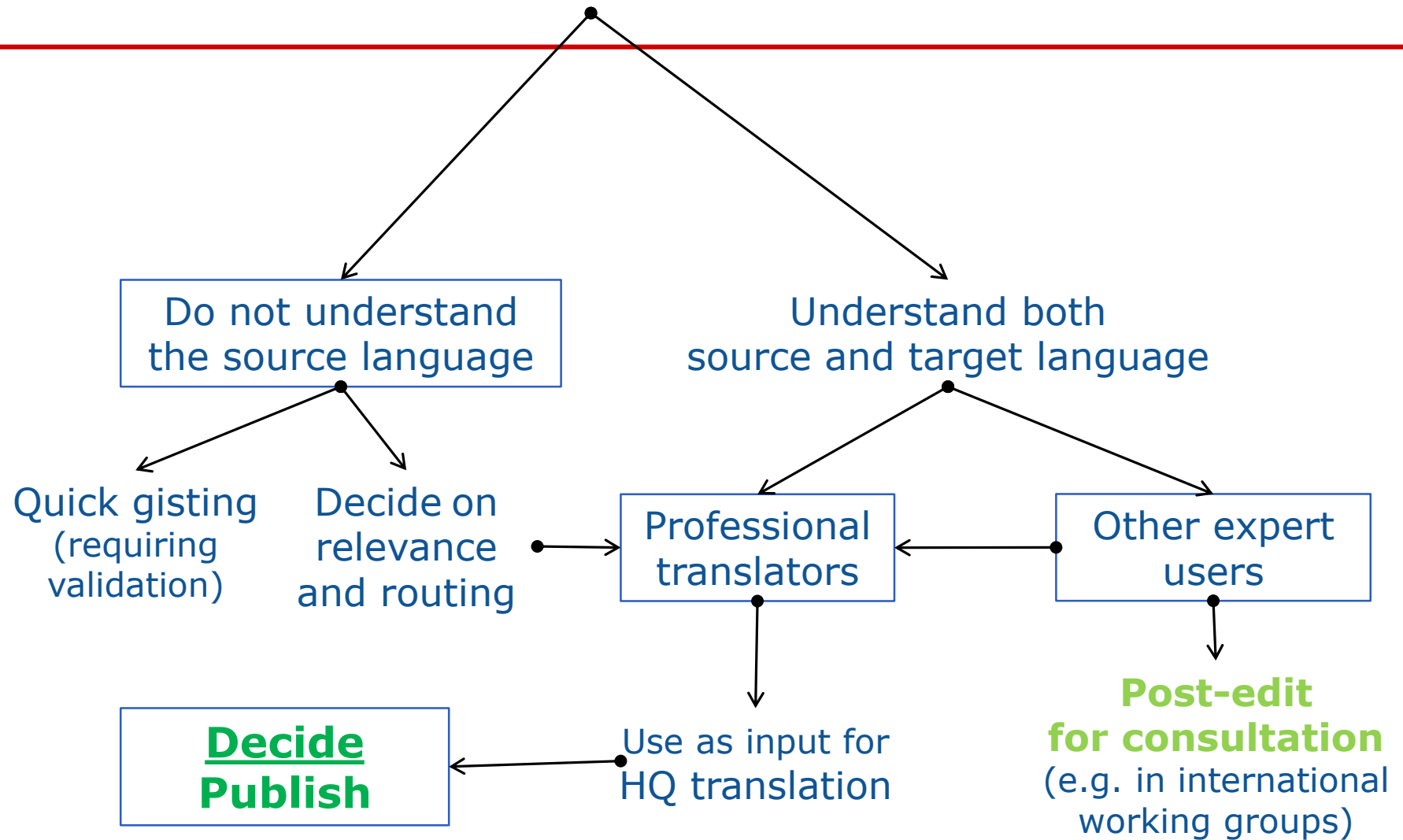
MT is the only viable solution for:

- quick and cheap **access to information** in foreign languages.
- understanding **information received** in a foreign language that otherwise could not be used or would require substantial time and costs to translate.
- making **multilingual use of websites** possible
- facilitating cross-lingual information **search** and **analytics**.



That is why machine translation (MT) is a **critically important technology for multilingual Europe**

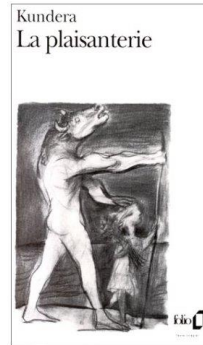
Machine Translation users



Translation is a complex process

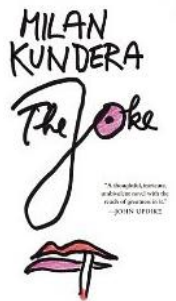


- Czech



ISBN: 978-2070366385 (1975)

- 1st French Edition



ISBN: 978-0060995058 (1993) / (1969)

- English



ISBN: 978-2070703739 (1985)

