

# Language Technologies in support of Multilingualism



*J. Mariani*  
LIMSI-CNRS &  
Director



Institute for Multilingual & Multimedia Information  
(IMMI)

# The Challenges of Multilingualism

- **Preserve the cultures (languages)**
  - Allow citizens to speak their own language
    - Study for the EC:
      - 90% of European citizens prefer to have access to a Web site in their native language
    - Only 30% of the Web in English (50% (2000), 35% (2004))
    - 50% of European citizens speak only one language
    - Only 3% of Japanese citizens fluently speak a foreign language
    - Only less than 5% of Indian citizens speak English

# The Challenges of Multilingualism

- **Allow for communication among humans**
  - European Union
    - 28 Member States, 24 official languages / 552 language pairs
    - 35 national languages, 60 to 220 languages spoken in Europe
    - 2,500 translators at the EC - 1.8 Mpages translated per year
    - Fully Multilingual EC: 8,500 translators – 6.8 Million documents
    - 30% European Parliament budget (300 M€) – 500 translators
    - Total cost for EU: 1.1 B€ per year = 2.2€ per European citizen
    - Study for the EC:
      - Economic barrier: Only 33% of EU citizens would agree to buy goods on the Internet in a foreign language
      - Cultural exchange: 80% of EU citizens think that websites in their language should be translated in foreign languages
  - Globalization of information
    - 72 hours of new video every minute on YouTube in all languages
    - > 6,000 languages = > 36,000,000 language pairs

# Examples of Needs

- Europe

- European Digital Library (Europeana)
  - 23 millions documents in 26 languages (2013)
  - Multilingual / crosslingual : need to have tools
- Business Intelligence (European Security Agency ENISA)
  - Multilingual alert and information exchange Platform for Members States
- European Patent Office
  - 3 languages (English, German, French) in order to decrease the costs
- European Commission, Parliament, Court of Justice (documents, reports, meetings)...
  - 1997 : 45% of source documents translated at the EC were in English and 40% in French
  - 2007 : 72% of source documents translated at the EC were in English and 12% in French!

# Examples of Needs

- **International**
  - Technical notices (aeronautics, car industry, consumer products...)
  - Dubbing and subtitling of audiovisuals
  - Translation of Texts, Videos, Radio & TV broadcasts on the Internet
  - Interpretation at meetings, conferences, workshops, courses
  - Interpretation in military and sanitary operations
  - Writing of scientific articles in native language

# Languages in scientific publications

- Publications and Citations Index (SCI)

	1980	1990	2000
English	85%	90%	96%
French	4%	2%	1%
German	5%	2.5%	1%
Spanish	0.7%	0.4%	0.3%
Japanese	0.7%	0.5%	0.3%
# docs.	555 000	690 000	950 000

	1990
English	97%
German	1%
French	0.6%
Japanese	0.1%

# Findings (1)

- Impossibility to answer (quickly/at all) all the numerous present and future needs with the present (and even future) Human Resources

# Findings (2)

- Considering Multilingualism is not the first priority in any economical sector
- But the sum of the small priorities in each of those economical sectors is very large
- It therefore necessitates a political thinking and a political action



# Findings (3)

- Multilingualism is necessary, but its cost is very important
- Benefiting from Language Technologies would facilitate multilingualism by
  - Decreasing its cost
  - Generalizing its use
  - If LT quality meets the user needs

# Findings (4)

- Languages that lack LT will be less and less used
  - Car GPS, Smartphone interaction, Internet search, Emergency access
- Languages that benefit from crosslingual LT will get more used
  - Machine Translation, Speech Translation

# Findings (5)

- LT aren't yet enough mature
  - Machine Translation didn't reach good enough quality for translating literary books or documents which request a good translation
  - However, it may already:
    - On one hand, help the human translator in his/her activity
    - On the other hand, provide an approximate translation for the general public, especially if free and online
- Therefore the research effort should be increasingly supported
  - Availability of Language Resources for all languages
  - Evaluation of Language Technology quality for all languages

# Situation in the EU

- Not yet a EC program addressing LT to support Multilingualism in the EU as a political issue (only as a research issue)
- Probably, EC doesn't have enough forces to address by itself all LT for all (EU official / European) languages

# Conclusion & Perspective

- Language Technologies are the only way to allow for full Multilingualism, in Europe and worldwide
- In order to develop LT, LR and LT evaluation are needed
  - Presently available for a very small set of languages
  - Other languages are in danger of digital extinction
- Proposal:
  - EC, MS and regions jointly support the development efforts
  - To share the LR and evaluate the LT within a common platform
  - Can be extended beyond Europe through collaboration with third parties