



# *ELRC France*

---

*Mon Europe, Ma Language*

---

**Philippe Gelin**

**Head of Sector Multilingualism**  
Directorate-General for Communications Networks, Content and Technology  
European Commission

---

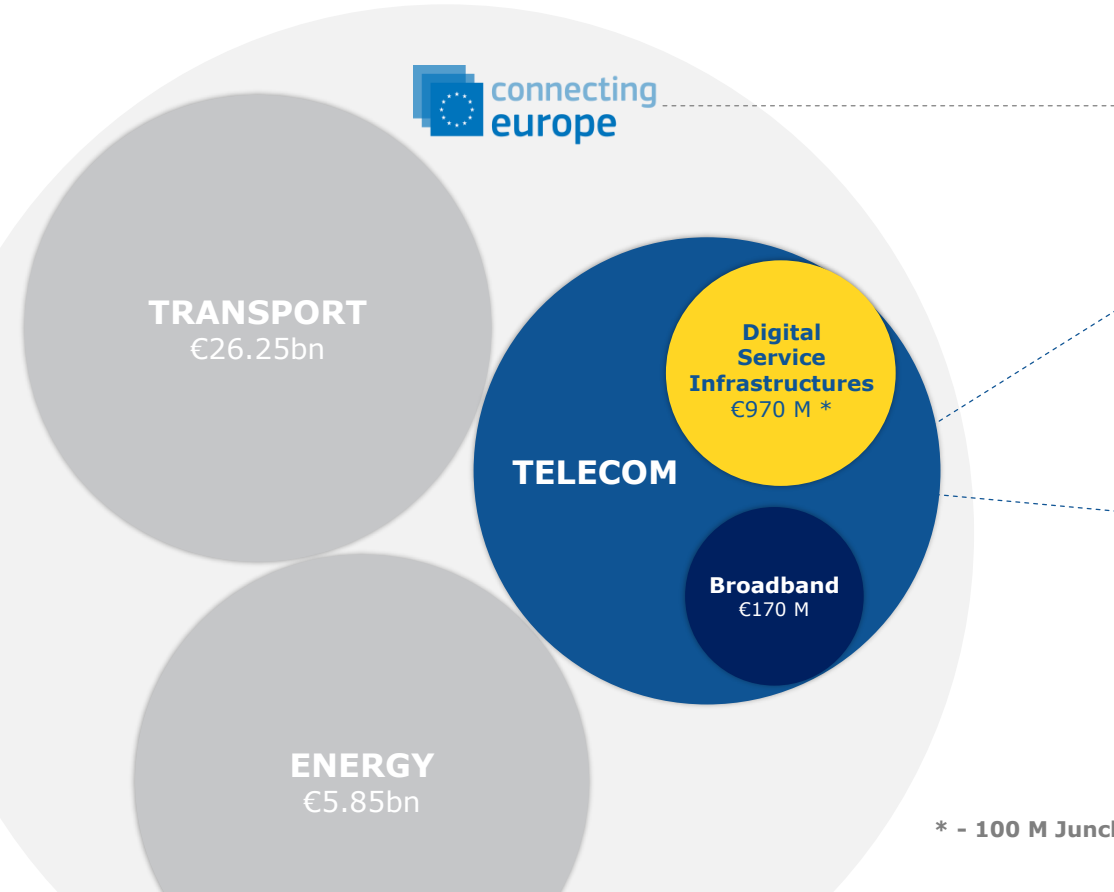
# CEF Automated Translation

Lowers language barriers in the Digital Single Market by:

- offering multilingual solutions for pan-European digital public services and public administrations
- enabling citizens and businesses to benefit from services in their own language



# The Connecting Europe Facility – funding programme 2014 -2020



## **CEF Regulation**

The Connecting Europe Facility (CEF) is a regulation that defines how the Commission can finance support for the establishment of trans-European networks to reinforce an interconnected Europe.

## **CEF Telecom Guidelines**

The CEF Telecom guidelines cover the specific objectives and priorities as well as eligibility criteria for funding of broadband networks and Digital Service Infrastructures (DSIs).

## **CEF Work Programmes**

Translates the CEF Telecom Guidelines in general objectives and actions planned on a yearly basis.

## **CEF Automated Translation**

helps European and national public administrations exchange information across language barriers in the EU, by providing machine translation capabilities that will enable all Digital Service Infrastructures to be multilingual.

\* - 100 M Juncker Package

# CEF Digital at a glance

Funding at EU LEVEL  
(Commission)  
Core Services

Justice, home affairs and  
citizens' rights



**eJustice Portal**



**ODR**

Science and  
Technology



**Open Data**

Business



**BRIS**

Employment and  
Social Rights



**EESSI**

etc.



**IDENTIFY**  
with eID



**SIGN**  
with eSignature



**INVOICE**  
with eInvoicing



**EXCHANGE**  
with eDelivery



**TRANSLATE**  
with eTranslation

Funding for the  
**MEMBER STATES**  
Generic Services



Typically 'deployment' projects  
at national level  
(up to 75% of eligible cost)

Funding at EU level (European Commission)

Automated Translation – CEF Building Block



Standards and technical specifications

Language  
Technology  
Specifications

a layer of technical specifications  
and standards that have to be  
complied with



Services / Software

MT Service  
eTranslation

Data Repository  
ELRC Share

Other services

a layer of sample software that  
complies with the specifications  
and is meant for reuse

Support / Conformance testing

Support

a layer of support services (e.g.  
conformance testing, help desks,  
onboarding services, etc.)



---

# CEF eTranslation

The same automatic translation for two complementary purposes

## DGT eTranslation

- Translators of the EU Institutions
- Digital services of the EU Institutions



## CEF eTranslation

- CEF Digital Service Infrastructures
- Pan-European digital public services
- **MS public administrations**



# At a Glance

← → ↻ 🏠 🔒 <https://webgate.ec.europa.eu/etranslation/translateTextSnippet.html> ★ P ⋮

📱 Apps 🌐 Google 📧 Gmail 🇪🇺 eTranslation is here! 📄 My IntraComm Wel... 🗺️ Space: Connect(ed)...



**Philippe GELIN**  
European Commission

[About](#) | [Help](#) | [Privacy policy](#)

Translate documents **Translate text** My translation requests My settings English ▾ Logout

Start typing your text

Your translation will appear here

E-mail me my translation 0 / 2500 🗑️ 📄 📎

From  To

Domain  ⓘ

# At a Glance

https://webgate.ec.europa.eu/etranslation/translateDocument.html

Apps Gmail Google Perso EU SM BRU



Philippe GELIN  
European Commission

[About](#) | [Help](#) | [Privacy policy](#)

[Translate documents](#)

[Translate text](#)

[My translation requests](#)

[My settings](#)

English

[Logout](#)



Hi

Drop files to upload (or click)

Supported formats:

From \*

To \*

Domain ⓘ

Cutting edge

Output format

Same as source

TMX

XLIFF

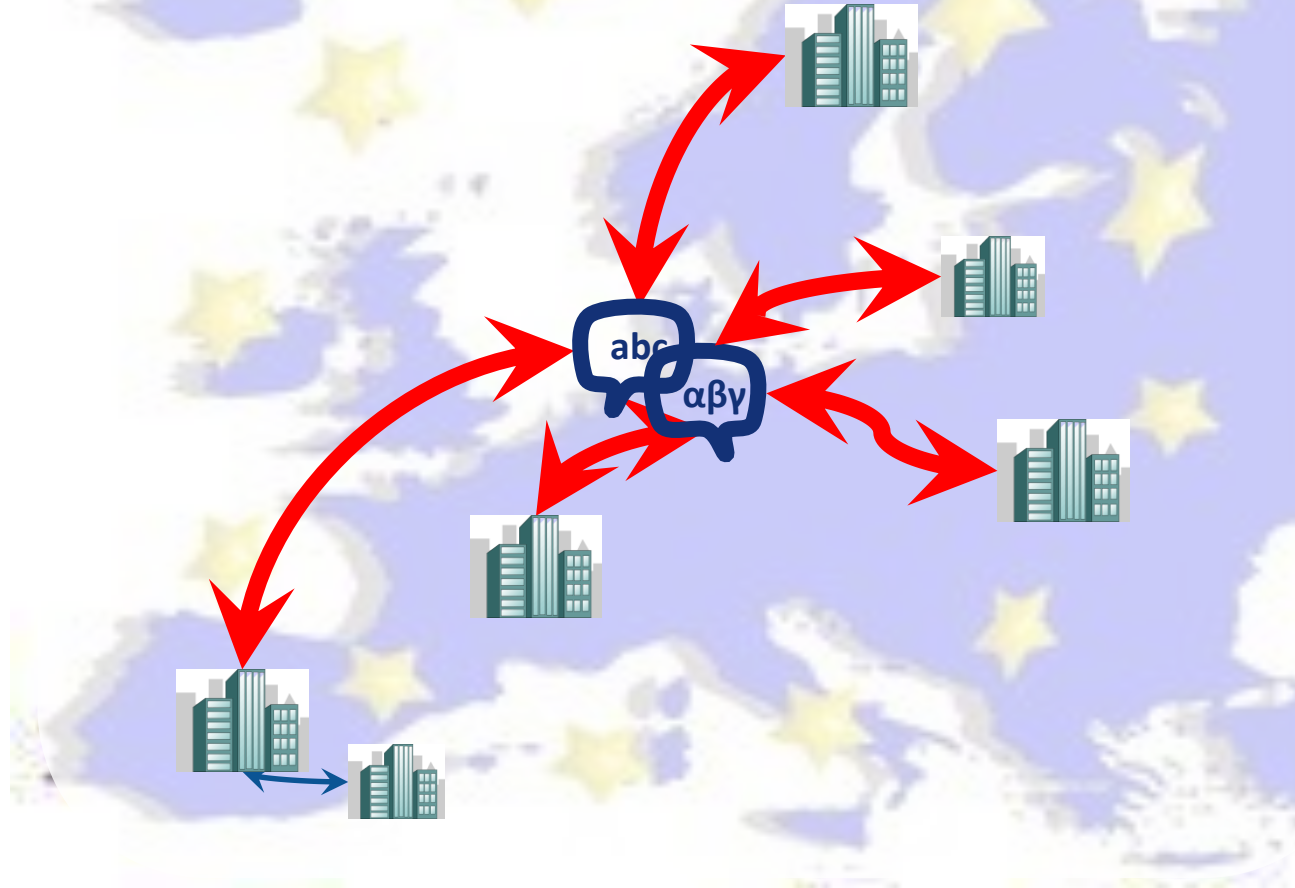
E-mail me my translation

Delete after download.

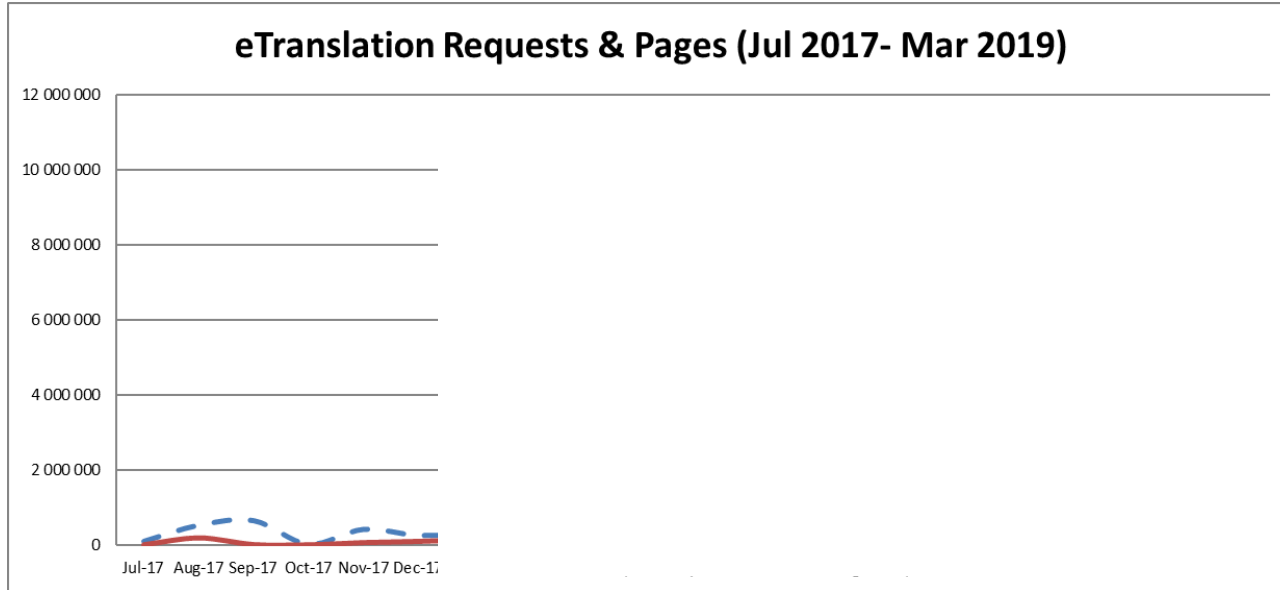
Translate document



# At a Glance – Machine <-> Machine



# eTranslation - State of play

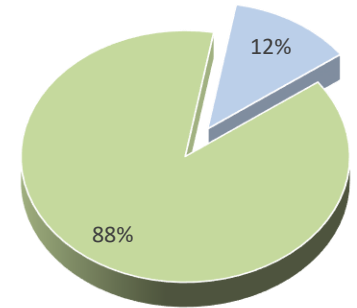


## Web Page use in Q1 2019

- 2.2 million pg. vs 18 million pg.
- incl. 554 661 pg. from MS

## Current record day

- 7 May - 875 773 pages.