- 2nd French Edition

Vision

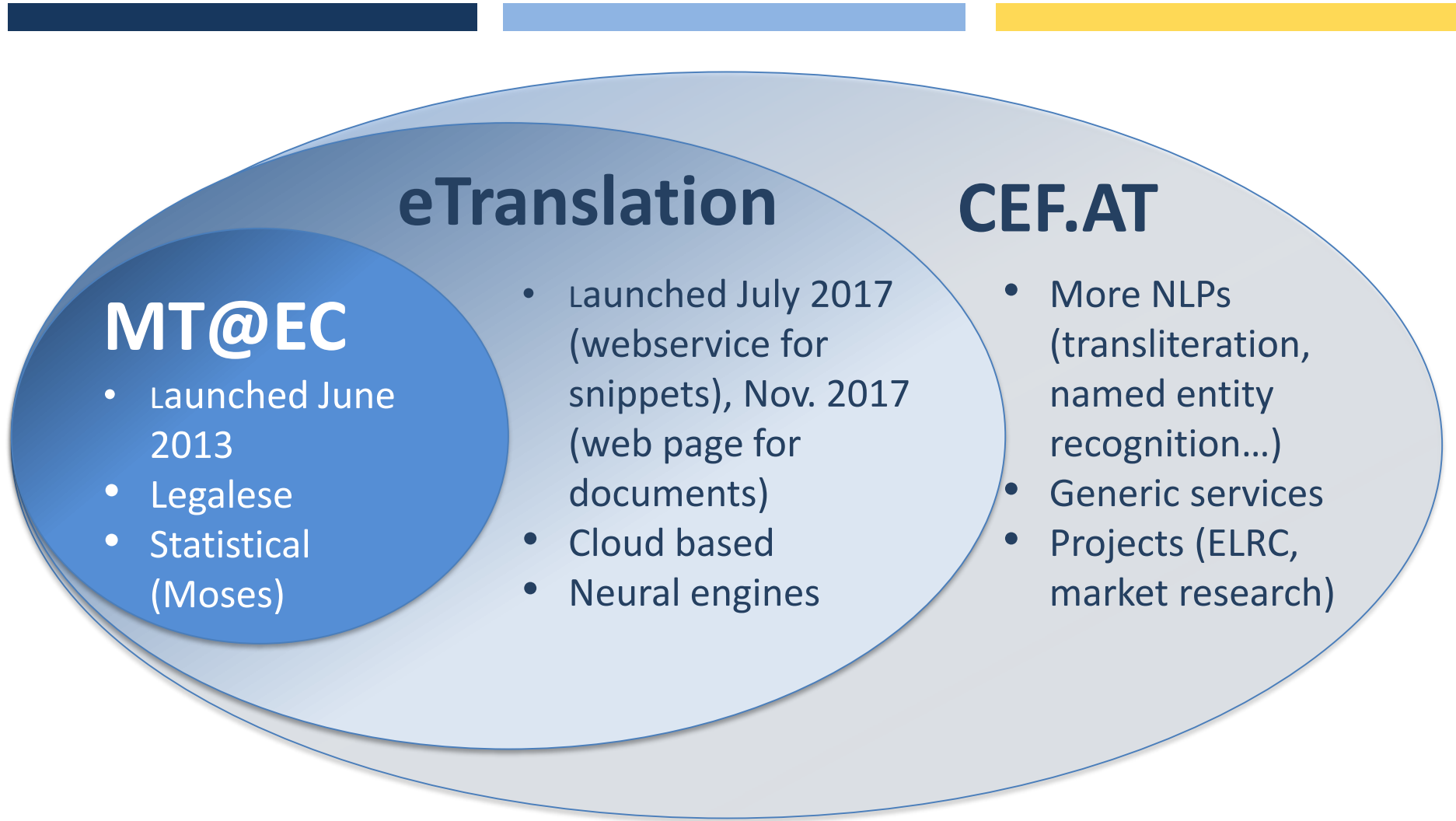


*Wouldn't it be great if I could start using a public service
in any Member State from any place
and obtain the information in my mother tongue?*



The CEF eTranslation platform @ work

Markus Foti
eTranslation Programme Manager
DGT R.3



- Available for:
 - individuals (submit documents through a web page)
 - machine-to-machine use
- Users:
 - Digital Service Infrastructures (EESSI, ODR, Open Data Portal, Europeana, etc.)
 - System suppliers (EURLex, N-Lex, Internal Market Information system...)
 - Individuals in public administrations
- Benefits:
 - Increase speed and productivity
 - Reduce costs
 - Facilitate information exchange



MT@EC - Machine Translation

Markus FOTI
European Commission

[About](#) | [Feedback](#) | [Help](#)



[My settings](#) English ▼ [Logout](#)

[Translate documents](#)

[Translate text](#)

[My translation requests](#)

Select at least one source file :

[Choose files...](#)

OR

[Drag files here.](#)

Translate from...

Translate into...

Domain

Output format:

☐ E-mail me my translation.

☐ Delete after download.

[Translate document](#)

Selected files

Drop files to upload (or click)

Supported formats:      

From *

To *

Domain

Output format

Same as source

TMX

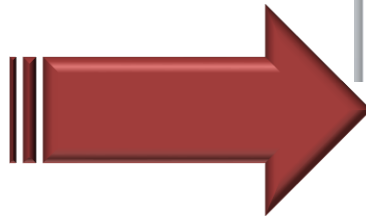
XLIFF

☒ E-mail me my translation

Translate document

One original,
many translations...

