Language Resource Board

—

Amsterdam, March 2019

—

June Lowery-Kingston

Head of Unit Accessibility, Multilingualism & Safer Internet
Directorate-General for Communications Networks, Content and Technology
European Commission

MULTILINGUALISM IN MFF

in Horizon Europe	**in Digital Europe Programme**

STATE OF THE ART TECHNOLOGY

R&I priorities + DATA & Infrastructure to support R&I

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*

European Commission

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 and Secure Society

Cultural Heritage

European Commission

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

European Commission

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, European libraries of algorithms Digital Innovation Hubs

European Commission

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, SMEs.

Where ?

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

Modernisation of administrations

European Commission

eTranslation – Statistics

- Neural Machine Translation in the 24 official languages of the EU (*24<->24*)

- Secured & scalable

- Domain specific engines:
  - *Public health, culture, banking, IP Law, Case law, … (not yet all 24<->24)*

- 51 digital services connected

- Translation volume:
  - 19 Mio pages in 2017
  - 40 Mio pages in 2018

- Latest daily record:

  19/03/19: 635 158 translated pages

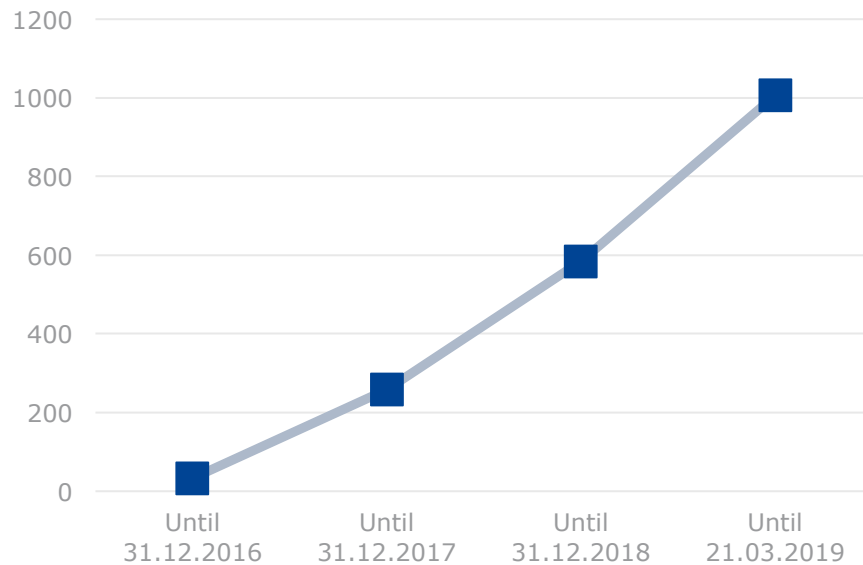 *20/03/19: 311 078 requests*
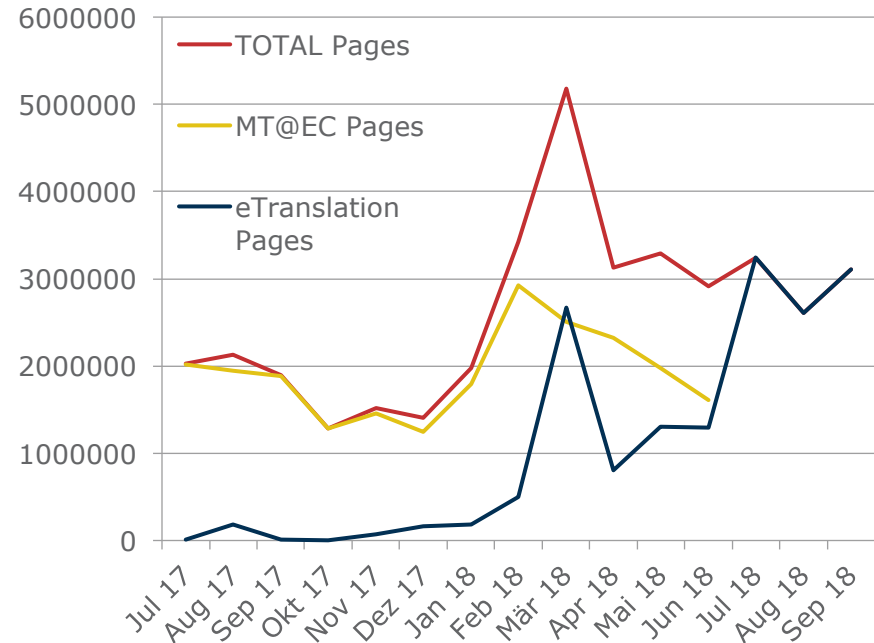
Now available to public administrations across Europe

European Language Resource Coordination

> *1000 language resources*
> **3 billion words**

No. of unique LR collected through ELRC-SHARE

1200	
1000	▪
800	
600	▪
400	
200	▪
0	▪

Until 31.12.2016   Until 31.12.2017   Until 31.12.2018   Until 21.03.2019

European Commission

LR collected within ELRC...