■ Web Page ■ Machine-to-machine

Funding at EU level (European Commission)

Automated Translation – CEF Building Block



Standards and technical specifications

Language  
Technology  
Specifications

a layer of technical specifications  
and standards that have to be  
complied with



Services / Software

MT Service  
eTranslation

Data Repository  
ELRC Share

Other services

a layer of sample software that  
complies with the specifications  
and is meant for reuse



Support / Conformance testing

Support

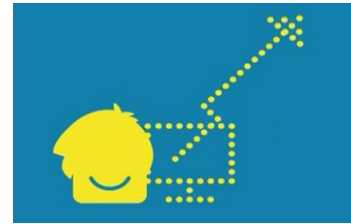
a layer of support services (e.g.  
conformance testing, help desks,  
onboarding services, etc.)



---

# Language resources: the key to success

- More data for all languages
- Better lexical coverage for all languages
- Machine translation adapted to your domain/subject matter



Through ELRC, the European Language Resource Coordination initiative, the Commission is looking for language resource holders who are willing and able to contribute data to improving the translation engines.

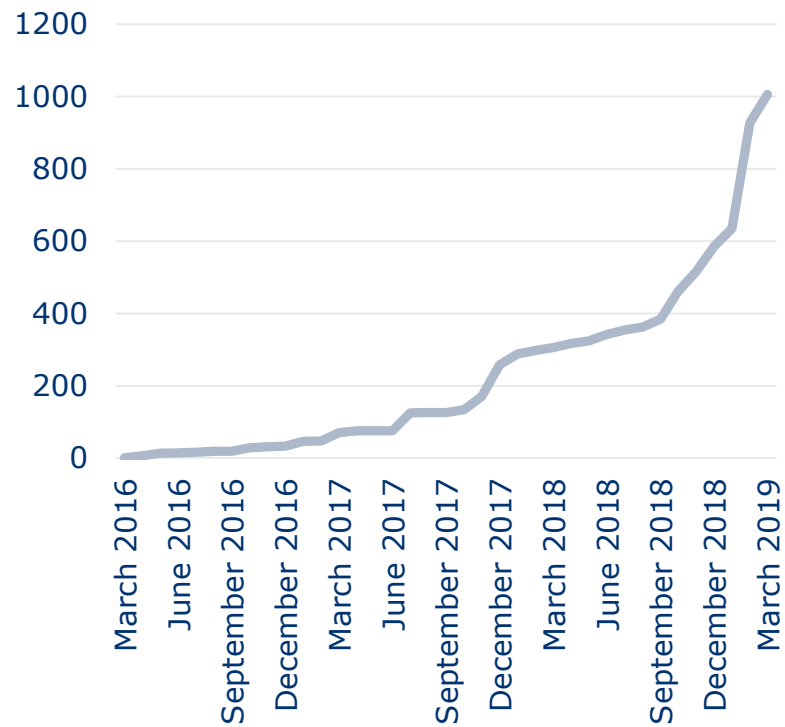
<http://www.lr-coordination.eu/>



# Language Resource Collection

- 1000 language resources
- Instrumental in building EN<>IS and EN<>NB engines
- Contributed to the development of the EN<>GA engine
- Main source for some of the domain-adapted engines
- Will be used for building domain generic engines
  
- Will be used by various other Natural Language Processing tools (summarisation etc.)

No. of unique LR collected through ELRC-SHARE by month





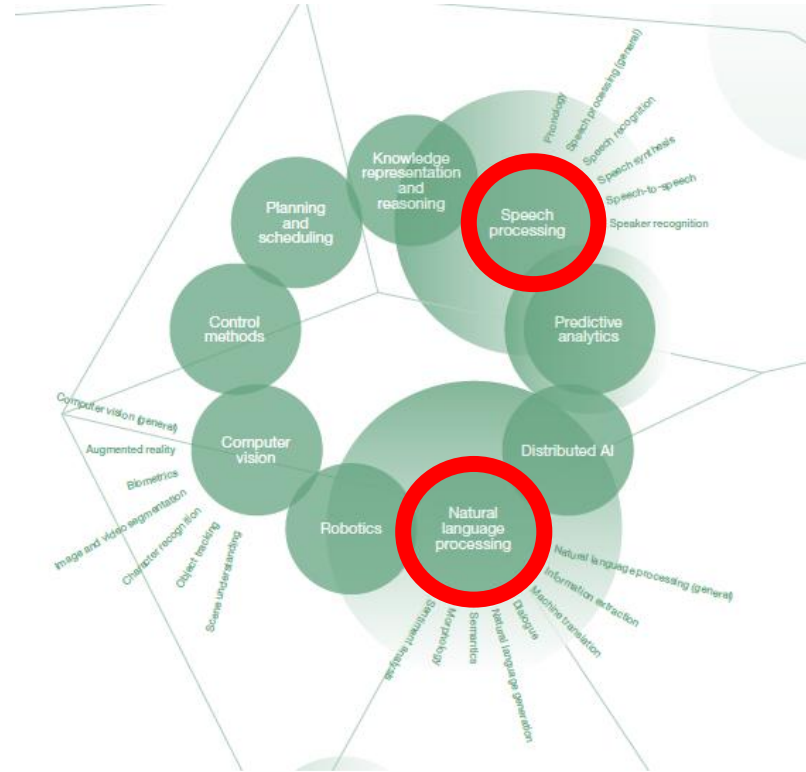
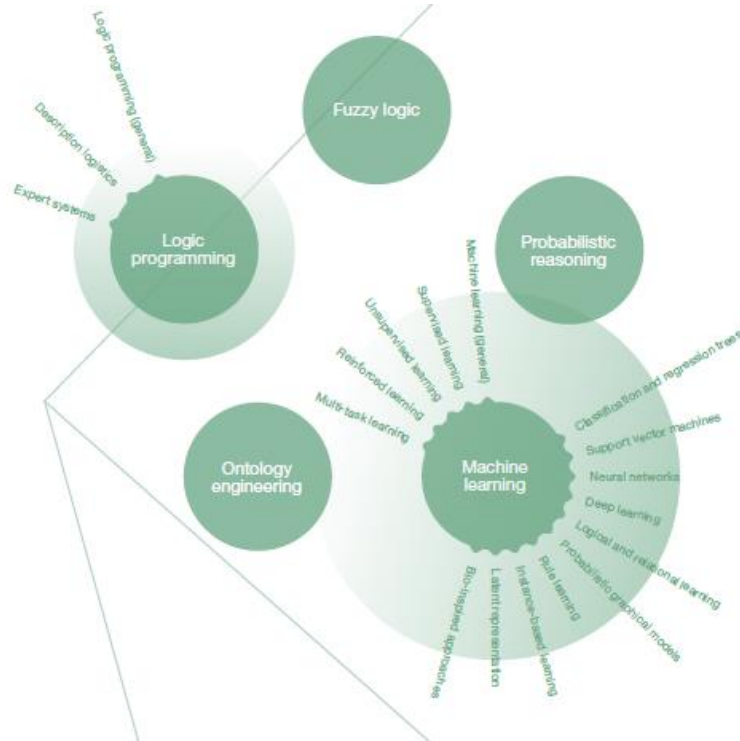
# ***Beyond Translation***

---

***The "new" paradigm in Language  
Technologies***

---

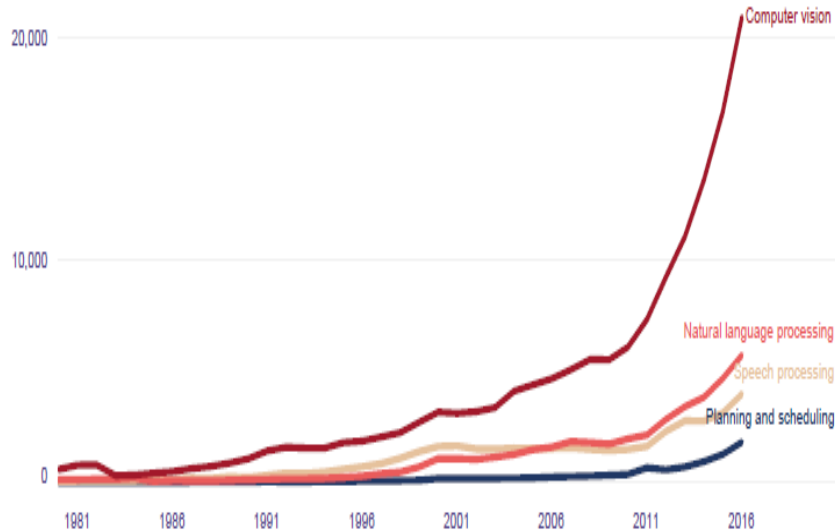
# AI Techniques & Fields of Applications



WIPO (2019). WIPO Technology Trends 2019: Artificial Intelligence.

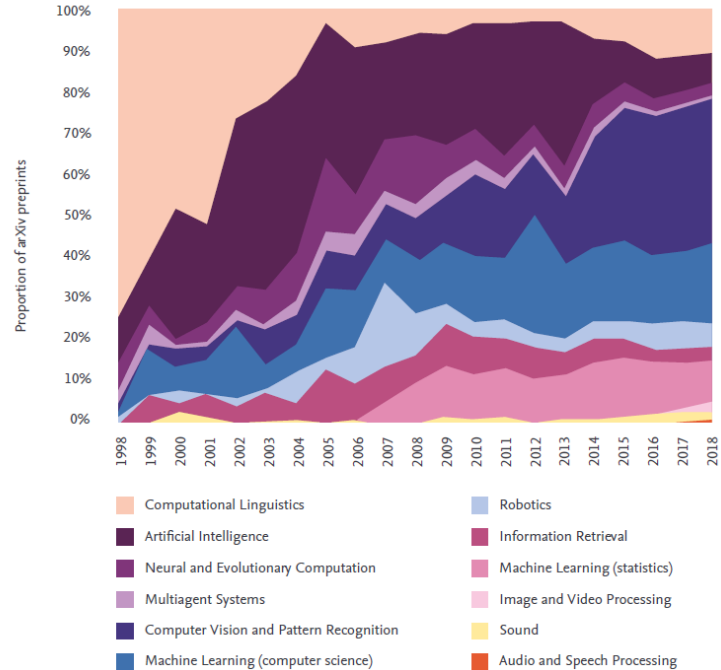
# AI & Language Technologies

Figure 3.11. Patent families for top functional applications by earliest priority year  
 Computer vision grew by an average of 23 percent annually between 2011 and 2016



Note: A patent may refer to more than one category

WIPO (2019). WIPO Technology Trends 2019: Artificial Intelligence.

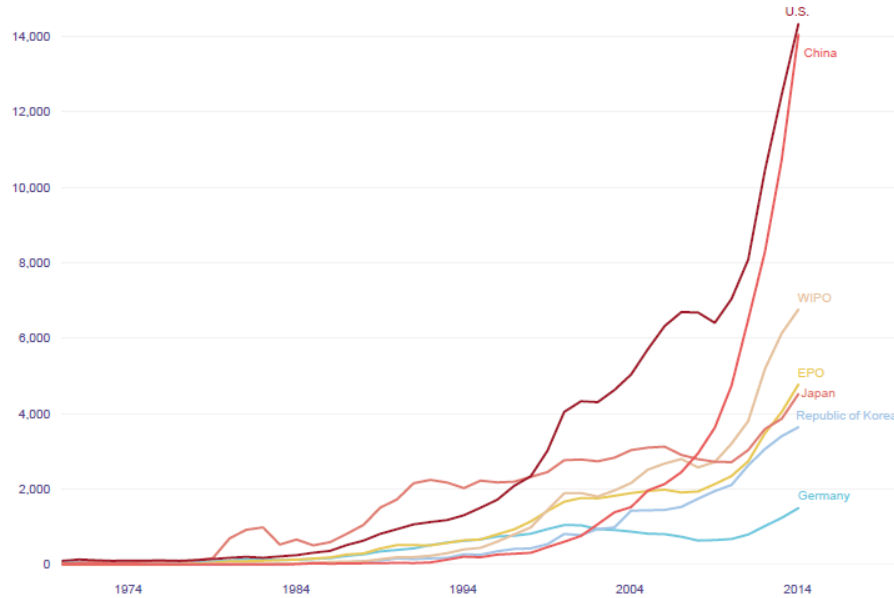


Artificial Intelligence: How knowledge is created, transferred, and used  
 Trends in China, Europe, and the United States (Elsevier, Nov. 2018)



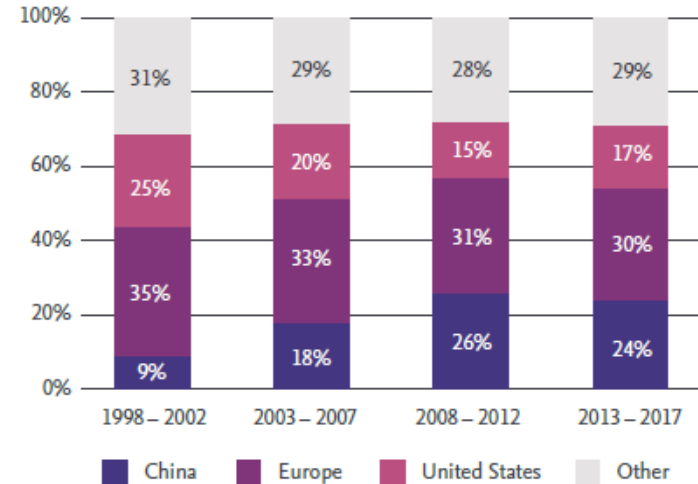
# AI Actors

**Figure 5.3. Number of patent applications for different offices by earliest priority date**  
*The number of patent applications filed in China grew by an average of 25 percent since 2009*



WIPO (2019). *WIPO Technology Trends 2019: Artificial Intelligence*.