English



Français

Deutsch

Italiano

Português

Polski

Lietuvių kalba

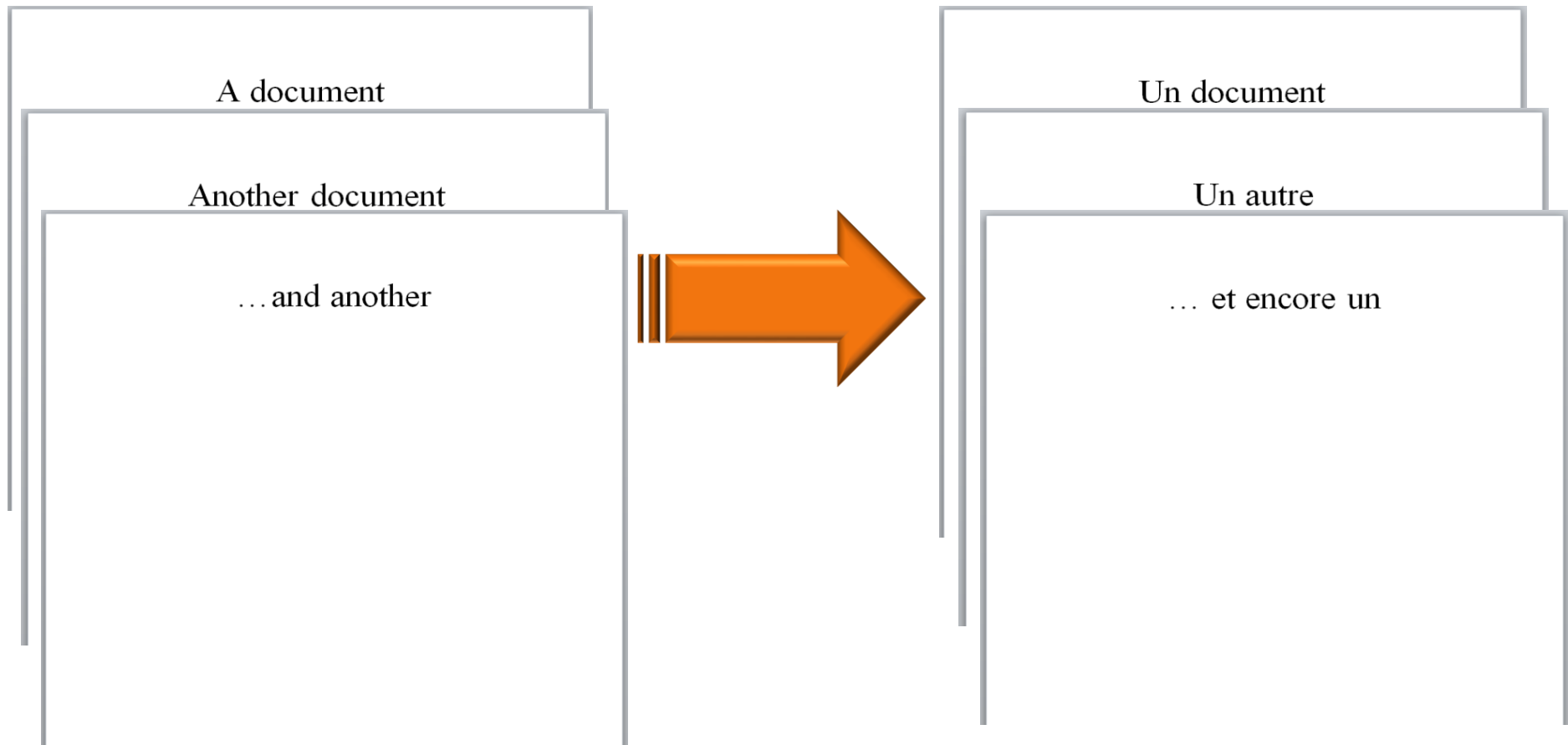
Ελληνικά

Български език

Svenska

All 24 EU languages

Many documents at once...



eTranslation protects your privacy

- All documents deleted after 24 hours or after delivery (on demand)
- IPR rights transferred
(Your translation belongs to YOU!)

No one

looking

at your translations!

- Open to DSIs and public administrations
 - Contact CEF-AT@ec.europa.eu with your request and use case
 - We will provide the technical documentation on how to connect
 - SOAP request or RESTful interface
 - Contact us for credentials
 - Adapt your service to multilingual use!