- were the main prerequisite for building EN<>IS and EN<>NB engines

- were used with a subset of EC data to build the new cultural domain DE engine

- contributed to the development of the EN<>GA engine (increase of BLEU scores)

- contributed to the development of the EUIPO engine

- are currently used for building engines for the WMT19 (EN<>DE, EN>LT and FR>DE)

- will be used for building general engines

- will be used by various other NLP tools (summarisation etc.)

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

European Commission
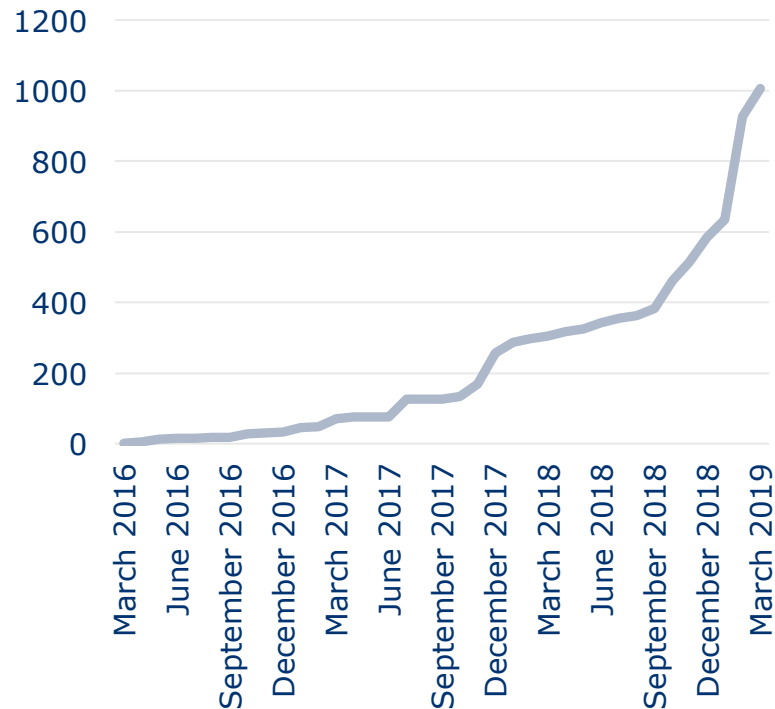
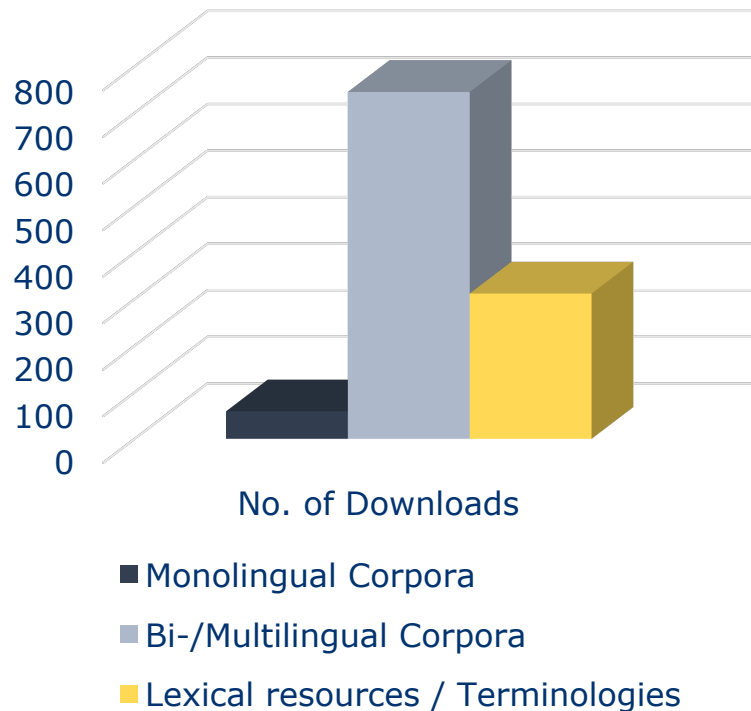# Requirements to build a Translation Model.

