

An Historical Perspective on Language Resources and Evaluation in Europe.



J. Mariani
LIMSI-CNRS &
Director



Institute for Multilingual & Multimedia Information
(IMMI)

LRE in Europe

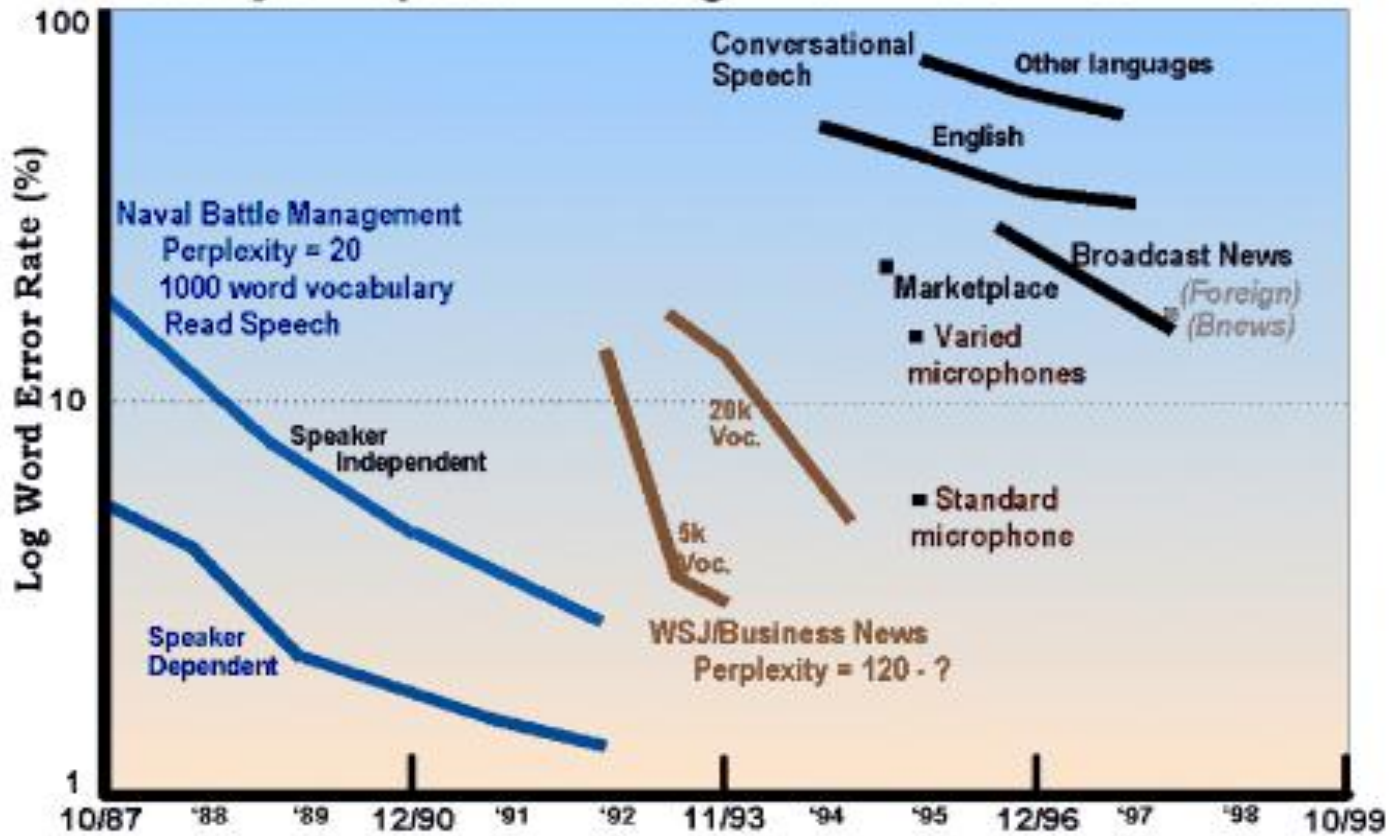
- Need to address Multilingualism in Europe
- Language Technologies (LT) can help allowing Multilingualism
- Necessity to have a platform to develop LT
 - Language Resources (LR)
 - Data: corpus, lexica, dictionaries, terminology...
 - Tools: annotation, transcription...
 - Meta-Resources: Standards, Metadata...
 - Language Technology Evaluation
- For all languages (11, 23, 60...)
- And all language pairs (110, 506, 3540...)