STRATEGIC
RESEARCH
AGENDA FOR
MULTILINGUAL
EUROPE 2020

edited by the
META Technology Council

<http://www.meta-net.eu/sra-en>

- ***Published in early 2013.***
- ***First strategic research agenda for our field.***
- ***Complex process of collecting and shaping technology visions.***
- ***Hundreds of researchers participated.***
- ***Very broad topics around multilingual Europe in general.***

MT

Text Analytics

Speech

Resources

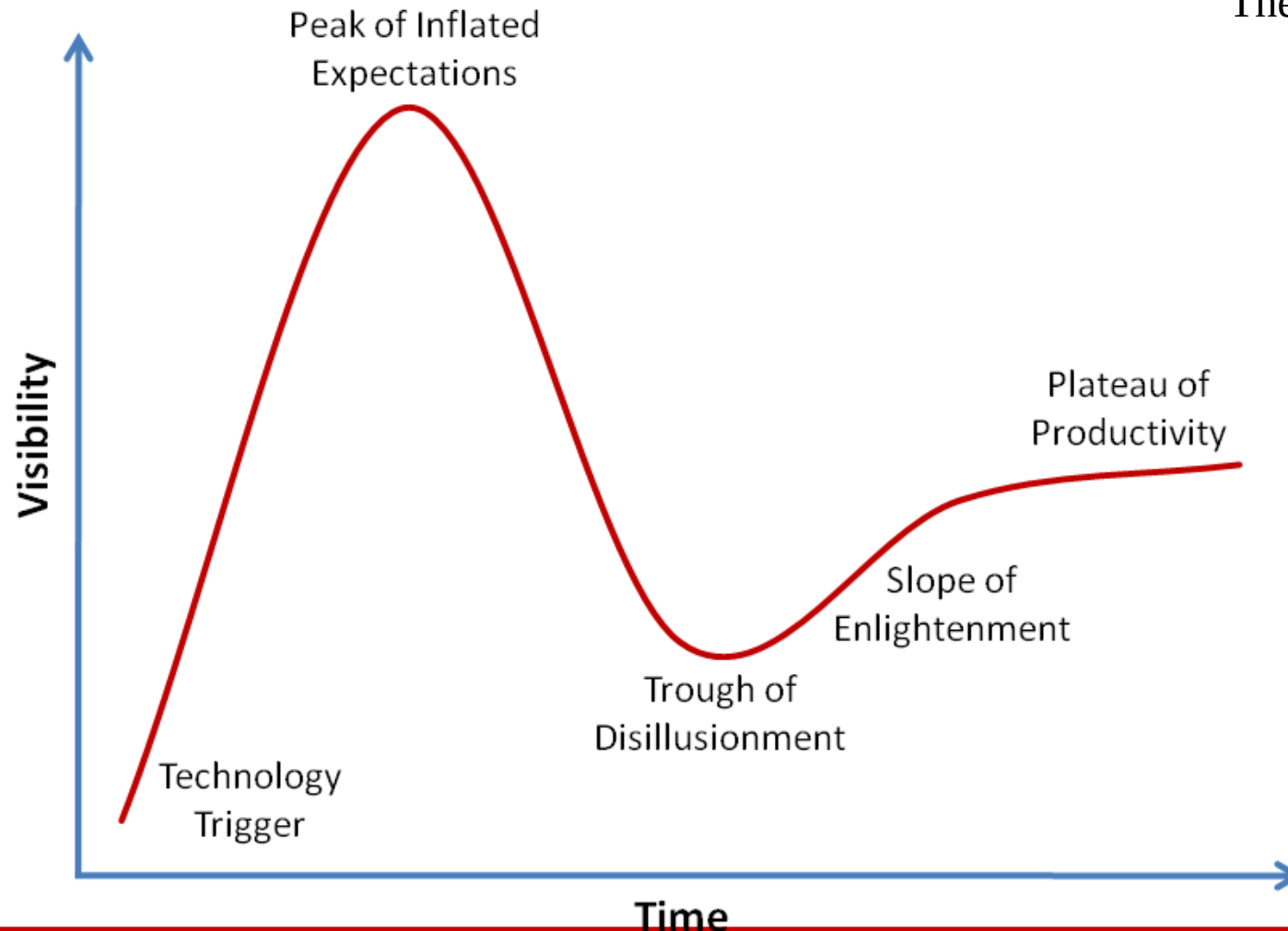
good	moderate	fragmentary	weak or no support through LT
English	French, Spanish	Catalan, Dutch, German, Hungarian, Italian, Polish, Romanian	Basque, Bulgarian, Croatian, Czech, Danish, Estonian, Finnish, Galician, Greek, Icelandic, Irish, Latvian, Lithuanian, Maltese, Norwegian, Portuguese, Serbian,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