- A minimum of 2.000.000 TUs is needed to build translation engine

- At least 100.000 TUs are needed for domain adaptation (i.e. to improve the engine in a particular domain)"
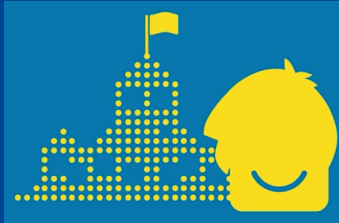

Resources

Language

Translation

European Commission
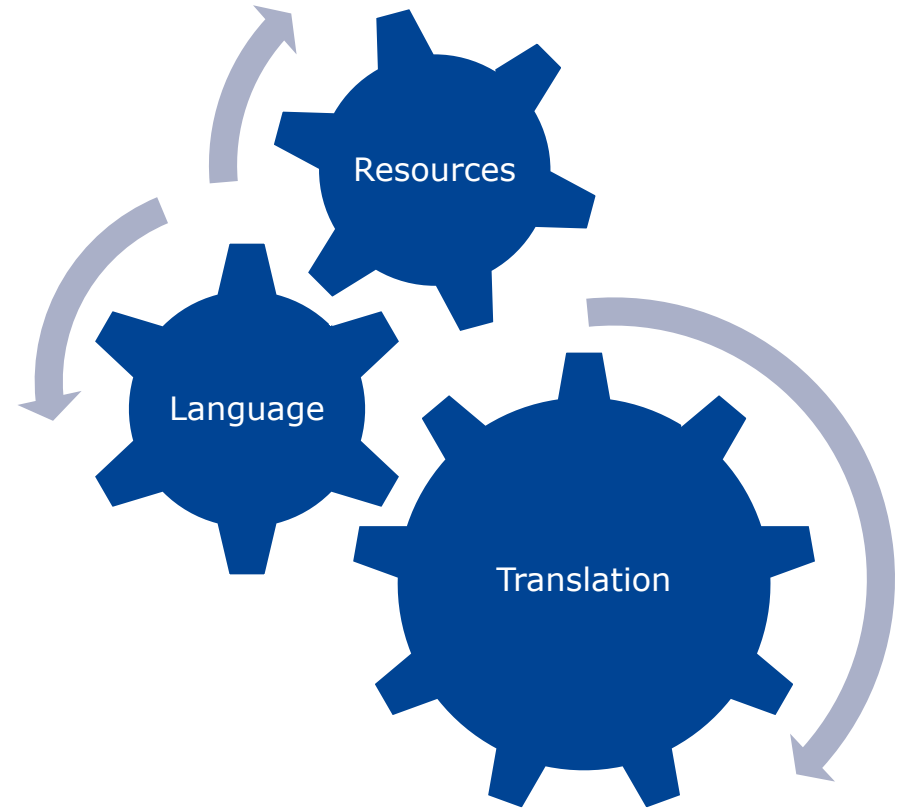
Thank you!

- LR are key for achieving high-quality output in MT in all EU languages

- They will also be indispensable input for other NLP tools in the future

No. of downloads of LRs by type

No. of Downloads

■ Monolingual Corpora

■ Bi-/Multilingual Corpora

■ Lexical resources / Terminologies

European Commission

CEF Telecom 2019 call for proposals

- Overall Budget: €4 million
- Indicative grants of €200,000 to €800,000   (5-20 projects)
- Co-funding: 75% of eligible costs
- Indicative duration: 2 years
- Deadline for submission: 14 May 2019

Three Priorities

1. Collect language resources

2. Collect language tools

3. Integrate and/or complement eTranslation into
   - CEF DSIs,
   - pan-European digital public services or
   - **Public administrations**

Freely and openly available

Sustainable

Conditions for participation

- Consortia must be composed of

  - at least **two public or private entities**,

  - based in **two or more Member States** and/or EEA countries participating in CEF Telecom program (EU, Norway, Iceland)

- Proposal should address at least **two languages**

- Consortia members need prior Member State/EEA country agreement

No research

Links & contacts:

CEF Call website:

https://ec.europa.eu/inea/en/connecting-europe-facility/cef-telecom/apply-funding/2019-automated-translation

eTranslation website:
http://ec.europa.eu/cefdigital/etranslation

eTranslation tool:
https://webgate.ec.europa.eu/etranslation

CNECT-LANGTECH@ec.europa.eu

June Lowery-Kingston HoU
Philippe Gelin HoS
Ainars Freimanis

European Commission

In a nutshell

- Help us spread the word …

  …in all official languages

Identify,

Share,

Benefit!

Now available to public administrations across Europe

Thank you!

Contact us

✉ CNECT-LANGTECH@ec.europa.eu