Europe on the international scene

NATO RSG10

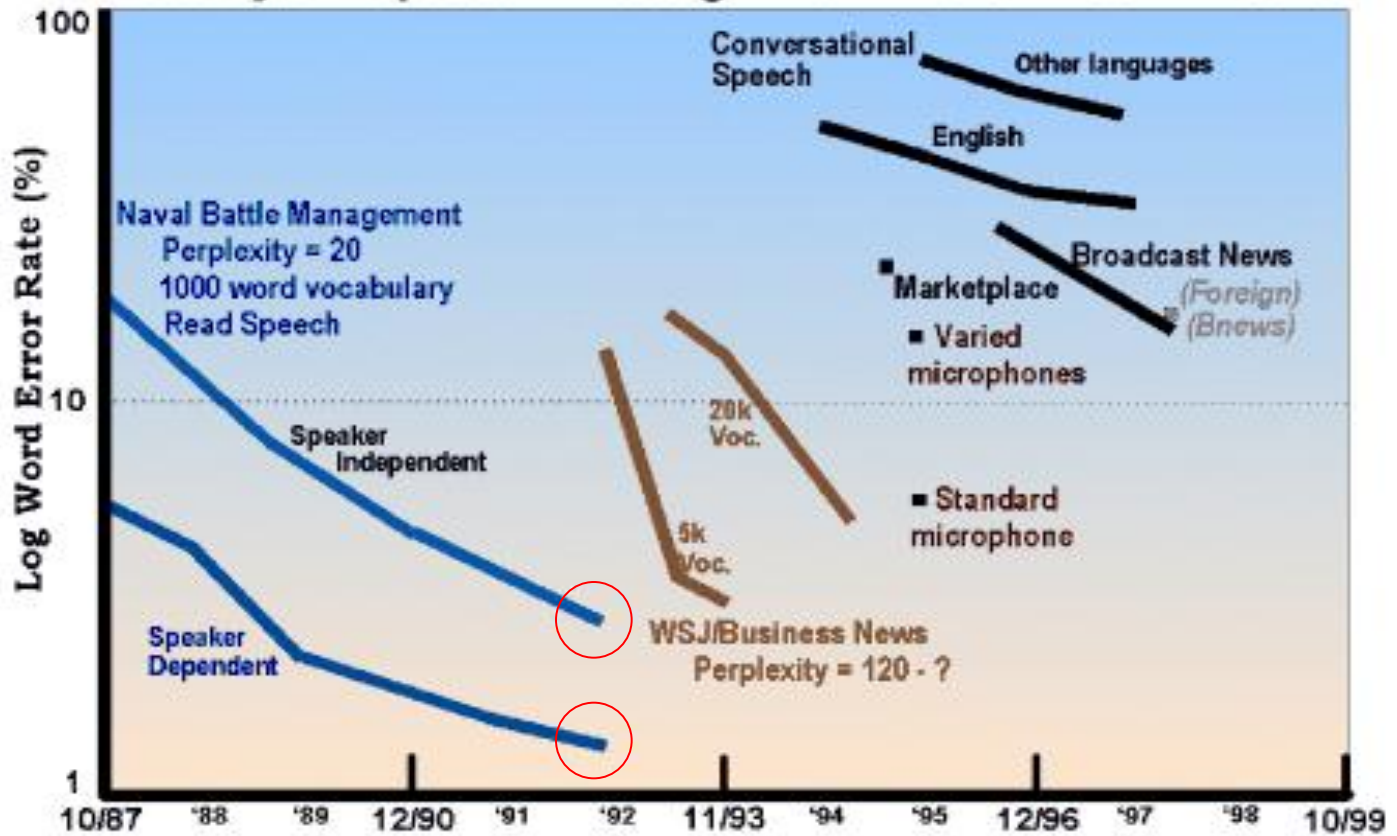
- First international multilingual comparative ASR evaluation campaign (1981)
- Isolated Digits 0-9 recorded on analog Magnetic Tapes
- 20 speakers
- 4 languages
 - English, French, German, Dutch
- 5 accents in English
 - (British, American, Dutch, German, French) English
- 5 systems tested
 - France, Germany, The Netherlands, UK, USA
 - Up to 15% WER