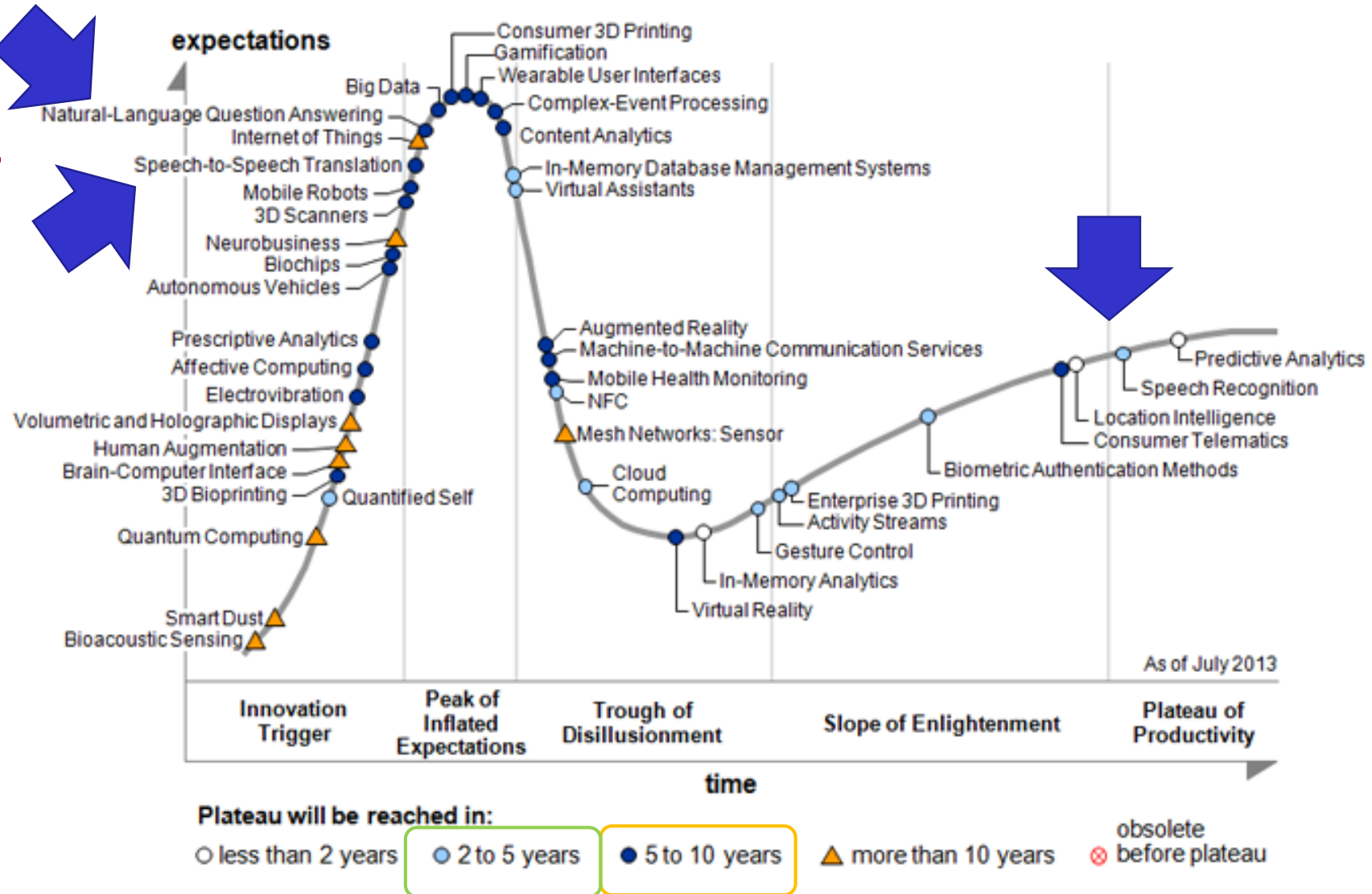
What are the trends ... Challenges for the next « decade! »

1. Introduction
 2. Languages: instruments of cultures, identity and business
 3. Language Technologies: some examples, illustrations
 4. Special focus on Automated Translation
 - Automated/Machine Translation , need for Language Resources (Data sets)
 - The MT@EC and the next generation (CEF-AT)
 - How can we help to improve it
 5. What are the trends ... Challenges for the next « decade! »
 6. Quick conclusions
 7. Q/A Session
-