**Share of world publications in AI**



**FIGURE 3.4**  
 Share of global publication output in AI (all document types) for periods 1998–2002, 2003–2007, 2008–2012, and 2013–2017, per region; source: Scopus.

*Artificial Intelligence: How knowledge is created, transferred, and used Trends in China, Europe, and the United States (Elsevier, Nov. 2018)*

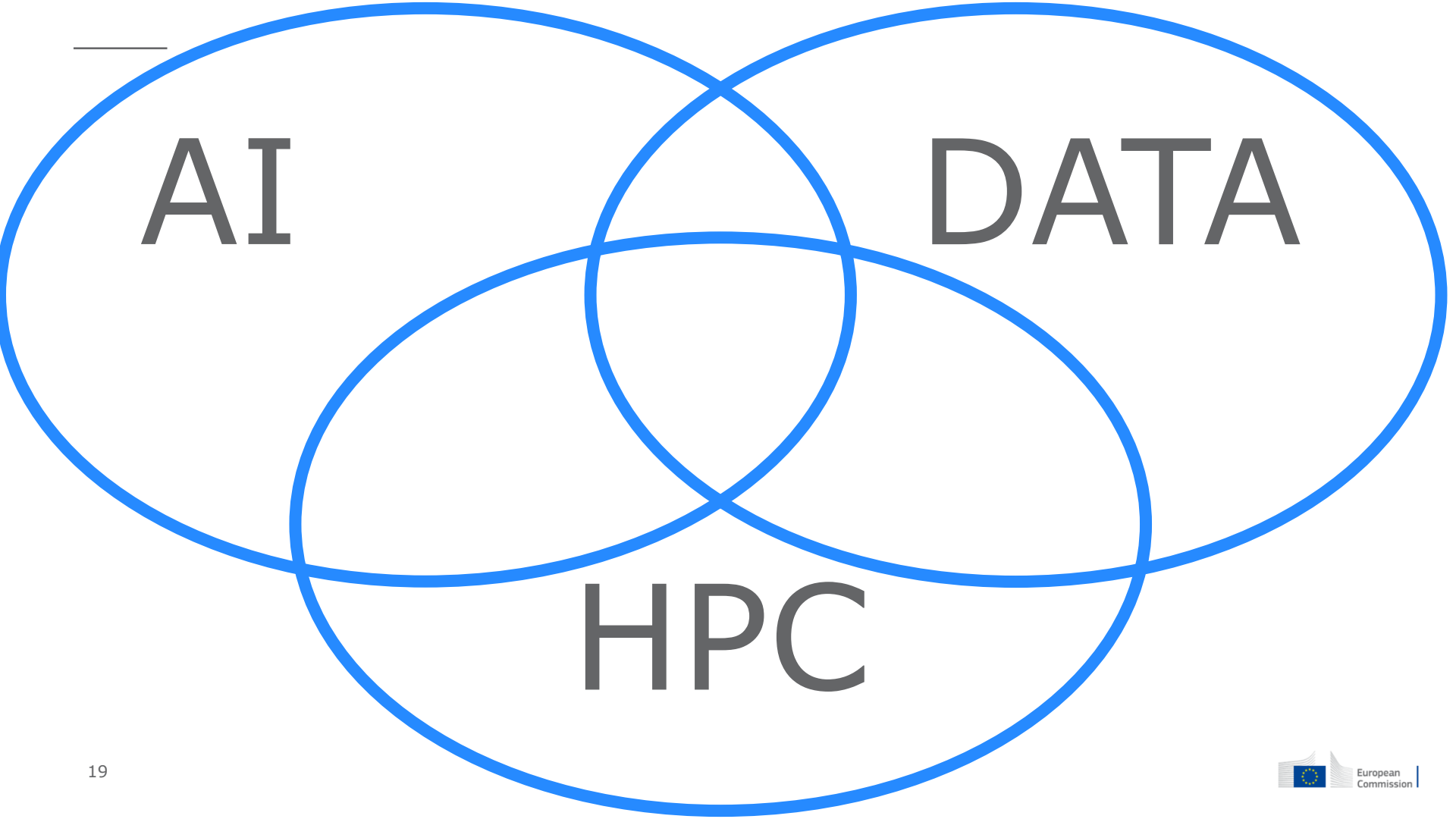


# *Why now?*

---

*Hype or Strike?*

---

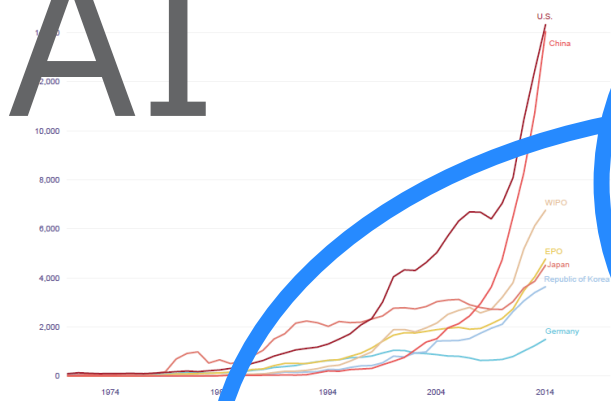


AI

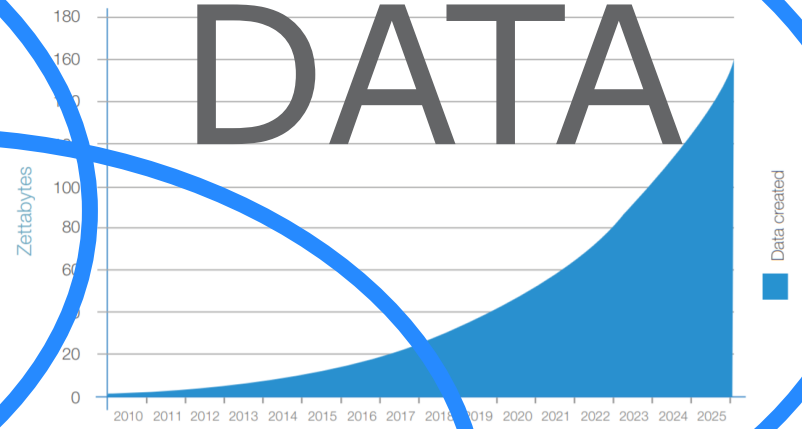
DATA

HPC

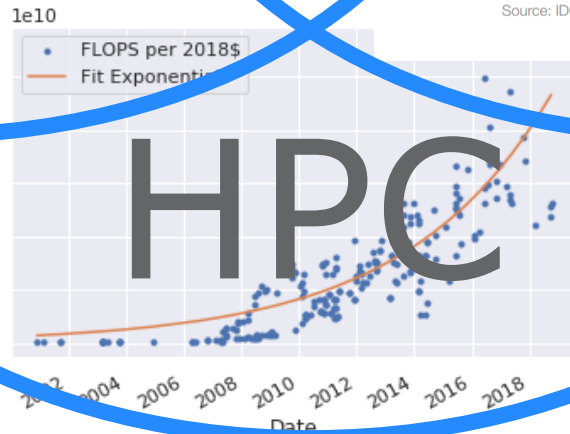
Figure 5.3. Number of patent applications for different offices by earliest priority date  
 The number of patent applications filed in China grew by an average of 25 percent since 2009



Annual Size of the Global Datasphere



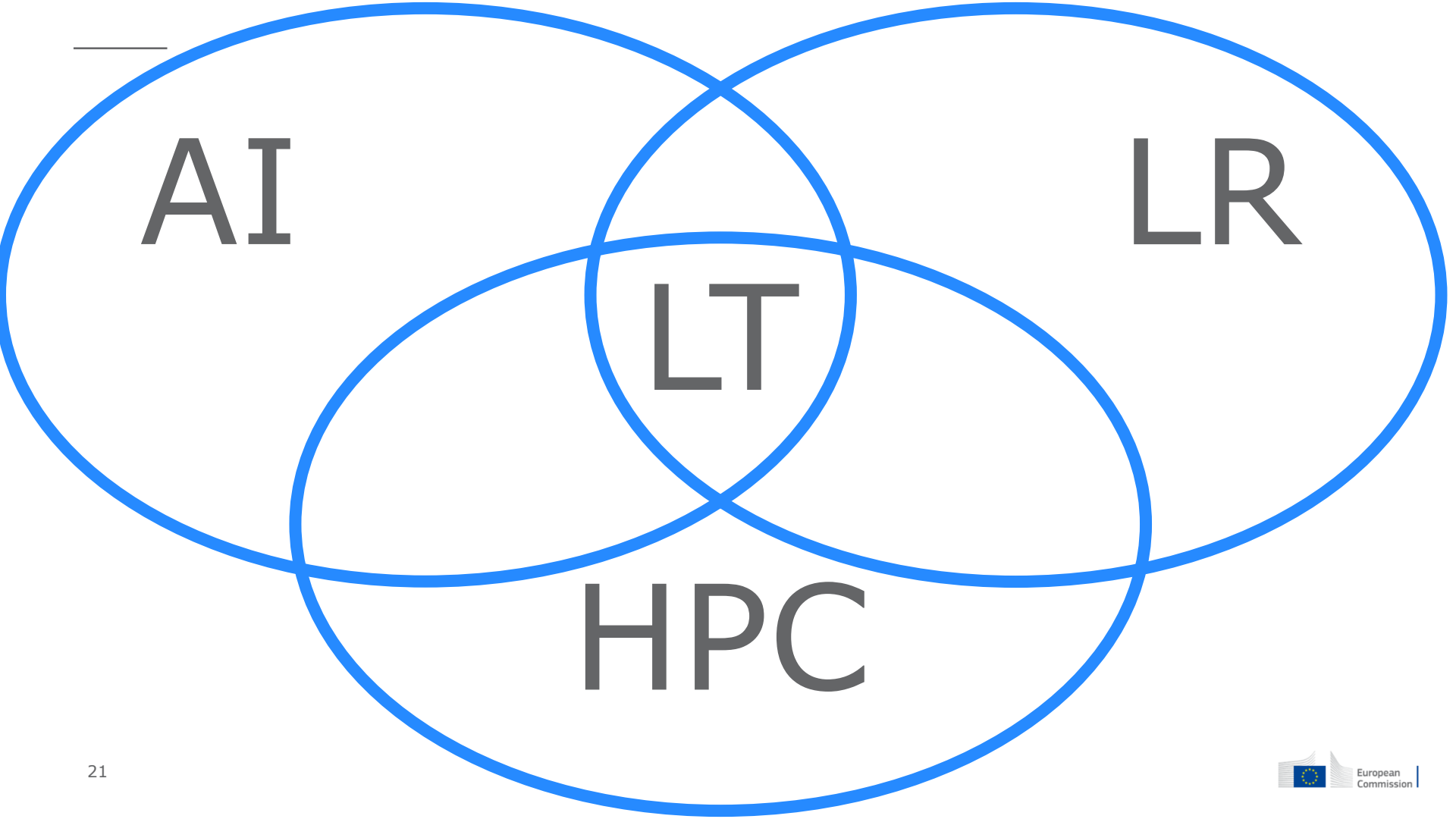
Source: IDC's Data Age 2025 study, sponsored by Seagate, April 2017



AI

DATA

HPC



---

# Shrinking Development Cycle: An example

[...] The development of Marian received funding from the European Union's Horizon 2020 Research and Innovation Programme under grant agreements

- 688139 (SUMMA; 2016- 2019),
- 645487 (Modern MT; 2015-2017),
- 644333 (TraMOOC; 2015-2017),
- 644402 (HimL; 2015-2017) [...]

Transition of **eTranslation from SMT to NMT from 2017 and 2018 to integrate** this new technology **into the DGT** translation tool and **provide services** across all CEF DSIs.



