

Multilingualism and Language Technologies in the EU

Javier Hernandez-Ros Acting Director European Commission DG CONNECT Directorate G – Media and Data



Overall Context

- Linguistic diversity is one of the EU's fundamental values and cultural cornerstones : multilingualism is at the very heart of the European idea.
- 24 official EU languages and dozens of regional and minority languages.
- Language barriers affect cross-border provision and access to online content and services.
- Multilingualism is key to seize the opportunities and benefits of the Digital Single Market.



EU actions

- Research & Innovation
- Deployment
 - Machine Translation
 - Language Resource Management
- Coordination



EU support to Research & Innovation

 Support to R&I projects in FP7 (2007-2013) and H2020 (2014-2020)

Over 200 MEUR funding to date for

- machine translation development
- computer-assisted translation
- speech technologies
- language resources
- Upcoming funding opportunities under H2020

Draft Workprogramme 2018-2020. Earmarked budget of 20 MEUR



EU R&I success stories

- Moses SMT toolkit development and adoption by industry
- <u>MateCat</u> computer-aided translation tool with parallel data gathering, and <u>MMT</u>, an adaptive domain-sensitive MT infrastructure
- QT21 neural machine translation, high performance results for some languages



EU R&I success stories

• Innovative SMEs developing/exploiting MT

Iconic Translation Machines, KantanMT, Tilde, etc.

- Mobilised LT community
 - EC coordination in language resources collection
 - Cracking the Language Barrier Initiative



CEF Automated Translation

Origins:

EC is the largest translation service in the world:

- ➔ need for quality MT facilitating translators' work
- \rightarrow need for secure and confidential MT environment
- \rightarrow **MT@EC** custom developed by EC for EC

CEF programme:

- Objective : create Digital public services.
- Existing MT@EC was used to create **CEF Automated Translation**

- As a multilingual enabler for public online services (EU and national)



CEF Automated Translation Actions

- Provide multilingual functionalities/support to pan-European digital public services (DSIs)
- Language data collection and sharing, in line with the PSI policy
- MT@EC partly funded by ISA programme → open to use by MSs public administrations

Used by

- all EU institutions and bodies
- Member States public administrations Spanish public administrations
 3rd most important users
- European online services funded or supported by the EU



European Language Resource Coordination

Language resources are crucial for the development of automated translation services and for increasing machine translation quality.

ELRC's aims and objectives are:

- awareness raising for public services on value of data
- language data collection
- engaging public sector for the identification and sharing of data for CEF Automated Translation

How you can engage:

- 2017-2019: ELRC workshops in all CEF countries, 2 conferences
- contribute language data
- helpdesks for legal and technical issues related to data collection/provision, on-site assistance possibility



Spanish Plan for the Advancement of Language Technologies

Objectives are in line with EU objectives to:

- Boost LT industry and LT innovation
- Value the existing high quality LT research community
- Expand benefits of LT through public services
- ➔ It reflects the importance of language technologies as a cross sectoral growth enabler for strategic economic sectors.
- ➔ Recognises the importance of public administrations' need for LT products and services as driving force for LT innovation.
- ➔ Offers an example of best practice to other EU MSs in elaborating their own national strategies.



CEF WP 2017 https://ec.europa.eu/inea/sites/inea/files/c 2017 696 f1 annex en v3 p1 875665.pdf

H2020 https://ec.europa.eu/programmes/horizon2020/en/whathorizon-2020

Contact

Javier Hernández-Ros javier.hernandez-ros@ec.europa.eu

Aleksandra Wesolowska <u>aleksandra.wesolowska@ec.europa.eu</u>