Where do we stand today ... techno trends

The Gartner Hype Cycle





Source: Gartner August 2013

The 2013 Emerging Technologies Hype Cycle highlights technologies

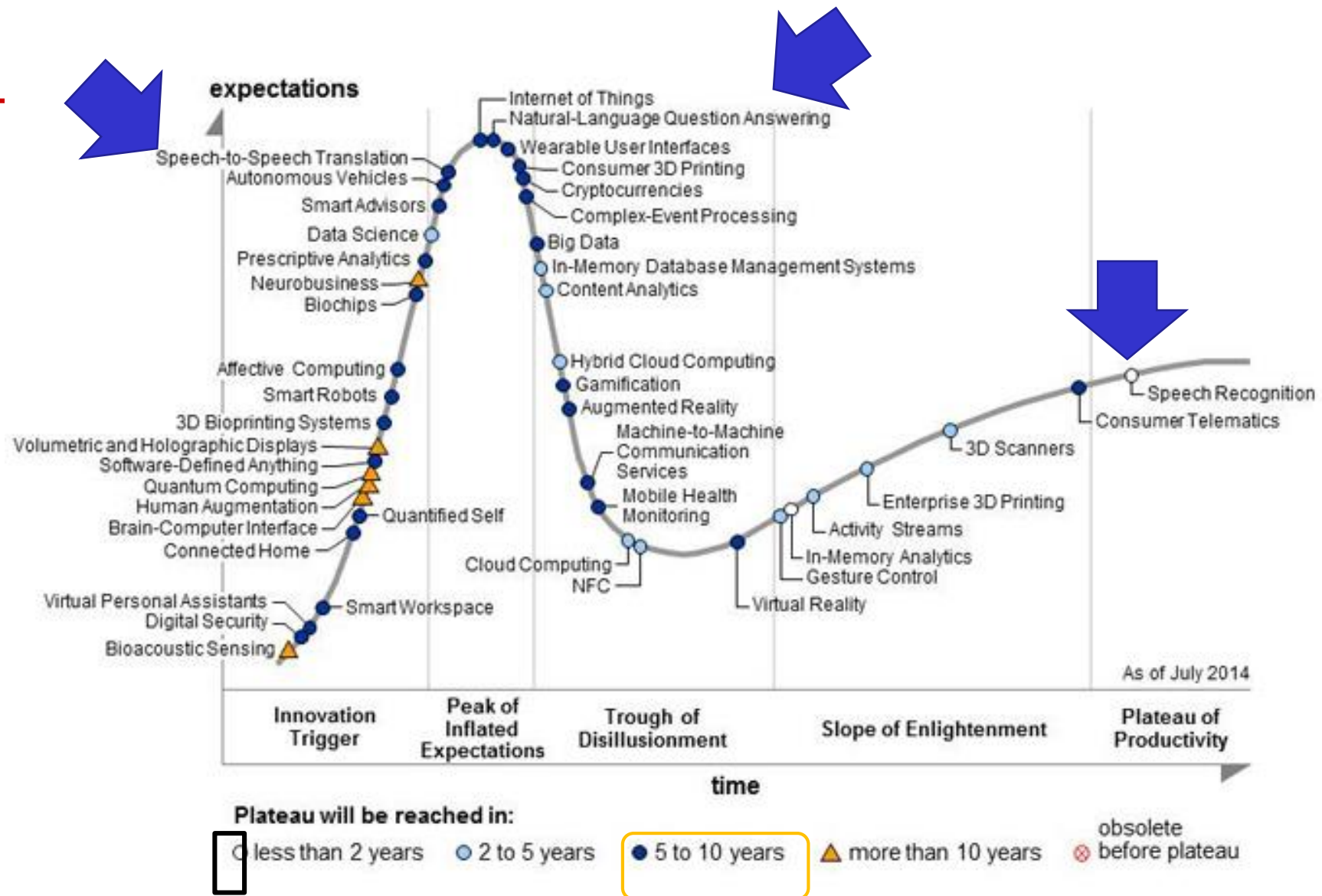
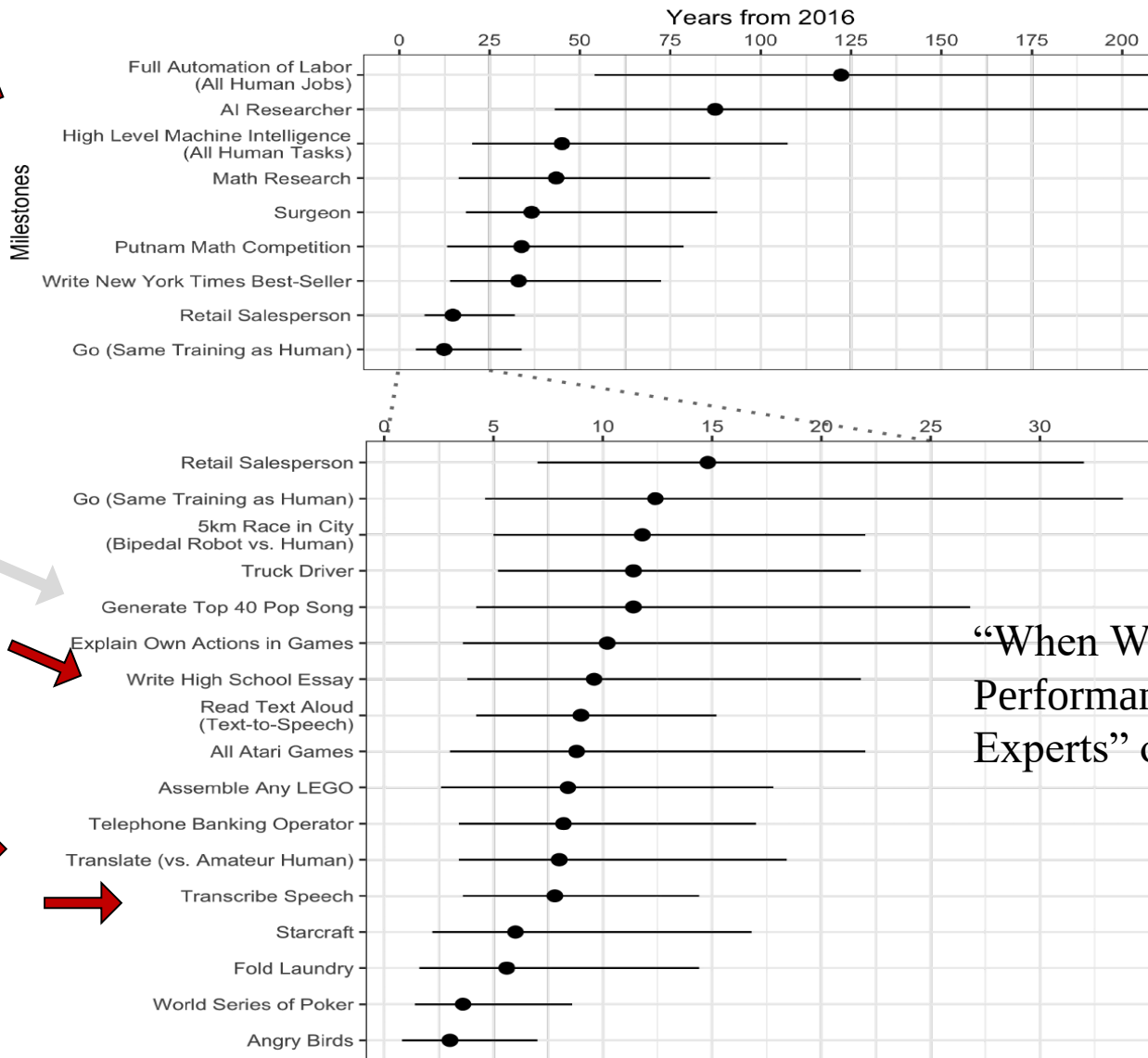