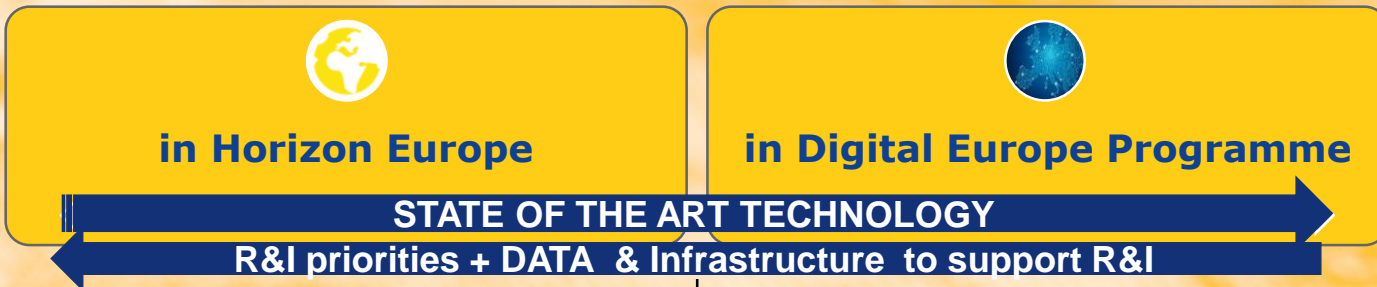
# *Future*

---

---



# MULTILINGUALISM IN MFF



## RESEARCH & INNOVATION

- **Cultural Heritage:**  
*Support Language Preservation*
- **Next Generation Internet:**  
*Support research to achieve digital language transparency*

## CAPACITY BUILDING

- *Artificial Intelligence – Natural Language Processing Tools & Resources*

## DEPLOYMENT & BEST USAGE

- ***Support public administrations and areas of public interest***
- ***Support deployment of Language Technologies***



# MULTILINGUALISM IN MFF



## in Horizon Europe

### Support Language Preservation

*Why: Most European languages are at risk of digital extinction*

*How: **Collect language resources***

*Impact: Digital Humanities and Language Technologies*

*Where ?*

*Inclusive Society*

*Cultural Heritage*

# MULTILINGUALISM IN MFF



## in Horizon Europe

**Support research to achieve digital language transparency**

*Why: EU competitiveness and independence to external technology providers*

*How: **Innovate in Language Technologies***

*Impact: keep EU Researchers Lower barriers in DSM*

**Where ?**

**Next Generation Internet  
Human-centric**

**Human-centric Internet  
applications**



# MULTILINGUALISM IN MFF



## in Digital Europe Programme

### ***Support deployment of Language Technologies***

*Why: Increase EU competition in a key technology*

*How: Provide languages data and tools to public and private and foster collaboration*

*Impact: Public and Private including SMEs*

### ***Where ?***

- ***Artificial Intelligence***
- ***European Data spaces***
- ***Digital Innovation Hubs***

### ***Language Technologies***

# MULTILINGUALISM IN MFF



## in Digital Europe Programme

***Support the modernisation of public administrations and areas of public interest***

*Why: offer EU citizen access to information in any language*

*How: Provide languages technologies enablers*

*Impact: Public administrations, Areas of public interest.*

***Where ?***

***Deployment, best use of digital capacities and Interoperability***

***Modernisation of administrations***

# MULTILINGUALISM IN MFF

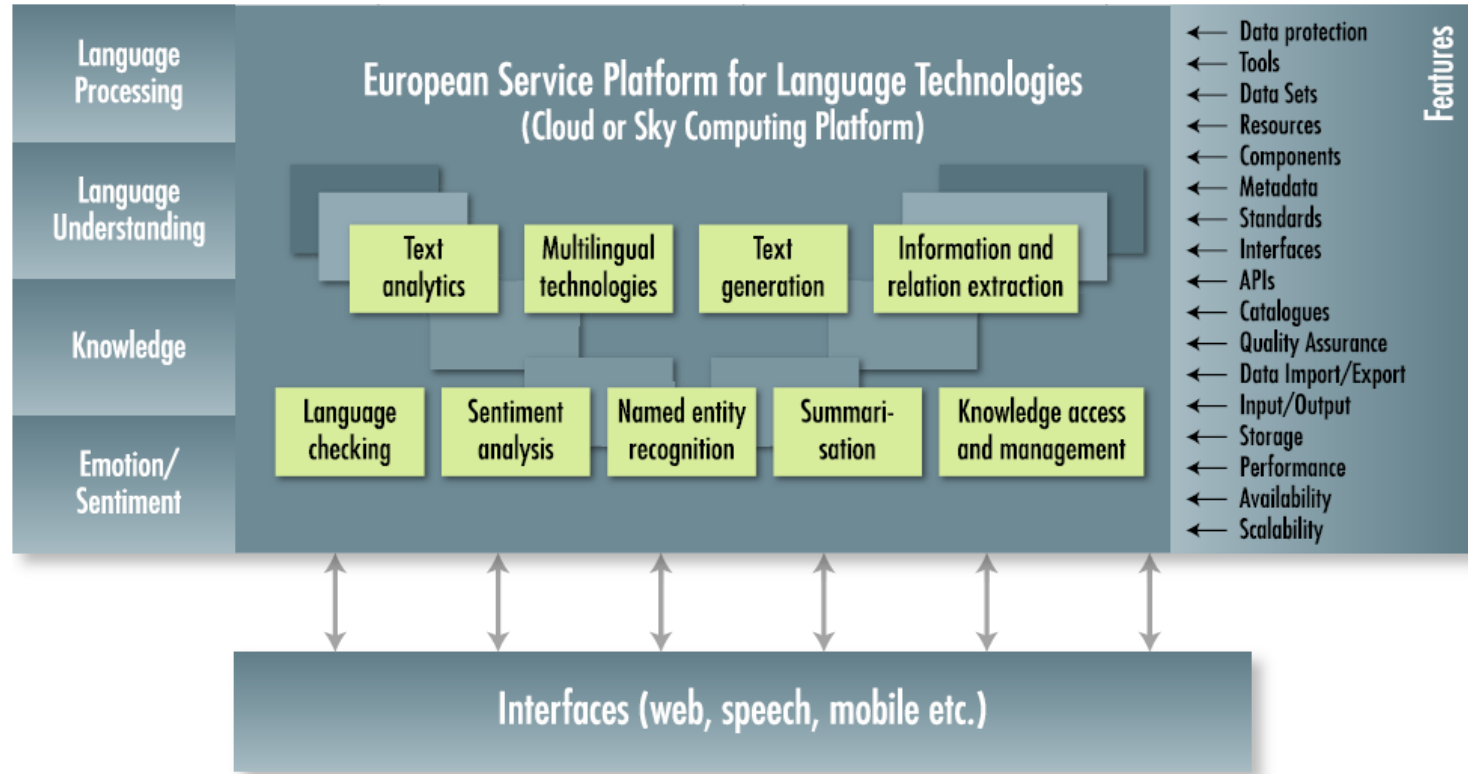


## in Digital Europe Programme

### *Language Technologies*

- *Ecosystem of technology providers*
- *Access to the necessary **data sets***
- *Access to the necessary **computing power***
- *Access to the necessary **algorithm***
- ***Localisation** and uptake of language technologies.*

# On its way : ELG – European Language Grid



From: META-NET  
SRA (2013)





***Examples***

---

***Inter Institutions***

---

Workspace - NewsDesk - Mozilla Firefox

https://newsdesk.emm4u.eu/ND1/workspace.html?e=1428793153960

Latest News [x] Email Target [x]

Total Articles: 4

Latest News (Latest News)

1 ANTI-CORRUPTION REPORT (2)

Parlamentswahl in Tschechien: Der Berlusconi von Prag (Parliamentary elections)

My Feeds

Refresh

Feeds

- Medical
- Ebooks
- AI/De
- Security
- Crime
- Ce
- Dr
- Fa
- Fo
- Hl
- HT
- DI
- DI

Saved Search

User defined

salbyron

### Text adopted - Application of the Postal Services Directive - P8\_TA-PROV(2016)0357 - Thursday, 15 September 2016 - Strasbourg - Provisional edition

A, whereas the postal market is still an area of the economy with strong prospects for growth and increasing competition, even though between 2012 and 2013 letter post services shrank by 4.85 % on average in the EU according to the European Commission Postal Statistics Database, which is in line...

Trigger words: European citizens, EU citizens  
[european\\_ONLINE SOURCES](#) lang= en pub: 7 Dec 2016, 16:03

### Investing in Europe's Youth: Questions and Answers

The recent Commission report on the implementation of the 'Youth Guarantee and the 'Youth Employment Initiative shows that the 'Youth Guarantee has become a reality across the EU and is yielding results. Around 9 million young people have taken up an offer under the Youth Guarantee scheme since its implementation.

Trigger words: European citizens, European citizenship  
[EUropa-IN ONLINE SOURCES](#) lang= en pub: 7 Dec 2016, 14:26

### Investing in Europe's Youth: Questions and Answers

What is the initiative 'Investing in Europe's Youth' about? Today the Commission has presented several key actions to support young people in Europe. This includes the launch of the European Solidarity Corps, which was announced by President Juncker in his State of the Union speech on 14 September 2016.

Trigger words: European citizens, European citizenship  
[europe-pressRelease\\_EU PRESS RELEASES](#) lang= en pub: 7 Dec 2016, 12:37

### EU agencies make the case for their effectiveness

Representatives from 45 EU agencies have issued four studies that seek to make the case for their effectiveness. Photo credit: Fotolia The move comes amid fresh questions in some quarters, including from some Eurosceptic parties such as UKIP, about the cost effectiveness of the agencies.

Trigger words: EU citizens, EU citizens  
[european\\_ONLINE SOURCES](#) lang= en pub: 7 Dec 2016, 11:59

### Saisie de 600 kilogrammes d'héroïne dans l'est de la Turquie - (Seized 600 kg of heroin in eastern turkey)

menara 12 Apr 2015 00:38, Language: fr en it de de

### GEO location

Map Satellite

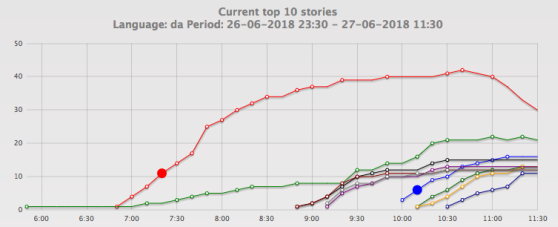
Google

ANTARCTICA Map data ©2016 Terms of Use

142 102 77 217 8 6

### Attributes

- 23 Citizenship Policy
- European Union
- European Council
- European Commission



### Norge sælger sine aktier og forlader SAS

Articles: 44 | Last update: Jun 27, 2018 11:09:00 AM | Start: Jun 27, 2018 6:54:00 AM | Sources: 19 | Peak: 1 | Current rank: 1

### SAS-aktien falder efter Norges salg af aktier med rabat

npinvestor Wednesday, June 27, 2018 11:09:00 AM CEST | info | en | other

Aktien i SAS falder fra åbningen omkring 3,5 pct. til 12,47 kr., efter at den norske stat har solgt alle sine 37,8 mio. aktier i lufttrafiksselskabet til institutionelle investorer. Norges aktiesalg skete til en kurs på 17,25 svenske kr. per aktie, hvilket omregnet til dansk valuta svarer til cirka 12,42 danske kr. ....

More articles...

### Politiet undersøger om brand på rådhus er påsat

Articles: 24 | Last update: Jun 27, 2018 11:24:00 AM | Start: Jun 27, 2018 5:30:00 AM | Sources: 15 | Peak: 3 | Current rank: 2

### Brand på Grenaa Rådhus i nat: Alle medarbejdere arbejder hjemme efter omfattende røgskader

ansvarer Wednesday, June 27, 2018 11:24:00 AM CEST | info | en | other

Grenaa: De seneste dage har politikerne i Norddjurs Kommune haft det hede med at forsøge at finde en løsning på det overophedede budget, hvor der over de næste fire år er underskud på 460 millioner kroner. I nat var brandfolk i aktion på rådhuset i Grenaa for at slukke en brand. Alarmen indløb klokken 3....

More articles...

### 4 min siden 3500 ællinger må aflives efter fund af mild fugleinfluenza

Articles: 16 | Last update: Jun 27, 2018 11:33:00 AM | Start: Jun 27, 2018 10:00:00 AM | Sources: 16 | Peak: 2 | Current rank: 3

3500 ællinger må aflives efter fund af mild fugleinfluenza

en Wednesday, June 27, 2018 10:24:00 AM CEST | info | en | other

Resultaterne viser, at der er tale om lavpatogen fugleinfluenza af typen H5. Lavpatogen betyder, at den er mild sammenlignet med højpatogene typer. Kun de færreste fjærkræ der af lavpatogen fugleinfluenza, men de kan smitte andre fjærkræ, og sygdommen kan ændre sig til den højpatogene type, der har en dødelighed på op imod 100 procent....

More articles...

### Trods protester udnævnes Babis til tjekkisk premierminister

Articles: 13 | Last update: Jun 27, 2018 11:24:00 AM | Start: Jun 27, 2018 10:17:00 AM | Sources: 13 | Peak: 6 | Current rank: 4

Trods protester udnævnes Babis til tjekkisk premierminister

en Wednesday, June 27, 2018 10:24:00 AM CEST | info | en | other

Babis, som leder centrumpartiet ANO, har forhandlet om en koalition med socialdemokraterne. Der skete et genembrud fredag, efter at partiet indgik et kompromis i en række personspørgsmål. I januar tabt Babis' regering en mistillidsafstemning i parlamentet. Alligevel har Babis sidet som fungerende premierminister siden....

More articles...

### Tools

Wednesday, June 27, 2018 11:49:00 AM CEST

RSS MAP

Facebook

subscribe | manage

info

Available on the App Store | Get it on Google play

### Languages

Select top stories in other languages.

en es cs da de el

en es et fr hr it

hu lt lv mt nl

pl pt ro ru sk sl

sv sw tr zh

Show additional languages

Interface: en - English

### Country Watch

The country most in the news at the moment.