History of Speech Recognition Benchmark Tests



Courtesy NIST/ Darpa

History of Speech Recognition Benchmark Tests



Courtesy NIST/ Darpa

Cocosda

- COordinating COmmittee for Speech Databases and speech I/O systems Assessment
 - Created 1991 (satellite event of the ESCA Eurospeech'91, Chiavari, Italy)
 - Convenors
 - J. Mariani, A. Fourcin, M. Liberman, Lin-Shan Lee, K. Choukri, D. Gibbon
 - Used to meet at Interspeech conferences
 - Now jointly meets with NLP at LREC
 - Presentations of reports on LRE for the different Regions (5) and Topics

ELRA

- European Language Resources Association
- Launched 1995, spin-off of EC Relator project
 - Presidents: A. Zampolli, J. Mariani, B. Maegaard, S. Piperidis
 - General Secretary: K. Choukri
- LR identification & distribution (ELDA)
- LR production, validation, LT evaluation
- > 1000 LRs in catalog
 - Spoken/written language, terminology, evaluation packages
- Launched LREC biennial international conference (1998-)
 - > 1200 attendees in Istanbul (May 2012)

France

FRANCIL

- Language Engineering network (1994–2000)
- Francophone Universities Association
- Activities related to the French language
 - Language Resources production
 - 7 LT evaluation campaigns (ARC)
 - Written / Spoken language processing
 - 2 M€ budget

ARC

- Written Language (4)
 - Text Information Retrieval (Amaryllis)
 - French-English Corpus Alignment (Arcade)
 - Automated Terminology Extraction
 - Working group on Text Understanding
- Spoken Language (3)
 - Newspaper Voice Dictation: BREF
 - Vocal Dialog
 - Text-to-Speech synthesis

- Techno-Langue National Program
 - Funded by 3 French ministries in charge of Research, Industry and Culture
 - 2 M€ / year over 3 years (2003-2005)
 - Language Resources
 - Dictionaries (bilingual), lexica, corpora, terminological databases, tools...
 - 8 evaluation campaigns

- Written Language (5)
 - Information Retrieval (Question & Answer) (EQUER)
 - Parallel Text Alignment (ARCADE2)
 - Automatic Terminology Extraction (CESART)
 - Syntactic Parsing (EASY)
 - Machine Translation (CESTA)
- Spoken Language (3)
 - Speech Transcription for the automatic Indexing of radio/TV broadcast (ESTER)
 - Vocal Dialog (MEDIA)
 - Text-to-Speech Synthesis (EVASY)

FRANCE

- ESTER2 (BN transcription evaluation, MoD/DGA)
- ANR (National Research Agency) projects
 - Passage (Syntactic parsers evaluation) (2006-2009)
 - Port-Media (Dialog portability evaluation) (2008-2011)
 - ETAPE (BN transcription evaluation) (2009-2012)
 - Repère challenge (Multimedia person ident.) (2010-)
 - ANR and MoD/DGA
 - CHIST-ERA
 - Camomile (2012-)
 - Collaborative Multilingual / Multimedia Annotation

FRANCE

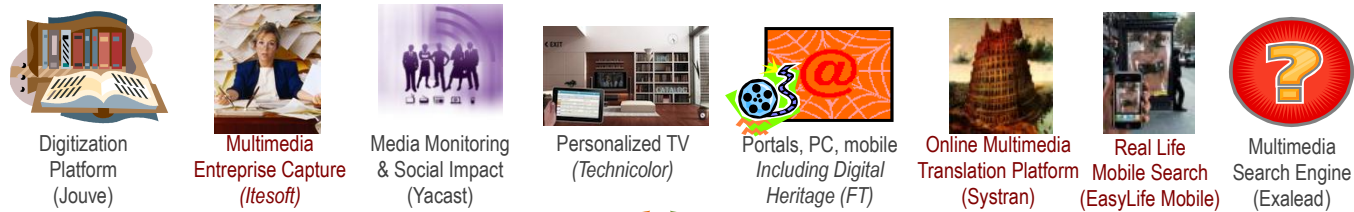


- Quaero Program (2008-2013)
 - Multilingual and Multimedia documents processing
 - Structured around:
 - development of technologies (~30),
 - for various medias (Speech, Text, Music, Image, Video),
 - in response to the needs of applicative projects (6->8),
 - based on the systematic use of corpus and evaluation.
 - 200 M€ budget over 5 years
 - 100 M€ public funding
 - 30 partners (Research/industry, France/Germany)

**Content!
Providers !**



Consumers!