Figure 1. Hype Cycle for Emerging Technologies, 2014

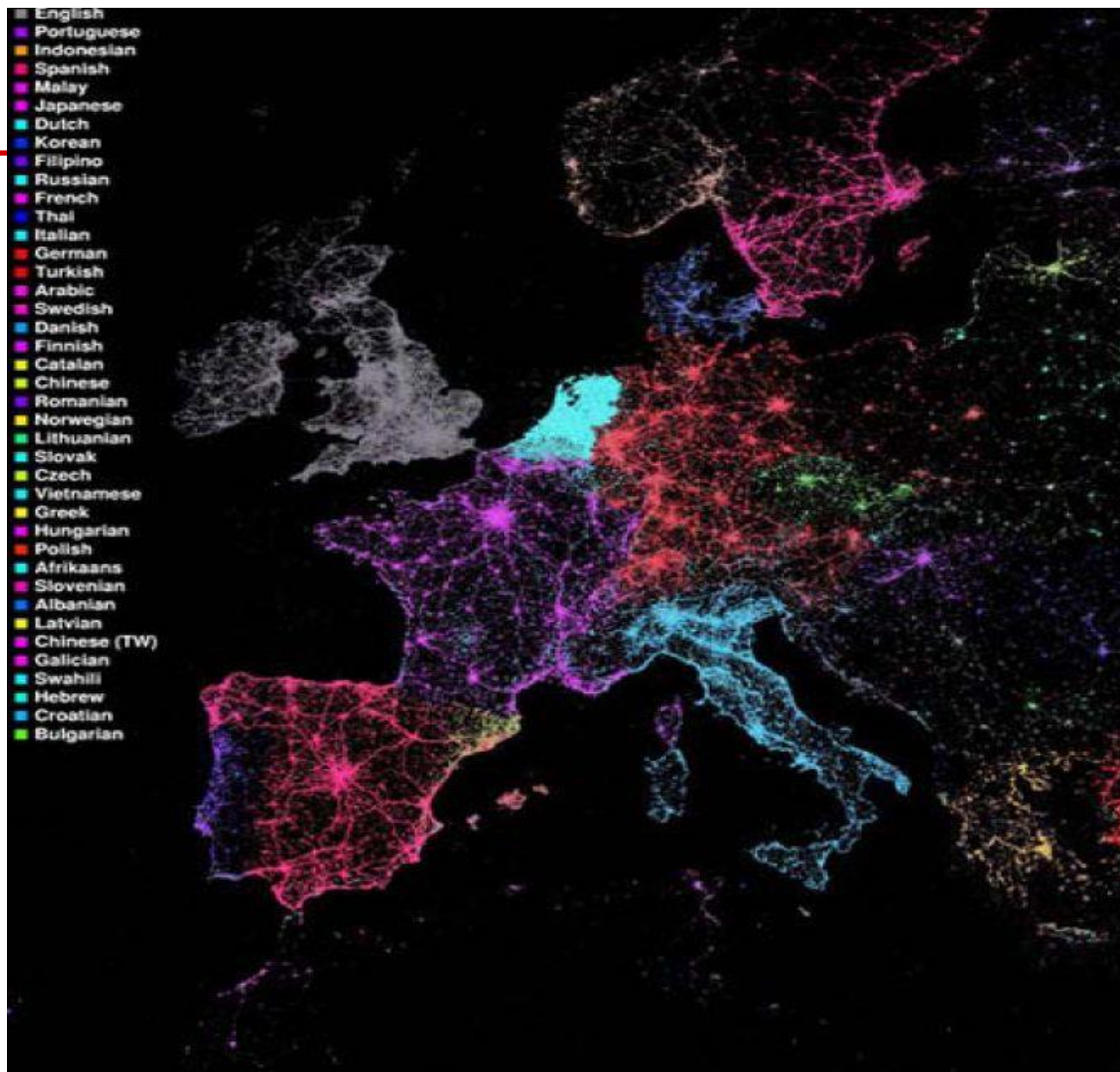


<https://slator.com/academia/yale-oxford-enter-business-predicting-end-human-translator/>

“When Will AI Exceed Human Performance? Evidence from AI Experts” on May 24, 2017.