Denmark (DK)

### News about Denmark

News sources from Denmark

### European Commission

Balance shift in SE Europe - Bulgaria wants the Burban natural gas, the EU is worried

en Wednesday, June 27, 2018 11:43:00

France summons US ambassador after spying revelations

321 TOTAL ARTICLES

Conflict

264 articles | 62 languages





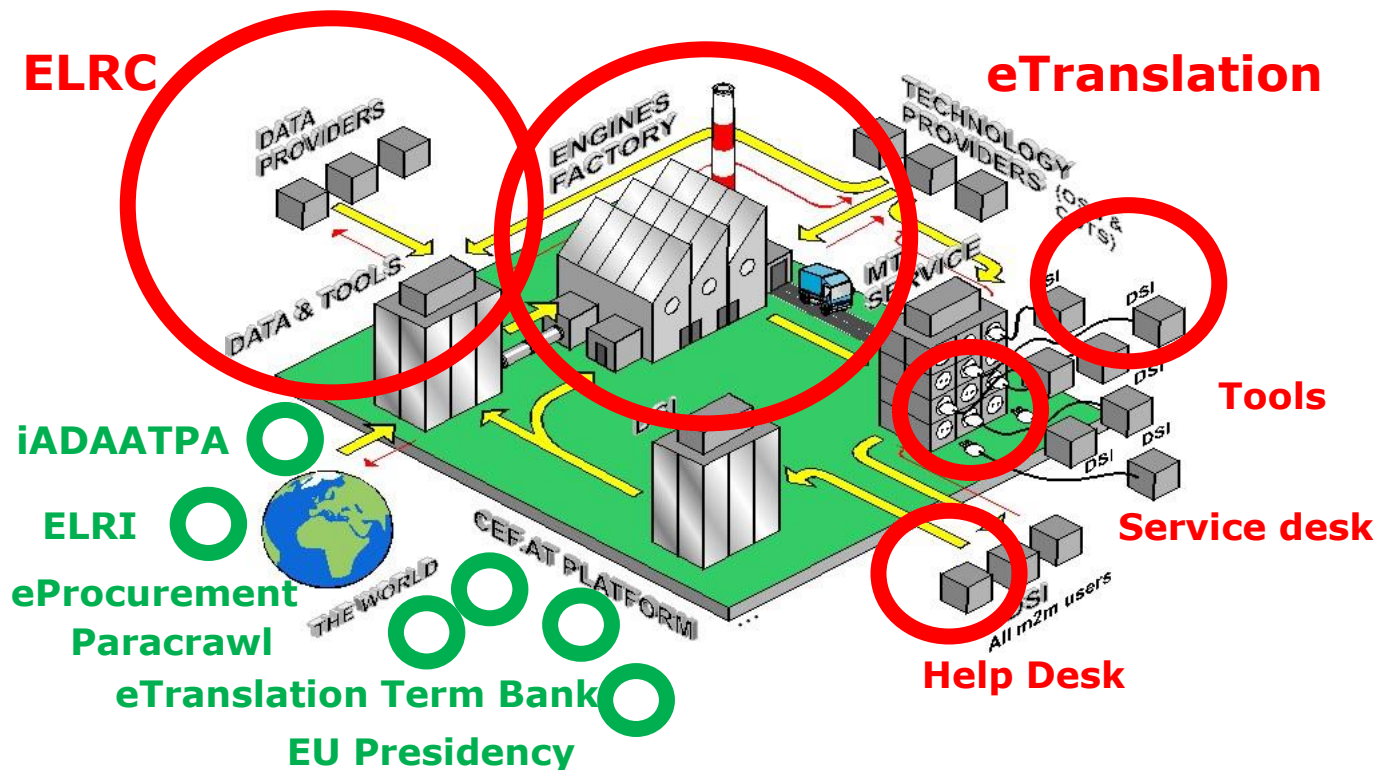
# *Examples*

---

# *Generic Services*

---

# CEF AT – Core and Generic Services



---

## Cross-border eProcurement notifications



- The interconnection of the **Croatian, Slovenian and Slovakian** national **eProcurement** systems to the Automated Translation (AT) Building Block for automatic translation of tendering information.
- The new architectures was programmed, tested and integrated into the national systems of the participating countries.
- As a direct result, each national platform was able to **publish tender notices and information** automatically translated into their respective languages from all of the partners' countries on their own user platform.

---

## CEF Automated Translation for the EU Council Presidency

Integration  
Projects

- Enable multilingual communication at the Presidency of the Council of the EU by integrating the CEF Automated Translation platform (AT platform) into a **Translation Toolkit** provided to the hosting countries of the EU Presidency events in 2017-2018 and beyond.
- Various digital environments – on mobile phones, on computer desktops, on the web, in the EU Council Presidency websites, and in translation productivity (CAT) tools.
- A set of highly customized Machine Translation (MT) engines for the EU Council Presidency domain and the official languages of the hosting countries in 2017-2018 (Estonia, Bulgaria and Austria).

---

# IADAATPA (Intelligent Automated Domain Adapted Automated Translation for Public Administrations)

Integration  
Projects

- The IADAATPA platform will include “**intelligent domain adaptation**” ensuring that translation requests are ‘routed’ to the most appropriate and relevant Machine Translation (MT) engine irrespective of the technology employed in building it.
- It will also provide a **Dashboard style Control Panel** displaying performance and routing statistics to validate the best association of MT engine and content type.
- As a result, the IADAATPA platform can be deployed across other digital services, by **other MT providers and technology developers.**

---

# APE-QUEST

## Automated Post-Editing (APE) & Quality Estimation (QE)

Integration  
Projects

- For Electronic Exchange of Social Security Information (EESSI) and Online Dispute Resolution (ODR).
- Quality Estimation to distinguish between high-quality, medium-quality, and low-quality MT output, as delivered by eTranslation
- Automated Post Editing on medium-quality MT output
- Applying the Unbabel PE workflow on low-quality output
- Applying QE on the result of this workflow to inform downstream applications.
- Covering English, French, Portuguese and Dutch languages.

---

## MICE

Integration  
Projects

- **Customisation** and improvement of CEF eTranslation Machine Translation (MT) output in specific domains.
- Automated translation of standards **and e-Business/e-Land register** information in **Belgium and Estonia**, respectively.
- MICE will provide a **software platform** for public services to use CEF eTranslation and customize it to their needs. The software will be made available as **Open Source package** that can be connected to the MT customization platforms, but also to other Open Source NLP frameworks.
- **Post-Editing** connectivity will be provided as part of the Action, so users can hook up the platform to the editing environment of their choice.

# NEC-TM Data: National European Central Translation Memory Data



- Setting up of a pan-European platform for the **sharing of translation memories** (TMs)
- Create a pan-European **data-sharing awareness campaign**
- Initially engaging national language programmes and data collection efforts in **Spain and Latvia, and national authorities in Croatia.**

• <http://www.nec-tm.eu/>

*(2017-EU-IA-0149)*

*Oct 2018 – Sept. 2020*

*€805,361*





---

## Multilingual resources for CEF AT in the legal domain



- Automatic translation on the body of national legislation in seven countries: **Bulgaria, Croatia, Hungary, Poland, Romania, Slovakia and Slovenia.**
- The project processes two resources available in all languages concerned
  - the multilingual ontology-based thesaurus **EUROVOC**
  - the corpora of all national legislation in the respective languages on the other.
- The project will result in valuable language corpora for the **legal domain**

---

## (Paracrawl I) Provision of Web-Scale Parallel Corpora for Official European Languages (Paracrawl I)



- Develop parallel corpora (collections of translated text) for all official EU languages, through **web crawling**.
- The **same type of large scale data** that is available to large commercial Machine Translation (MT) engines.
- For **8 languages** (i.e. English, German, Spanish, French, Polish, Italian, Portuguese, and Czech) the parallel corpora will have **more than 1 billion tokens**.
- For the **16 other languages**, the action aims to collect **more than 100 million tokens**.

---

## (Paracrawl II) Broader Web-Scale Provision of Parallel Corpora for European Languages



- Collect translated sentences from the web for all **24 official EU languages plus Icelandic, Norwegian, Basque, Catalan/Valencian, and Galician.**
- The Action builds on the CEF funded Action 2016-EU\_IA-0114.
- The resulting parallel corpora will be made **freely available.**
- To aid in customising machine translation systems, **a free software tool will perform domain filtering and weighting of the corpus.** The tool will accept examples of in-domain data provided by the user and filter the large corpus by relevance to that domain.

(2017-EU-IA-0178)

Sept. 2018 – Sept. 2020

€907,976



- Build **Terminological Resources (TRs)** to the language resource pipeline of the CEF's Automated Translation (AT) platform for the three sector-specific domains of **health, business legislation, and customer protection** to support the CEF DSIs for: eHealth, e-Justice, and Online Dispute Resolution (ODR).
- A minimum of 150 TRs covering all official languages of the EU plus Norwegian and Icelandic will be provided by the end of the Action, comprising an already existing baseline list.

---

# European Language Resource Infrastructure (ELRI)



- The provision of an infrastructure to help **collect, prepare and share language resources** that can in turn improve translation services.
- Providing **national public relay stations** where resources remain under member states' laws and regulations until further clearance is negotiated and granted.
- ELRI targets resources that are relevant to **Digital Service Infrastructures** and involves public institutions and public translation centres in **France, Ireland, Portugal and Spain** as a first step, with a future extension to additional member states as a key objective.



***Examples***

---

***Research***

---

---

# GoURMET – Global Under-Resources Media Translation



- Development of a high-quality machine translation for **low-resource language pairs and domains**
- Significantly improve the robustness and applicability of neural machine translation
- Development of **tools for media analysts and journalists**;
- Sustainable, maintainable platform and services

<https://gourmet-project.eu/project-info/>

*(ICT-29-2018-825299)*

*Jan. 2019 – Dec. 2021*

*€ 2,906,099*



---

# Bergamot – Browser-based Multilingual Translation



- **Client-side machine translation** in a web browser such as Mozilla Firefox.
- **Preserve citizens privacy**
- Increases the uptake of language technologies in Europe in various sectors that require confidentiality.
- Expose **quality estimates**.
- Enable users to confidently generate text in a language they do not speak,
- Enabling **cross-lingual online form filling**.
- **Adapting translation** on the fly using local information too private to upload the cloud.
- <https://browser.mt/>

*(ICT-29-2018-825303)*

*Jan. 2019 – Dec. 2021*

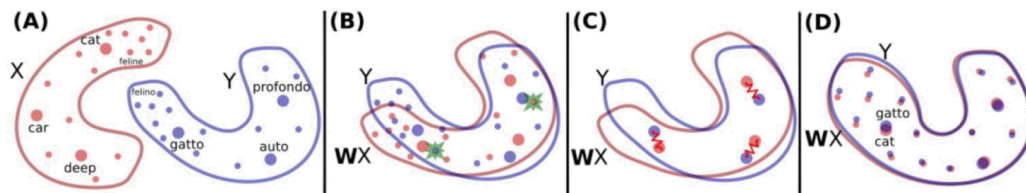
*€ 2,999,096*



# EMBEDDIA – Cross-Lingual Embeddings for Less-represented Languages in European News Media



- Creation of new tools to allow high quality **transformations** (not translations) between languages
- Use **cross-lingual embeddings**



- Coupled with deep neural networks to **allow existing monolingual resources** to be used across languages
- **Near real-time applications**, without the need for large computational resources

<http://embeddia.eu/>

---

# ELITR – European Live Translator



- **Automatic subtitling system** of live meetings and conference presentations
- Design and implement **automatic minuting**, i.e. create structured summaries from automatic transcript of a discussion.
- Research towards providing of **spoken language translation** (interpreting).
- Advance the state of the art to deliver machine translation of high-quality at the level of whole documents and scale up to multilingual settings with dozens of source and/or target languages.
- The technologies will be tested by the **Supreme Audit Office of the Czech Republic** and **alfaview®**, a German online conferencing system.

<https://elitr.eu/>

# COMPRISE - Cost-effective, Multilingual, Privacy-driven voice-enabled Services



- **Private-by-design** methodology and tools for **voice adaptation technology**
- **Anonymisation** of speech (removing name, address,...)
- Software development kit interoperating with a cloud-based resource platform: **smart consumer apps, e-commerce, and e-health.**
- Allow virtually unlimited collection of real-life non-private quality speech and language data;
- Enable businesses to quickly **develop multilingual voice-enabled services** in many languages;
- Allow all **citizens to transparently access contents and services available in other languages by voice interaction in their own language**; result in cost savings for both technology providers and users.

<http://www.comprise2020.eu/>

*(ICT-29-2018-825081)*

*Dec. 2018 – Nov. 2021*

*€ 3,201,016*



EN

## Eric Peters Speech

Statement on the Commission's work on language technologies at the EP Conference on Language Equality in the Digital Age

Introductory remarks

- Good afternoon, Ladies and Gentlemen,
- I am Eric Peters, senior advisor in the Cabinet of Mariya Gabriel, the European Commissioner for Digital Economy and Society.
- Let me begin by thanking Mrs. EVANS, the rapporteur, for her very good report on Language Equality in the Digital Age, as well as Jordi Solé and Ana Miranda for having organised the event and invited me to be with you today on behalf of the Commissioner.
- **In a European Union founded on the motto 'United in diversity', respect for linguistic diversity is a fundamental value that the Union must respect.** The crucial importance of language technologies in today's digital environment is undisputed.
- The European Commission therefore shares the view of the European Parliament that overcoming language barriers in the digital environment is essential for an inclusive society, a vibrant digital single market, and for our unity in diversity. It also agrees on the nature of the obstacles hampering language equality identified in this report.