**Validated!
Technologies!**

**Annotated!
Content!**

Raw Data!



**Core Technologies Cluster (CTC)
(LIMSI-CNRS)!**



**Corpus!
(RWTH)!**

Europe



EC projects on LR/Evaluation

- SAM (1987-1989)
- SQALE (1993-1995)
- EAGLES (1993-1998), ISLE (2000-2003)
- DISC (1997-1998, 1999-2000)
- ELSE (1998-1999)
- CLASS (2000-2003)
- CLEF (2000-)
- KHRESMOI (2010-2014)
- PROMISE (2010-2013), etc.

European LR projects



- Speechdat series (1997-2003)
 - Speechdat: Voice driven teleservices
 - Speechdat (II): speaker verification
 - Speechdat (M): fixed and mobile phone
 - Speechdat Car: in-vehicle and mobile phone
 - Speecon: Consumer devices
 - Speechdat (E): Eastern Europe
 - SALA: South America
 - LILA: Asia
 - Orientel: Mediterranean and Middle East...

European Commission



- 7th Framework Program (2007-2013)
 - 170 M€ for LT
 - Several projects on Infrastructures for Language Resources

CLARIN



- “Common Language Resources and technology Infrastructure”
 - ESFRI (European Strategy Forum on Research Infrastructure)
 - Distribution of LR and tools for Human and Social Sciences
 - 4.1 M€ EC funding
 - 2008-2010 : preparatory phase
 - 200 M€ planned budget (mostly funded by MS)
 - 2008-2020 : preparatory + construction + exploitation
 - 176 member institutions (33 countries)
 - ERIC (9 MS)

FLaReNet



- “Fostering Language Resources Network”
 - Thematic Network E-Content+
 - 0.9 M€ EC funding (2008-2011)
 - Promote Language Resources (data/tools) for research in Automatic Language Processing (ICT)
 - Think Tank
 - Forums (Vienna (2009), Barcelona (2010), Venice (2011))
 - LRE Map starting at LREC 2010 (4,200 entries)
 - 99 institutional members (31 countries)
 - 398 individual members
 - Link EU-US : INTEROP (NSF)



- Multilingual Europe Technology Alliance
 - Network of Excellence
 - 6 M€ funding (2010-2013)
 - Machine Translation / Multilingual Language Technologies



- ## 3 Pillars

- ### META-RESEARCH

- Pushing the research frontiers (MT)

- ### META-SHARE

- Open Resource Infrastructure: On-line LR distribution

- ### META-VISION

- Charting, Strategy, Roadmap, Promotion, Education
- LT Strategic Research Agenda for EC Horizon2020 FP
- Language White Papers (30 EU languages)
- Language Matrices: Language Coverage
 - » From LRE Map data : Need for a LR identifier

Language Matrix

Spoken Language Data

| | Bulgarian | Czech | Danish | Dutch | English | Estonian | Finnish | French | German | Greek | Hungarian | Irish | Italian | Latvian | Lithuanian | Maltese | Polish | Portuguese | Romanian | Slovak | Slovene | Spanish | Swedish | Other Europe | Asturian | Basque | Catalan | Galician | Arabic | Hindi | Japanese | Korean | Mandarin | Other | Multilingual | L.I. | N.A. | Total | |
|-------------------------------|-----------|-------|--------|-------|---------|----------|---------|--------|--------|-------|-----------|-------|---------|---------|------------|---------|--------|------------|----------|--------|---------|---------|---------|--------------|----------|--------|---------|----------|--------|-------|----------|--------|----------|-------|--------------|------|------|-------|---|
| Corpus | 1 | 4 | 7 | 12 | 80 | 1 | 1 | 35 | 22 | 2 | 2 | 0 | 9 | 0 | 0 | 0 | 7 | 7 | 1 | 0 | 1 | 18 | 5 | 26 | 0 | 6 | 4 | 3 | 15 | 1 | 12 | 3 | 11 | 31 | 2 | 4 | 3 | 336 | |
| Lexicon | 0 | 1 | 0 | 0 | 4 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 11 | 0 | 0 | 0 | 18 | |
| Grammar/Language Model | 0 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 2 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 6 |
| Acoustic and language models | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| Ontology | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 1 | |
| Statistical Speech Resource | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 |
| Total | 1 | 5 | 7 | 12 | 87 | 1 | 1 | 35 | 25 | 2 | 2 | 0 | 9 | 0 | 0 | 0 | 7 | 7 | 1 | 0 | 1 | 19 | 6 | 26 | 0 | 7 | 4 | 3 | 16 | 1 | 12 | 3 | 11 | 42 | 2 | 5 | 3 | 363 | |