Challenges ... Trends

- More (and more) Data
 - New techniques Neural Networks approaches (less data needed to start)
 - Other technologies as seen in the Gartner Hype Cycle will emerge (e.g. affective computing)
 - More work on the less resourced languages (those without market & lucrative segments)
 - Only 400 languages have more than 1M speakers
-

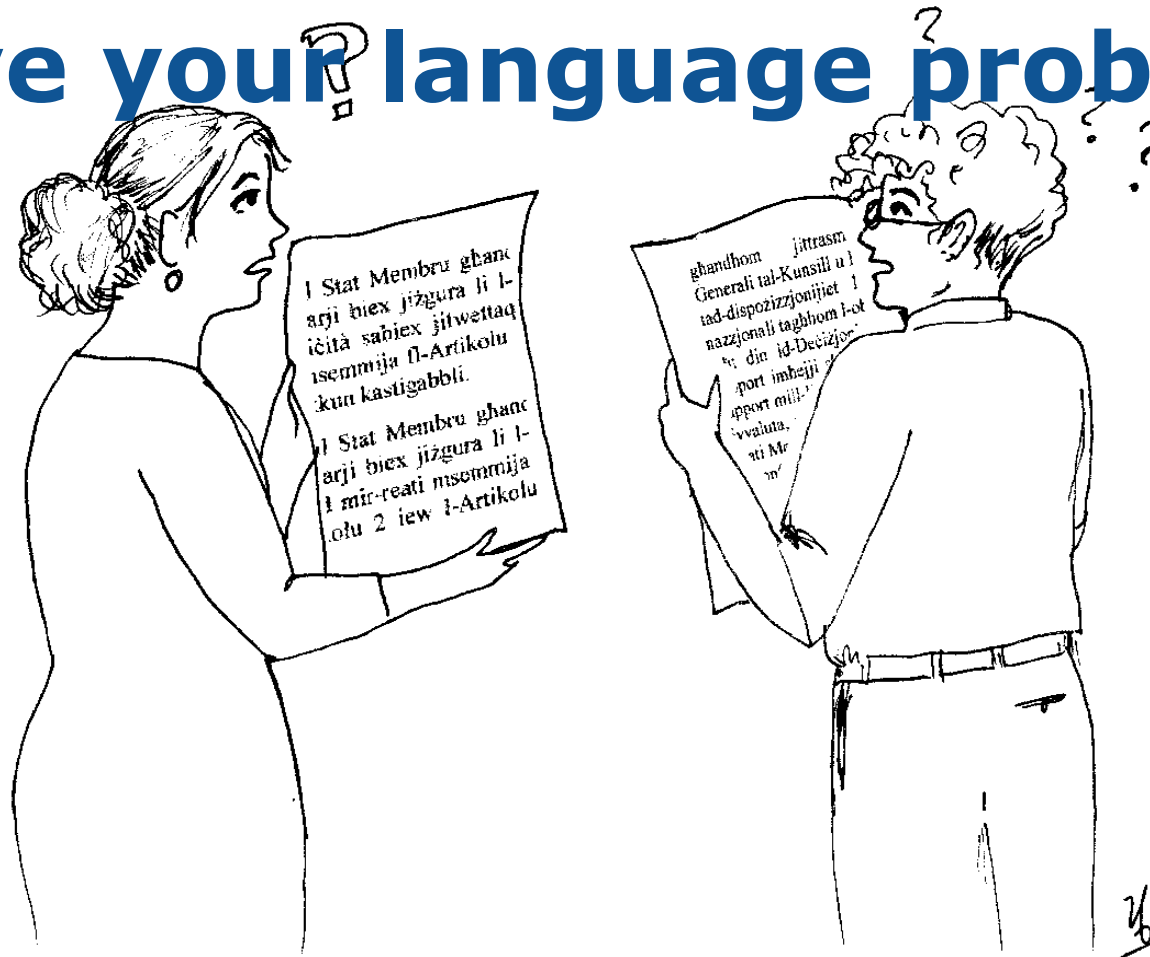


THE EUROPEAN PARLIAMENT
VOTED IN FAVOR OF THE
RESOLUTION ON "**LANGUAGE
EQUALITY IN THE DIGITAL AGE**"
ON 11. SEPTEMBER

“18. Calls on the Commission and the Member States to develop strategies and policy action to facilitate multilingualism in the digital market; requests, in this context, that the Commission and the Member States define the minimum language resources that all European languages should possess, such as data sets, lexicons, speech records, translation memories, annotated corpora and encyclopaedic content, in order to prevent digital extinction;”

“21. Calls on the Commission to make as a priority of language technology those Member States which are small in size and have their own language, in order to pay heed to the linguistic challenges that they face;”

Solve your [?]language [?]problems!



Oh! One last slide ...

What happens to translators?

- Machine Translation makes mistakes (like humans) – important texts must always be checked and validated by people!
 - Translation automation mainstreams multilingualism: people grow used to finding all content in their own languages.
 - Translation automation multiplies translation volumes to levels never seen or dreamed of before.
 - There is an increasing need for language professionals...
 - ...and their job descriptions get more interesting and varied, new professions emerge:
 - High Quality translations, quality control, content management, editing, stakeholder relations, corpus linguistics, computational linguistics, configuration and optimization of translation systems...
 - Translators have a bright future!
-