"FR"

## Discours de M. Eric Peters

Déclaration sur les travaux de la Commission sur les technologies linguistiques lors de la conférence du Parlement européen sur l'égalité des langues à l'ère numérique

Remarques liminaires

- Mesdames et Messieurs,
- Je suis Eric Peters, conseiller principal au sein du cabinet de Mariya Gabriel, commissaire européenne pour l'économie et la société numériques.
- Permettez-moi de commencer par remercier Mme EVANS, rapporteure pour son très bon rapport sur l'égalité des langues à l'ère numérique, ainsi que M. Jordi Solé et Ana Miranda, qui ont organisé la manifestation et m'ont invité à être avec vous aujourd'hui au nom du commissaire.
- **Dans une Union européenne fondée sur la devise «Unie dans la diversité», le respect de la diversité linguistique est une valeur fondamentale que l'Union doit respecter.** L'importance cruciale des technologies linguistiques dans l'environnement numérique actuel n'est pas contestée.
- La Commission européenne partage donc l'avis du Parlement européen selon lequel la suppression des barrières linguistiques dans l'environnement numérique est essentielle pour une société inclusive, un marché unique numérique dynamique et pour notre unité dans la diversité. Elle convient également de la nature des obstacles qui entravent l'égalité entre les langues identifiées dans le présent rapport.

---

## A few ideas....

- Speech to Speech Translation
- Explainability  
(to get an inside of the Blackbox)
- Multitask learning Same model -> multiple tasks  
( Translation+ Q&A + clustering +...)
- Multilingual model
- UnBias models. (Gender,...)



# ***Ways Forward***

---

## ***From Translation to Language Technologies***

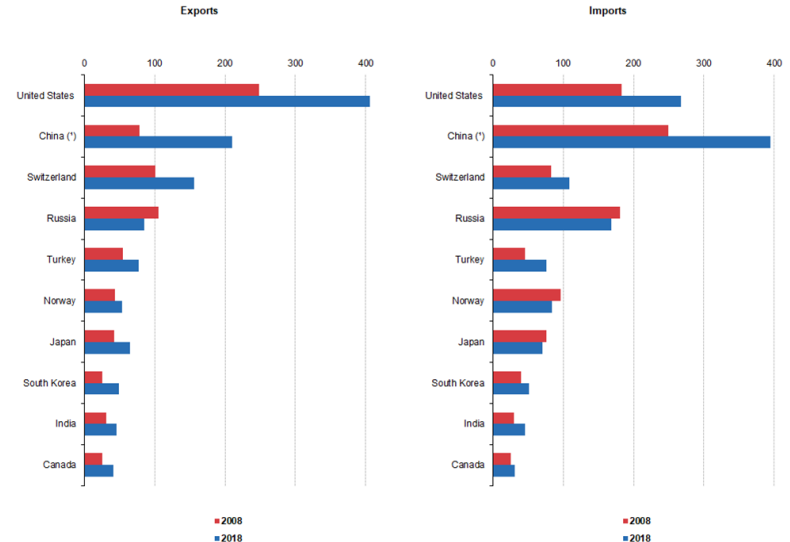
---

# Extending language coverage

eTranslation covers 26 languages: the official 24 EU languages + Icelandic and Norwegian

- Extension with economically relevant languages
  - **Chinese**
  - **Russian**
  - **Japanese**
  - **Korean**
- Extension with socially relevant non EU indigenous language
  - **Arabic**
  - **Turkish**
  - **Ukrainian**
- Extension with socially relevant EU indigenous languages

Extra EU-28 trade in goods by main trading partners, EU-28, 2008 and 2018  
(billion EUR)



Note: partners are sorted according to the sum of imports and exports in 2017.  
(\*) Excluding Hong Kong.  
Source: Eurostat (online data code: ext\_it\_maineu)

eurostat

CEF Automated Translation Survey

<https://ec.europa.eu/eusurvey/runner/CEFAutomatedTranslationPlatformSurvey2019>

---

# Extending domain coverage

- Generic domain engines
- Other possible domains
  - **Science and Technology**
  - Intellectual property and trademarks
  - Chemistry and pharmaceutical
  - Medicine and healthcare
  - Judicial and law enforcement
  - Agricultural, fisheries and food
  - Business
  - Economy, finance and tax
  - Social media interaction (messages, blogs, memes)
  - Customer support and consumer protection
  - Transport and travel
  - Culture, education and youth
  - Climate action
  - Energy and natural resources



---

# Additional applications

## Wrappers of the Automatic Translation services

- Websites Automatic Translation
- Easy to use Human Computer Assisted Translation (CAT) tool
  - including translation tools for social media

## Text Analysis with privacy focus

- Anonymization tools
  - including Named Entity Recognition
- Document classification
- Disinformation automatic detection

## Speech enabled services

- Speech transcription
- Subtitling
- Speech to speech translation



***Quality***

—  
***Try it.***  
—

# Achieving Human Parity on Automatic Chinese to English News Translation

Hany Hassan, Anthony Aue, Chang Chen, Vishal Chowdhary, Jonathan Clark, Christian Federmann, Xuedong Huang, Marcin Junczys-Dowmunt, William Lewis, Mu Li, Shujie Liu, Tie-Yan Liu, Renqian Luo, Arul Menezes, Tao Qin, Frank Seide, Xu Tan, Fei Tian, Lijun Wu, Shuangzhi Wu, Yingce Xia, Dongdong Zhang, Zhirui Zhang, and Ming Zhou

Microsoft AI & Research

## Abstract

Machine translation has made rapid advances in recent years. Millions of people are using it today in online translation systems and mobile applications in order to communicate across language barriers. The question naturally arises whether such systems can approach or achieve parity with human translations. In this paper, we first address the problem of how to define and accurately measure human parity in translation. We then describe Microsoft's machine translation system and measure the quality of its translations on the widely used WMT 2017 news translation task from Chinese to English. We find that our latest neural machine translation system has reached a new state-of-the-art, and that the translation quality is at human parity when compared to professional human translations. We also find that it significantly exceeds the quality of crowd-sourced non-professional translations.

## 1 Introduction

Recent years have seen human performance levels reached or surpassed in tasks ranging from games such as Go [32] to classification of images in ImageNet [20] to conversational speech recognition on the Switchboard task [49].

In the area of machine translation, we have seen dramatic improvements in quality with the advent of attentional encoder-decoder neural networks [34, 3, 38]. However, translation quality continues to vary a great deal across language pairs, domains, and genres, more or less in direct relationship to the availability of training data. This paper summarizes how we achieved human parity in translating text in the news domain, from Chinese to English. While the techniques we used are not specific to the news domain or the Chinese-English language pair, we do not claim that this result necessarily generalizes to other language pairs and domains, especially where limited by the availability of data and resources.

Translation of news text has been an area of active interest in the Machine Translation community for over a decade, due to the practical and commercial importance of this domain, the availability of abundant parallel data on the web (at least in the most popular languages) and a long history of government-funded projects and evaluation campaigns, such as NIST-OpenMT<sup>1</sup> and GALE<sup>2</sup>. The annual evaluation campaign of the WMT (Conference on Machine Translation) [6], has also focused on news translation for more than a decade.

Defining and measuring human quality in translation is challenging for a number of reasons. Traditional metrics of translation quality, such as BLEU [28], TER [33] and METEOR [10] measure translation quality by comparison with one or more human reference translations. However,

# Parvenir à la parité humaine au sujet de la traduction automatique en anglais vers l'anglais

Hany Hassan, Anthony Aue, Chang Chen, Vishal Chowdhary, Jonathan Clark, Christian Federmann, Xuedong Huang, Marcin Junczys-Dowmunt, William Lewis, Mu Li, Shujie Liu, Renqian Luo, Arul Menezes, Tao Qin, Frank Seide, Xu Tan, Fei Tian, Lijun Wu, Shuangzhi Wu, Yingce Xia, Dongdong Zhang, Zhirui Zhang et Ming Zhou

Microsoft AI & Recherche

## Résumé

La traduction automatique a rapidement progressé ces dernières années. Des millions de personnes y utilisent aujourd'hui des systèmes de traduction en ligne et des applications mobiles pour faire connaître les barrières linguistiques. La question se pose naturellement de savoir si ces systèmes peuvent approcher ou atteindre la parité avec des traductions humaines. Dans le présent document, nous aborderons tout d'abord le problème de la manière de définir et de mesurer avec précision la parité humaine dans la traduction. Nous décrivons ensuite le système de traduction automatique de Microsoft et mesurons la qualité de ses traductions sur le produit de la traduction de l'actualité (WMT 2017 de chinois vers l'anglais). Nous constatons que notre dernier système de combinaison de la machine neuronale a atteint un état de pointe nouveau et que la qualité de la traduction est atteinte à la parité humaine par rapport aux traductions humaines professionnelles. Nous constatons également qu'il dépasse de manière significative la qualité des traductions non professionnelles provenant de l'externalisation ouverte (crowd-sourcing).

## 1 Introduction

Ces dernières années ont vu des niveaux de performance humaine atteints ou dépassés dans les tâches allant de jeux tels que Go [32] à la classification des images dans ImageNet [20] pour la reconnaissance de la parole à la conversation dans la tâche standard [49].

Dans le domaine de la traduction automatique, nous avons connu une amélioration spectaculaire de la qualité avec l'arrivée de réseaux neuronaux attentionnés [34, 3, 38]. Toutefois, la qualité de la traduction continue de varier considérablement entre les paires de langues, les domaines et les genres, plus ou moins en relation directe avec la disponibilité des données de formation. Le présent document résume la manière dont nous avons obtenu la parité humaine pour traduire du texte dans le domaine de l'actualité, du chinois à l'anglais. Bien que les techniques que nous utilisons ne soient pas spécifiques au domaine de l'information ou à la combinaison linguistique chinoise, nous ne prétendons pas que ce résultat généralise nécessairement les autres paires de langues et domaines, en particulier lorsque la disponibilité des données et des ressources est limitée.

La traduction de textes d'actualité a été un domaine d'intérêt actif dans la sphère de traduction

Translation quality outside Legal domain

---

# Demo links

Website translation: <http://www.lr-coordination.eu/>

eTranslation snippets:

<https://webgate.ec.europa.eu/etranslation/translateTextSnippet.html>

[N-LEX:](#)

Bulgarian Presidency: <https://eu2018bg.bg/en/translation>

List of CEF projects:

<https://ec.europa.eu/inea/en/connecting-europe-facility/cef-telecom/projects-by-dsi/etranslation>

## Statement on the Commission's work on language technologies at the EP Conference on Language Equality in the Digital Age

### Introductory remarks

- Good afternoon, Ladies and Gentlemen,
- I am Eric Peters, senior advisor in the Cabinet of Mariya Gabriel, the European Commissioner for Digital Economy and Society.
- Let me begin by thanking Mrs. EVANS, the rapporteur, for her very good report on Language Equality in the Digital Age, as well as Jordi Solé and Ana Miranda for having organised the event and invited me to be with you today on behalf of the Commissioner.
- In a European Union founded on the motto 'United in diversity', respect for linguistic diversity is a fundamental value that the Union must respect. The crucial importance of language technologies in today's digital environment is undisputed.
- The European Commission therefore shares the view of the European Parliament that overcoming language barriers in the digital environment is essential for an inclusive society, a vibrant digital single market, and for our unity in diversity. It also agrees on the nature of the obstacles hampering language equality identified in this report.

## Déclaration sur les travaux de la Commission sur les technologies linguistiques lors de la conférence du Parlement européen sur l'égalité des langues à l'ère numérique

### Remarques liminaires

- Mesdames et Messieurs,
- Je suis Eric Peters, conseiller principal au sein du cabinet de Mariya Gabriel, commissaire européenne pour l'économie et la société numériques.
- Permettez-moi de commencer par remercier Mme EVANS, rapporteure pour son très bon rapport sur l'égalité des langues à l'ère numérique, ainsi que M. Jordi Solé et Ana Miranda, qui ont organisé la manifestation et m'ont invité à être avec vous aujourd'hui au nom du commissaire.
- Dans une Union européenne fondée sur la devise «Unie dans la diversité», le respect de la diversité linguistique est une valeur fondamentale que l'Union doit respecter. L'importance cruciale des technologies linguistiques dans l'environnement numérique actuel n'est pas contestée.
- La Commission européenne partage donc l'avis du Parlement européen selon lequel la suppression des barrières linguistiques dans l'environnement numérique est essentielle pour une société inclusive, un marché unique numérique dynamique et pour notre unité dans la diversité. Elle convient également de la nature des obstacles qui entravent l'égalité entre les langues identifiées dans le présent rapport.

Statement on the Commission's work on language technologies at the EP Conference on Language Equality in the Digital Age

#### Introductory remarks

- Good afternoon, Ladies and Gentlemen,
- I am Eric Peters, senior advisor in the Cabinet of Mariya Gabriel, the European Commissioner for Digital Economy and Society.
- Let me begin by thanking Mrs. EVANS, the rapporteur, for her very good report on Language Equality in the Digital Age, as well as Jordi Solé and Ana Miranda for having organised the event and invited me to be with you today on behalf of the Commissioner.
- In a European Union founded on the motto 'United in diversity', respect for linguistic diversity is a fundamental value that the Union must respect. The crucial importance of language technologies in today's digital environment is undisputed.
- The European Commission therefore shares the view of the European Parliament that overcoming language barriers in the digital environment is essential for an inclusive society, a vibrant digital single market, and for our unity in diversity. It also agrees on the nature of the obstacles hampering language equality identified in this report.

