



Language Technologies in Europe: trends and future perspectives

European Commission

Márta Nagy-Rothengass, DG *CONNECT*

Data Value Chain Unit

Warsaw, 27 September 2012

Outline

- **Language technology in CONNECT**
- **Language technology is important for the European Data Value Chain**
- **Language technology landscape in EU**

Conclusions



EUROPEAN COMMISSION

**The Communication Networks, Content & Technology
Directorate General**

Directorate Media and Data

G1: Converging Media and Content
G2: Creativity

G3: Data Value Chain

G4: Inclusion, Skills and Youth
G5: Administration and Finance

Since July 1st 2012



EUROPEAN COMMISSION

The Information Society and Media Directorate General

**Directorate Digital Content & Cognitive
Systems (based in Luxembourg)**

E1: Language Technologies, Machine Translation
E2: Technologies for Information Management
E3: Cultural Heritage & Technology Enhanced Learning
E4: Access to Information
E5: Cognitive Systems & Robotics
E6: eContent and Safer Internet
E7: Administration and Finance



The "language technology" portfolio will continue to exist, but under the broader "Data value" context

- A new opportunity to explore synergies between the "Data" challenge, Semantic Web, and language technologies
- A portfolio of 130+ projects with 250+ MEUR funding, 1500+ FTE
- We should make more effort to enhance multidisciplinary working together and profit from each other's achievements

Why does the EU support language technologies?

- The **European Union** brings together **500 million people** speaking **many languages...**
 - less than half of Europeans understand English
 - the Web and the online market is more multilingual than ever
- Towards a ***digital single market*** where content & services can flow freely
 - ✓ support cross-border exchanges between public online services
 - ✓ eCommerce
 - ✓ ease internationalisation of SMEs...

Language technologies in ICT policy context



- **A strategy for smart, sustainable and inclusive growth**
- **A vision to achieve high levels of employment, a low carbon economy, productivity and social cohesion, to be implemented through concrete actions at EU and national levels.**

➤ **One of the seven flagship initiatives of Europe 2020, set out to define the key enabling role that the use of ICTs will have to play if Europe wants to succeed in its ambitions for 2020.**

➤ **The overall aim [...] is to deliver sustainable economic and social benefits from a digital single market [...]**

➤ **Support research and innovation**

"The Commission is invited to make rapid progress in key areas of the digital economy to ensure the creation of the Digital Single Market by 2015, including [...] the availability of public sector Information."

Conclusions of the European Council (4 February 2011)

"DATA Value" Vision

to **enable** and **foster** best possible social and commercial added value

based on intelligent use, management and re-use of data sources in Europe.

This will lead to

- **increased business intelligence and efficiency** of private and public sectors
- **world class applications**
- **new business opportunities** involving SMEs - (open) data friendly policy and business environment

Main pillars of the "Data value chain" thinking

- 1. Creation of "data value" friendly policy environment**
- 2. Multilingual Data and service infrastructure**
- 3. European Research and innovation support of Language Technologies**

1, Creation of "data value" friendly policy environment:

- Fostering of (Open) **Data policy**
- Adoption of the **revised Directive on the re-use of Public Sector Information (PSI)** and the **Commission decision on re-use of its own information**
- **Implementation of PSI** policy across Europe by ensuring compliance and the development of **soft law instruments** (e.g. guidelines on licensing and charging)
- **Stakeholder** involvement and engagement

2, Multilingual (Open) Data and service infrastructure

- Development of **European Digital Service Infrastructure** and **fostering new services** in relation to
 - Building and reuse of language resources
 - Open Data portals at local, regional, national and European level
 - "Multilingual access to online services"
- Leading by best practice examples ...

... progressive deployment of a multilingual Open Data and service Infrastructure

- 2012: Launch of **European Commission Open Data Portal**
- 2013: **pan-European Open Data Portal** in place
- 2014-2020: **European Digital Service Infrastructure** (Commission proposals for new Multiannual Financial Framework – “Connecting Europe Facility”)

3, European Research and innovation support of Language Technologies

- first programmes in 1980's
- a fresh start since 2008
 - explosion of online content, esp. social networks
 - sizeable industry of language services
 - promising technical advances
- approx **150 M€ funding for LT projects** in the last 5 years
- nearly **60 LT projects** underway

Knowledge/information management, semantic technologies (70 projects)

- making sense of large amounts of information
- visualisation
- exploiting big/volatile data
 - specific Software architectures, data structures
 - specific Hardware
 - exploitation of data on the fly (rather than storage)

Main themes of our current projects

- machine translation (text, speech)
- speech recognition, dialogue systems
- Semantic technologies, information extraction, question answering, data mining in multilingual context
- annotation of content (text, multimedia, objects)
- summarization, digestion, visualisation – of information

II/ Language technology landscape in EU

- **Speech recognition and dialogue systems**
 - difficult...
- **Machine translation**
 - strong and mature, needs further push for quality & coverage
- **Analytics, semantic technologies, IE, IR**
 - growing, emerging, exciting
- **Language resources – necessary for all the above**
 - strong basis, requires continuity & coordination

Automatic Speech recognition (ASR) and dialogue technology

- **Europe has the best research and academic teams...**
...but has practically lost the ASR industry,
ASR industry is dominated by US & Japan
- **EU needs to find its niche – and fill the language gaps**
 - US giants may not be interested in small markets and languages
 - Opportunities: ground-breaking research on dialogue, open source development, technology transfer to SMEs...
- **Strategic importance for mobile devices & cars**
 - fast & reliable interaction without keyboard
 - safe operation while driving

Machine translation (MT)

- **Europe has the best research and academic teams...**
...and dynamic, networked industry, mostly SMEs
- **EU has many success stories in MT**
 - MOSES Open Source toolkit, and services building on it (e.g. LetsMT!)
 - Networking of MT industry & research in EU
 - Effective technology transfer
- **Remaining challenges**
 - Improve quality, break the glass ceiling
 - Cover all EU languages – both directions!
 - Consolidate the fragmented industry & research

Analytics

- **Growing importance** due to need **to analyse huge amounts** of online textual streams (web, news, social media) **in multiple languages**
- **EU** has (some) **industry**, mostly (specialized) **SMEs**
- **Recent developments:**
 - need for multilingual and cross-lingual analytics
 - multilingual semantic web & Linked Open Data
- **Remaining challenges**
 - Analyse noisy, multilingual streams (e.g. Twitter)
 - Cover all EU languages

Language resources (LR)

- **Long history – recent revival (2009-) through series of EU projects (e.g. META-NET)**
- **Indispensable raw material** for practically all **Language Technologies**
- **Recent developments:**
 - Automated acquisition and processing of LRs
 - Pan-European consolidation of LRs in process
- **Remaining challenges**
 - Make LRs accessible, usable, standardized, machine-readable
 - Cover all EU languages
 - Build LR and tools infrastructure for plug & play use

Conclusions

- There are **new opportunities arising** from the new EC focus and structure, use it!
- **'Big data' has the future:** LT is part of it
- **Towards the Digital Single Market in Europe we need LTs** for reducing the barriers
- **European digital service infrastructure** would help unleashing the potential of language resources (strategic importance of CEF)
- **There are good funding opportunities. They and CEF** will be addressed by Kimmo Rossi in the afternoon session

Further info

- ***Experts data base:***

<https://cordis.europa.eu/emmfp7/>

- ***Unit – Data Value Chain***

URL: <http://cordis.europa.eu/info-management/>

<http://cordis.europa.eu/fp7/ict/language-technologies/>

eMail to: cnect-g3@ec.europa.eu



European
Commission

Thank you!