



H2020 ICT WP 2014-2015

ICT 17

Cracking the language barrier

Aleksandra Wesolowska

Unit G.3 - Data Value Chain

H2020 Info Day, Luxembourg 15 January 2014



European Digital Single Market is fragmented by language barriers

- online commerce, social communication and exchange of cultural content hampered by the national/linguistic borders
- insufficient quality and coverage of current MT solutions
- no easy customization

Approach

Explore new avenues, methods, approaches to achieve ***significant improvement in translation quality*** in **fully automatic MT**



ICT 17 Cracking the language barrier (15 MEUR)

- (a) Research and innovation action **(4 MEUR)**
 - (b) Innovation actions **(10 MEUR)**
 - (c) CSA **(1 MEUR)**
-
- Call deadline: 23 April 2014, 17:00 Brussels time



- special focus on the official EU languages "facing digital extinction"
– 21 languages with "weak or no support" or "fragmentary support" of MT solutions (see META-NET Language White Papers)
- close collaboration and clustering with other H2020 actions and among all actions supporting a language resource infrastructure (META, CLARIN-ERIC, Connecting Europe Facility, national programs, structural funds...)
- for (a) and (b) **research** and/or **innovation** focus



- no research on computer-aided translation
- no post-editing of MT output
- content analytics only if it's a use case for fully automatic MT

- no minimum number of languages but the choice of languages is important (representative subset of European languages)
- emphasis on difficult, small EU languages as **target** language
- for (a) and (b) no prescribed action size = fit for purpose;
"projects requesting small EU contribution": 2-4 MEUR



one **research and innovation action**

- a multidisciplinary research action → new paradigm in MT
- focus on points where current systems fail
 - quality
 - adaptation
 - expensive to improve
 - need of large not available corpora
- test & validate against agreed benchmarks
- use existing and emerging structures (developed under CSA)



- self-learning/self-improving **fully automatic systems**, making best use of available data and language resources
- systems dealing with huge volumes, high variety of languages and text styles
- systems delivering results in reasonable time
- high quality = full compliance with user requirements
- preferably language independent approach
- focus on difficult target languages
- research focus but clear innovation potential



a few **innovation actions**

- test, validate, evaluate quality improvement in realistic use situations - experiment opportunities for new approaches to automatic translation
- use situations: possibly online services, CEF, other digital infrastructures, e-commerce... → bottlenecks in digital single market
- address "poorly served" languages
- connect, contribute & make use of platforms and infrastructures for language resources, open data...
- limited research and development activities are possible



- pilots in high-quality fully-automatic machine translation in any application area
- clear problem definition: *well documented market needs* and users
- clear business goals
- impact criterion weighted at 1.5 when establishing ranking



one **CSA**

- promote a common infrastructure for MT benchmarking
- promote optimal use of language resources from various sources (Member States, internet, business, research, administrations...)
- organisational and coordination aspects between all projects & actions in MT & LR area in all phases of lifecycle (research & innovation in H2020, deployment in CEF...)
- interoperability & metadata harmonisation
- observatory in the LT area
- single beneficiary or multi-party grant



- **By 2025**, an online EU **internal market free of language barriers**, delivering automatic translation quality, equal to currently best performing language pair/direction, in most relevant use situations and for at least 90% of the EU official languages
- **Significant improvement in quality, coverage and technical maturity of MT** for at least half of the 21 EU languages that currently have "weak or no support" or "fragmentary support" of MT solutions, according to the META-NET Language White Papers
- Large contributions of language resources and language technology tools (from all EU **Member States** and Associated countries) to **a single platform** for sharing, maintaining and making use of language resources and tools; establishing **widely agreed benchmarks for machine translation quality** and stimulating competition between methods and systems

→ proposals are expected to **contribute** to these impacts



Thank you!