Erklärung zu den Arbeiten der Kommission zu Sprachtechnologien auf der Konferenz des EP zur Gleichstellung der Sprachen im digitalen Zeitalter

#### Einleitende Bemerkungen

Guten Nachmittag, meine Damen und Herren,

- Ich bin Eric Peters, leitender Berater im Kabinett von Mariya Gabriel, für die digitale Wirtschaft und Gesellschaft zuständiges Mitglied der Europäischen Kommission.
- Wir danken Frau EVANS, der Berichterstatterin für ihren sehr guten Bericht über die Gleichstellung der Sprachen im digitalen Zeitalter, sowie Jordi Solé und Ana Miranda für die Organisation der Veranstaltung und bitten mich, heute im Namen des Kommissars zu sein.
- In einer Europäischen Union, die sich auf das Motto „In Vielfalt geeint“ gründet, ist die Achtung der sprachlichen Vielfalt ein Grundwert, den die Union achten muss. Die entscheidende Bedeutung der Sprachtechnologien in der heutigen digitalen Umgebung ist unbestritten.
- Die Europäische Kommission teilt daher die Auffassung des Europäischen Parlaments, dass die Überwindung von Sprachbarrieren im digitalen Umfeld für eine inklusive Gesellschaft, einen lebendigen digitalen Binnenmarkt und unsere Einheit in Vielfalt von wesentlicher Bedeutung ist. Er stimmt auch der Art der Hindernisse zu, die die sprachliche Gleichstellung behindern, wie sie in diesem Bericht aufgezeigt werden.

Statement on the Commission's work on language technologies at the EP Conference on Language Equality in the Digital Age

#### Introductory remarks

- Good afternoon, Ladies and Gentlemen,
- I am Eric Peters, senior advisor in the Cabinet of Mariya Gabriel, the European Commissioner for Digital Economy and Society.
- Let me begin by thanking Mrs. EVANS, the rapporteur, for her very good report on Language Equality in the Digital Age, as well as Jordi Solé and Ana Miranda for having organised the event and invited me to be with you today on behalf of the Commissioner.
- In a European Union founded on the motto 'United in diversity', respect for linguistic diversity is a fundamental value that the Union must respect. The crucial importance of language technologies in today's digital environment is undisputed.
- The European Commission therefore shares the view of the European Parliament that overcoming language barriers in the digital environment is essential for an inclusive society, a vibrant digital single market, and for our unity in diversity. It also agrees on the nature of the obstacles hampering language equality identified in this report.

Declaración sobre el trabajo de la Comisión en materia de tecnologías del lenguaje en la Conferencia del PE sobre la igualdad lingüística en la era digital

#### Observaciones introductorias

- Buenas tardes, señoras y señores:

Soy Eric Peters, asesor principal del Gabinete de la comisaria Mariya Gabriel, la comisaria europea de Economía y Sociedad Digitales.

- comenzamos a dar las gracias a la Sra. EVANS, ponente, por su excelente informe sobre la igualdad lingüística en la era digital, así como a Jordi Solé y Ana Miranda por haber organizado el acto y a invitar a mí en el día de hoy en nombre del Comisario.
- En una Unión Europea basada en el lema «Unida en la diversidad», el respeto de la diversidad lingüística es un valor fundamental que la Unión debe respetar. La importancia crucial de las tecnologías del lenguaje en el entorno digital actual es indiscutible.

Por consiguiente, la Comisión Europea comparte la opinión del Parlamento Europeo de que la superación de las barreras lingüísticas en el entorno digital es esencial para una sociedad inclusiva, un mercado único digital dinámico y nuestra unidad en la diversidad. También está de acuerdo con la naturaleza de los obstáculos que obstaculizan la igualdad lingüística que se señalan en el presente informe.

Statement on the Commission's work on language technologies at the EP Conference on Language Equality in the Digital Age

#### Introductory remarks

- Good afternoon, Ladies and Gentlemen,
- I am Eric Peters, senior advisor in the Cabinet of Mariya Gabriel, the European Commissioner for Digital Economy and Society.
- Let me begin by thanking Mrs. EVANS, the rapporteur, for her very good report on Language Equality in the Digital Age, as well as Jordi Solé and Ana Miranda for having organised the event and invited me to be with you today on behalf of the Commissioner.
- In a European Union founded on the motto 'United in diversity', respect for linguistic diversity is a fundamental value that the Union must respect. The crucial importance of language technologies in today's digital environment is undisputed.
- The European Commission therefore shares the view of the European Parliament that overcoming language barriers in the digital environment is essential for an inclusive society, a vibrant digital single market, and for our unity in diversity. It also agrees on the nature of the obstacles hampering language equality identified in this report.

Dichiarazione sull'attività della Commissione nel settore delle tecnologie del linguaggio in occasione della conferenza del PE sull'uguaglianza linguistica nell'era digitale

#### Osservazioni introduttive

- buongiorno, signore e signori,
- Sono Eric Peters, consulente di alto livello nel gabinetto di Mariya Gabriel, commissario europeo responsabile per l'Economia e la società digitali.
- iniziamo ringraziare l'on. EVANS, il relatore, per la sua ottima relazione sull'uguaglianza linguistica nell'era digitale, nonché Jordi Solé e Ana Miranda per aver organizzato l'evento, e mi ha invitato a partecipare in data odierna a nome del commissario.
- In un'Unione europea fondata sul motto "Unita nella diversità", il rispetto della diversità linguistica è un valore fondamentale che l'Unione deve rispettare. L'importanza cruciale delle tecnologie del linguaggio nell'ambiente digitale odierno è incontestabile.
- La Commissione europea condivide pertanto l'opinione del Parlamento europeo secondo cui il superamento delle barriere linguistiche nell'ambiente digitale è essenziale per una società inclusiva, un mercato unico digitale dinamico e per la nostra unità nella diversità. Concorda inoltre sulla natura degli ostacoli che ostacolano l'uguaglianza linguistica individuati nella presente relazione.



---

Close to perfect ?.... Well...

Statement on the Commission's work on language technologies at the EP Conference on Language Equality in the Digital Age

#### Introductory remarks

- Good afternoon, Ladies and Gentlemen,
- I am Eric Peters, senior advisor in the Cabinet of Mariya Gabriel, the European Commissioner for Digital Economy and Society.
- Let me begin by thanking Mrs. EVANS, the rapporteur, for her very good report on Language Equality in the Digital Age, as well as Jordi Solé and Ana Miranda for having organised the event and invited me to be with you today on behalf of the Commissioner.
- In a European Union founded on the motto 'United in diversity', respect for linguistic diversity is a fundamental value that the Union must respect. The crucial importance of language technologies in today's digital environment is undisputed.
- The European Commission therefore shares the view of the European Parliament that overcoming language barriers in the digital environment is essential for an inclusive society, a vibrant digital single market, and for our unity in diversity. It also agrees on the nature of the obstacles hampering language equality identified in this report.

Ráiteas ar obair an Choimisiúin maidir le teicneolaíochtaí teanga ag Comhdháil PE faoi Chomhionannas Teanga sa Ré Dhigiteach

#### Réamhrá

- Dea-iarnóin, a Dhaoine Uaisle,
- Is mise Eric Pecer, comhairleoir sinsearach sa Cabinet, mar atá Mariya Gabriel, an Coimisinéir Eorpach um an nGeilleagar Digiteach agus an tSochaí Dhigiteach.
- Cur tús leis an scéal seo, mar shampla, tá an rapóirtéir, an rapóirtéir, i gcomhair a tuairisce an-mhaith ar Chomhionannas Teanga sa Ré Dhigiteach, chomh maith le Jorí Solé agus Ana anda don imeacht a d'éagraigh sibh sa lá atá inniu ann, agus d'áiríodh iarr sí orm a bheith in éineacht leat inniu thar ceann an Choimisinéara.
- In Aontas Eorpach bunaithe ar an mana 'Aontaithe san Éagsúlacht', is luach bunúsach é urraim don éagsúlacht teanga nach mór don Aontas a urramú. Tá sé beartaithe gan amhras a bheith ann maidir le teicneolaíochtaí teanga i dtimpeallacht dhigiteach an lae inniu.
- Tá an Coimisiún Eorpach, dá bhrí sin, ar aon intinn le Parlaimint na hEorpa go bhfuil bacainní teanga sa timpeallacht dhigiteach ríthábhachtach do shochaí chuimsitheach, do mhargadh aonair digiteach bríomhar, agus don aontacht san éagsúlacht. Aontaíonn sé freisin leis an gcineál constaicí a chuireann bac ar an gcomhionannas teanga a shainiú sa tuarascáil seo.

"GA"

Óráid Eric Claair

Ráiteas ar obair an Choimisiúin maidir le teicneolaíochtaí teanga ag Comhdháil PE faoi Chomhionannas Teanga sa Ré Dhigiteach

Réamhrá

- Dea-iarnóin, a Dhaoine Uaisle,
- Is mise Eric Pecер, comhairleoir sinsearach sa Cabinet, mar atá Mariya Gabriel, an Coimisinéir Eorpach um an nGeilleagar Digiteach agus an tSochaí Dhigiteach.
- Cur tús leis an scéal seo, mar shampla, tá an rapóirtéir, an rapóirtéir, i gcomhair a tuairisce an-mhaith ar Chomhionannas Teanga sa Ré Dhigiteach, chomh maith le Jorí Solé agus Ana anda don imeacht a d’eagraigh sibh sa lá atá inniu ann, agus d’&apos; iarr sí orm a bheith in éineacht leat inniu thar ceann an Choimisinéara.
- In Aontas Eorpach bunaithe ar an mana ‘Aontaithe san Éagsúlacht’, is luach bunúsach é urraim don éagsúlacht teanga nach mór don Aontas a urramú. Tá sé beartaithe gan amhras a bheith ann maidir le teicneolaíochtaí teanga i dtimpeallacht dhigiteach an lae inniu.
- Tá an Coimisiún Eorpach, dá bhrí sin, ar aon intinn le Parlaimint na hEorpa go bhfuil bacainní teanga sa timpeallacht dhigiteach ríthábhachtach do shochaí chuimsitheach, do mhargadh aonair digiteach bríomhar, agus don aontacht san éagsúlacht. Aontaíonn sé freisin leis an gcineál constaicí a chuireann bac ar an gcomhionannas teanga a shainaithnítear sa tuarascáil seo.

“EN”

Address (if applicable)

Statement on the work of the Commission on language technologies for the EP Conference on Language Equality within the Digital Age

Introduction

- Good afternoon, ladies and Gentlemen,
- I’m Eric Peer, senior adviser in Cabinet, as Mariya Gabriel, European Commissioner for Digital Economy and Society.
- Start this story, for example, your rapporteur, the rapporteur, is for a very good report on Language Equality within the Digital Age, as well as his personnel for your event today, and to your event’ Ask me to be accompanied by request on behalf of the Commissioner.
- In a European Union based in the motto of ‘United in Diversity’, respect for linguistic diversity is a fundamental value that the Union must respect. The creation of language technologies in this modern digital environment is today’s digital environment.
- The European Commission is, therefore, in the sense of the European Commission that in the digital environment, language barriers in the digital environment are essential to a inclusive digital single market, and the unity in diversity. It also agrees with the type of obstacle which hinders the identity of language identified in this report.

## Statement on the Commission's work on language technologies at the EP Conference on Language Equality in the Digital Age

### Introductory remarks

- Good afternoon, Ladies and Gentlemen,
- I am Eric Peters, senior advisor in the Cabinet of Mariya Gabriel, the European Commissioner for Digital Economy and Society.
- Let me begin by thanking Mrs. EVANS, the rapporteur, for her very good report on Language Equality in the Digital Age, as well as Jordi Solé and Ana Miranda for having organised the event and invited me to be with you today on behalf of the Commissioner.
- In a European Union founded on the motto 'United in diversity', respect for linguistic diversity is a fundamental value that the Union must respect. The crucial importance of language technologies in today's digital environment is undisputed.
- The European Commission therefore shares the view of the European Parliament that overcoming language barriers in the digital environment is essential for an inclusive society, a vibrant digital single market, and for our unity in diversity. It also agrees on the nature of the obstacles hampering language equality identified in this report.

## Verklaring over het werk van de Commissie inzake taaltechnologieën op de conferentie van het Europees Parlement over de gelijkheid van talen in het digitale tijdperk

### Inleidende opmerkingen

- Goedemiddag, Ladies en heren,
- ik ben Eric Peters, senior adviseur in het kabinet van Mariya Gabriel, Europees commissaris voor Digitale Economie en Samenleving.
- Laat mij beginnen met een dankwoord aan mevrouw EVANS, de rapporteur, voor haar zeer goed verslag over de gelijkheid van talen in het digitale tijdperk, en Jordi Solé en Ana Miranda voor het organiseren van het evenement en verzocht mij vandaag namens de commissaris mee te praten.
- In een Europese Unie, gebaseerd op het devies „In verscheidenheid verenigd“, is de eerbiediging van taalkundige verscheidenheid een fundamentele waarde die de Unie moet respecteren. Het cruciale belang van taaltechnologieën in de huidige digitale omgeving staat buiten kijf.
- De Europese Commissie deelt daarom het standpunt van het Europees Parlement dat het overwinnen van taalbarrières in de digitale omgeving van essentieel belang is voor een inclusieve samenleving, een dynamische digitale interne markt en voor onze eenheid in verscheidenheid. Zij is het ook eens over de aard van de belemmeringen die de gelijkheid van talen in dit verslag in de weg staan.

Verklaring over het werk van de Commissie inzake taaltechnologieën op de conferentie van het Europees Parlement over de gelijkheid van talen in het digitale tijdperk

### Inleidende opmerkingen

- Goedemiddag, Ladies en heren,
- ik ben Eric Peters, senior adviseur in het kabinet van Mariya Gabriel, Europees commissaris voor Digitale Economie en Samenleving.
- Laat mij beginnen met een dankwoord aan mevrouw EVANS, de rapporteur, voor haar zeer goed verslag over de gelijkheid van talen in het digitale tijdperk, en Jordi Solé en Ana Miranda voor het organiseren van het evenement en verzocht mij vandaag namens de commissaris mee te praten.
- In een Europese Unie, gebaseerd op het devies „In verscheidenheid verenigd“, is de eerbiediging van taalkundige verscheidenheid een fundamentele waarde die de Unie moet respecteren. Het cruciale belang van taaltechnologieën in de huidige digitale omgeving staat buiten kijf.
- De Europese Commissie deelt daarom het standpunt van het Europees Parlement dat het overwinnen van taalbarrières in de digitale omgeving van essentieel belang is voor een inclusieve samenleving, een dynamische digitale interne markt en voor onze eenheid in verscheidenheid. Zij is het ook eens over de aard van de belemmeringen die de gelijkheid van talen in dit verslag in de weg staan.

Statement on the Commission's work on language technologies at the European Parliament conference on language equality in the digital era

### Opening remarks

- Good afternoon, Ladies and Messrs.
- I am Eric Peters, Senior Advisor in the Cabinet of Mariya Gabriel, European Commissioner for Digital Economy and Society.
- Let me start by thanking Ms EVANS, the rapporteur, for her very good report on language equality in the digital age, and Jordi Solé and Ana Miranda for organising the event and asked me to speak on behalf of the Commissioner today.
- In a European Union, based on the motto "United in diversity", respect for linguistic diversity is a fundamental value that the Union must respect. The crucial importance of language technologies in the current digital environment is undisputed.

Therefore, the European Commission shares the view of the European Parliament that overcoming language barriers in the digital environment is essential for an inclusive society, a dynamic digital single market and our unity in diversity. It also agrees with the nature of the obstacles to the equality of languages in this report.

Statement on the Commission's work on language technologies at the EP Conference on Language Equality in the Digital Age

#### Introductory remarks

- Good afternoon, Ladies and Gentlemen,
- I am Eric Peters, senior advisor in the Cabinet of Mariya Gabriel, the European Commissioner for Digital Economy and Society.
- Let me begin by thanking Mrs. EVANS, the rapporteur, for her very good report on Language Equality in the Digital Age, as well as Jordi Solé and Ana Miranda for having organised the event and invited me to be with you today on behalf of the Commissioner.
- In a European Union founded on the motto 'United in diversity', respect for linguistic diversity is a fundamental value that the Union must respect. The crucial importance of language technologies in today's digital environment is undisputed.
- The European Commission therefore shares the view of the European Parliament that overcoming language barriers in the digital environment is essential for an inclusive society, a vibrant digital single market, and for our unity in diversity. It also agrees on the nature of the obstacles hampering language equality identified in this report.

Lausunto komission kieliteknologiaa koskevasta työstä Euroopan parlamentin kielten tasa-arvoa digitaalisella aikakaudella käsitelleessä EP:n konferenssissa

#### Alustavia huomioita

- Hyvä iltapäivä Hyvät naiset ja herrat.
- Eric Peters, Mariya Gabrielin kabinetin vanhempi neuvonantaja, digitaalitaloudesta ja -yhteiskunnasta vastaava Euroopan komission jäsen.
- Let aluksi kiittänyt Eva EVANSia, joka oli esittelijän mielestä hyvä raportti kielitentasa-arvosta digitaalisella aikakaudella, sekä Jordi Solé ja Ana Miranda, jotka olivat järjestäneet tapahtuman ja kutsuivat minut tänään sinuun komission jäsenen puolesta.
- EU:n tunnuslauseeseen "Moninaisuudessaan yhtenäinen" perustuvan Euroopan unionin kielellisen monimuotoisuuden kunnioittaminen on perustavaa laatua oleva arvo, jota unionin on kunnioitettava. Kieliteknologian ratkaiseva merkitys nykypäivän digitaalisessa ympäristössä on kiistaton.
- Euroopan komissio on siksi Euroopan parlamentin kanssa samaa mieltä siitä, että kieliesteiden voittaminen digitaalisessa ympäristössä on olennaisen tärkeää osallistavan yhteiskunnan, elinvoimaisten digitaalisten sisämarkkinoiden ja moninaisuudessaan yhtenäisyytemme kannalta. Komitea on samaa mieltä myös siitä, että tässä kertomuksessa yksilöidyt kielelliset esteet ovat luonteeltaan esteitä.

Lausunto komission kieliteknologiaa koskevasta työstä Euroopan parlamentin kielten tasa-arvoa digitaalisella aikakaudella käsitelleessä EP:n konferenssissa

Alustavia huomioita

- Hyvä iltapäivä Hyvät naiset ja herrat.
- Eric Peters, Mariya Gabrielin kabinetin vanhempi neuvonantaja, digitaalitaloudesta ja -yhteiskunnasta vastaava Euroopan komission jäsen.
- Let aluksi kiittänyt Eva EVANSia, joka oli esittelijän mielestä hyvä raportti kielientasa-arvosta digitaalisella aikakaudella, sekä Jordi Solé ja Ana Miranda, jotka olivat järjestäneet tapahtuman ja kutsuivat minut tänään sinuun komission jäsenen puolesta.
- EU:n tunnuslauseeseen "Moninaisuudessaan yhtenäinen" perustuvan Euroopan unionin kielellisen monimuotoisuuden kunnioittaminen on perustavaa laatua oleva arvo, jota unionin on kunnioitettava. Kieliteknologian ratkaiseva merkitys nykypäivän digitaalisessa ympäristössä on kiistaton.
- Euroopan komissio on siksi Euroopan parlamentin kanssa samaa mieltä siitä, että kieliesteiden voittaminen digitaalisessa ympäristössä on olennaisen tärkeää osallistavan yhteiskunnan, elinvoimaisten digitaalisten sisämarkkinoiden ja moninaisuudessaan yhtenäisyytemme kannalta. Komitea on samaa mieltä myös siitä, että tässä kertomuksessa yksilöidyt kielelliset esteet ovat luonteeltaan esteitä.

Opinion on the Commission's work on language technologies in the EP Conference on Equality of Languages in the Digital Age

Preliminary remarks

- Good afternoon, ladies and gentlemen.
- Senior Adviser, Commissioner for Digital Economy and Society, Private Office of the Cabinet of Ms Gabriel, Mariya Gabriel, Cabinet.
- You initially thanked Eva EVANSia, who, in the rapporteur's view, had a good report on the value of the negative information in the digital age, and Jordi Solé and Ana Miranda, who had organised the event and invited me to you today on behalf of the Commissioner.
- The respect of the linguistic diversity of the European Union under the EU motto "United in diversity" is a fundamental value to be respected by the Union. The crucial role of language technologies in today's digital environment is beyond doubt.
- The European Commission therefore agrees with the European Parliament that overcoming language barriers in the digital environment is essential for an inclusive society, a vibrant digital single market and our unity in diversity. The Committee also agrees that the linguistic barriers identified in this report are by nature barriers.

Statement on the Commission's work on language technologies at the EP Conference on Language Equality in the Digital Age

#### Introductory remarks

- Good afternoon, Ladies and Gentlemen,
- I am Eric Peters, senior advisor in the Cabinet of Mariya Gabriel, the European Commissioner for Digital Economy and Society.
- Let me begin by thanking Mrs. EVANS, the rapporteur, for her very good report on Language Equality in the Digital Age, as well as Jordi Solé and Ana Miranda for having organised the event and invited me to be with you today on behalf of the Commissioner.
- In a European Union founded on the motto 'United in diversity', respect for linguistic diversity is a fundamental value that the Union must respect. The crucial importance of language technologies in today's digital environment is undisputed.
- The European Commission therefore shares the view of the European Parliament that overcoming language barriers in the digital environment is essential for an inclusive society, a vibrant digital single market, and for our unity in diversity. It also agrees on the nature of the obstacles hampering language equality identified in this report.

A Bizottság nyelvi technológiákkal kapcsolatos munkájáról szóló nyilatkozat az EP „Digitális korban” című európai parlamenti konferencián

#### Bevezető megjegyzések

- Helyes délután, Hölgyeim és Uraim!
  - Eric Peters, Mariya Gabriel vezető tanácsadója, a digitális gazdaságért és társadalomért felelős európai biztos vagyok.
  - Elkezdem, hogy köszönetet mondok EVANS asszonynak, az előadónak, hogy nagyon jó jelentést készítsen a nyelvi egyenlőségről a digitális korban, valamint Jordi Solé és Ana Miranda azért, hogy megszervezte a rendezvényt, és felkérte, hogy a biztos képviselőjeként vegyen részt Önökkel.
- Az „Egyesülve a sokféleségben” jelmondaton alapuló Európai Unióban a nyelvi sokszínűség tiszteletben tartása olyan alapvető érték, amelyet az Uniónak tiszteletben kell tartania. A nyelvi technológiáknak a mai digitális környezetben betöltött kulcsfontosságú szerepe vitathatatlan.
- Az Európai Bizottság ezért osztja az Európai Parlament azon nézetét, hogy a nyelvi akadályok leküzdése a digitális környezetben alapvető fontosságú a befogadó társadalom, az élénk digitális egységes piac, valamint a sokszínűségben való egységünk szempontjából. Egyetért továbbá az e jelentésben meghatározott nyelvi egyenlőséget akadályozó tényezők jellegével.



A Bizottság nyelvi technológiákkal kapcsolatos munkájáról szóló nyilatkozat az EP „Digitális korban” című európai parlamenti konferencián

#### Bevezető megjegyzések

- Helyes délután, Hölgyeim és Uraim!
- Eric Peters, Mariya Gabriel vezető tanácsadója, a digitális gazdaságért és társadalomért felelős európai biztos vagyok.
- Elkezdem, hogy köszönetet mondok EVANS asszonynak, az előadónak, hogy nagyon jó jelentést készítse a nyelvi egyenlőségről a digitális korban, valamint Jordi Solé és Ana Miranda azért, hogy megszervezte a rendezvényt, és felkérte, hogy a biztos képviselőként vegyen részt Önökkel.

Az „Egyesülve a sokféleségben” jelmondaton alapuló Európai Unióban a nyelvi sokszínűség tiszteletben tartása olyan alapvető érték, amelyet az Uniónak tiszteletben kell tartania. A nyelvi technológiáknak a mai digitális környezetben betöltött kulcsfontosságú szerepe vitathatatlan.

Az Európai Bizottság ezért osztja az Európai Parlament azon nézetét, hogy a nyelvi akadályok leküzdése a digitális környezetben alapvető fontosságú a befogadó társadalom, az élénk digitális egységes piac, valamint a sokszínűségben való egységünk szempontjából. Egyetért továbbá az e jelentésben meghatározott nyelvi egyenlőséget akadályozó tényezők jellegével.

Statement by the EP on its work on language technologies at the EP conference on “The Digital Age”

#### Introductory remarks

- Good afternoon, ladies and gentlemen!
- Senior adviser to Mariya Gabriel, Senior Commissioner for the Digital Economy and Society.
- I would like to thank Ms EVANS, rapporteur, to draw up a very good report on linguistic equality in the digital age, and Jordi Solé and Ana Miranda to organise the event and invited him to represent the Commissioner.

Respect for linguistic diversity in the European Union, based on the motto ‘United in diversity’, is a fundamental value that the Union must respect. The key role of language technologies in today’s digital environment is undeniable.

The European Commission therefore shares the view of the European Parliament that overcoming language barriers in the digital environment is crucial for an inclusive society, a vibrant digital single market and our unity in diversity. It also agrees with the nature of the obstacles to language equality set out in this report.

---

# Exemples

## good terminology

- voice-activated controls
- commandes vocales
- sprachgesteuerten Kontrolle

## almost good coreference resolution

- Language Technologies .... They aim
- Les technologies linguistiques .... Elles visent
- Sprachtechnologien .... Sie dienen

## Good translation of idioms

- low hanging fruit
- fruits mûrs
- wenig hängendes Obst

## almost good terminology (langage/langue)

- natural language processing
- traitement du langue naturel
- Verarbeitung natürlicher Sprache

## Very good word order

- (connected) cars are being fitted
- voitures (connectées) sont équipées
- (vernetzte) Fahrzeuge ausgerüstet werden

## Good translation of abbreviations

- module de traduction automatique du CEF (Connecting Europe Facility)
- MIE (Le Mécanisme pour l'Interconnexion en Europe)
- des Blocks „Automated Translation“ der Fazilität „Connecting Europe“

# Links & contacts



## **PHILIPPE GELIN**

European Commission

DG CONNECT - Communications Network, Content and Technology

Dir G: Data

Unit G3: Accessibility, **Multilingualism** and Safer Internet

[Philippe.Gelin@ec.europa.eu](mailto:Philippe.Gelin@ec.europa.eu)