LR Impact Factor (LRIF) Data

| | Bulgarian | Czech | Danish | Dutch | English | Estonian | Finnish | French | German | Greek | Hungarian | Irish | Italian | Latvian | Lithuanian | Maltese | Polish | Portuguese | Romanian | Slovak | Slovene | Spanish | Swedish | Other Europe | Asturian | Basque | Catalan | Galician | Arabic | Hindi | Japanese | Korean | Mandarin | Other | Multilingual | L.I. | N.A. | LRIF* | |
|-------------------------------|-----------|-------|--------|-------|---------|----------|---------|--------|--------|-------|-----------|-------|---------|---------|------------|---------|--------|------------|----------|--------|---------|---------|---------|--------------|----------|--------|---------|----------|--------|-------|----------|--------|----------|-------|--------------|------|------|-------|----|
| WordNet | 0 | 0 | 0 | 0 | 46 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 47 |
| Europarl | 1 | 0 | 13 | 16 | 38 | 0 | 12 | 26 | 21 | 11 | 0 | 0 | 12 | 0 | 0 | 0 | 0 | 13 | 0 | 0 | 0 | 20 | 20 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 44 |
| Wikipedia | 0 | 0 | 0 | 1 | 13 | 0 | 0 | 5 | 3 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 2 | 0 | 1 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 3 | 3 | 22 | |
| Penn Treebank | 0 | 0 | 0 | 0 | 15 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 1 | 0 | 0 | 0 | 0 | 15 | |
| Arabic Treebank (ATB) | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 12 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 1 | 13 | |
| British National Corpus (BNC) | 0 | 0 | 0 | 0 | 13 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 13 |
| Penn Discourse Treebank | 0 | 0 | 0 | 0 | 10 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 10 |
| Sparse data | 30 | 45 | 37 | 77 | 942 | 18 | 20 | 182 | 174 | 29 | 34 | 2 | 107 | 9 | 12 | 3 | 36 | 46 | 32 | 8 | 19 | 138 | 55 | 210 | 1 | 26 | 21 | 8 | 103 | 40 | 120 | 13 | 131 | 343 | 31 | 87 | 74 | 2038 | |

Parallel Corpora for MT

| | | | | | | | | | | | | | | | | | | | | | | | |
|--|-----|-----|-----|-----|-----|-----|----|----|----|----|----|----|---|---|---|---|----|---|---|----|---|---|---|
| | 638 | 90 | 83 | 55 | 40 | 29 | 28 | 12 | 8 | 8 | 10 | 9 | 7 | 7 | 6 | 6 | 5 | 4 | 4 | 5 | 3 | 3 | 2 |
| | 90 | 359 | 34 | 24 | 29 | 11 | 11 | 11 | 7 | 6 | 6 | 6 | 7 | 5 | 5 | 5 | 4 | 4 | 4 | 3 | 3 | 3 | 2 |
| | 83 | 34 | 306 | 17 | 16 | 11 | 12 | 10 | 8 | 6 | 6 | 6 | 7 | 5 | 5 | 6 | 4 | 4 | 4 | 2 | 3 | 3 | 2 |
| | 52 | 24 | 17 | 207 | 14 | 10 | 9 | 10 | 7 | 5 | 5 | 4 | 6 | 5 | 5 | 5 | 3 | 4 | 4 | 2 | 3 | 3 | 2 |
| | 39 | 29 | 16 | 14 | 187 | 8 | 7 | 10 | 8 | 6 | 6 | 4 | 6 | 5 | 5 | 5 | 3 | 4 | 4 | 3 | 3 | 3 | 2 |
| | 29 | 11 | 11 | 10 | 8 | 102 | 12 | 4 | 6 | 5 | 6 | 4 | 6 | 5 | 5 | 5 | 3 | 4 | 4 | 2 | 3 | 3 | 2 |
| | 26 | 11 | 12 | 9 | 7 | 12 | 89 | 4 | 6 | 4 | 5 | 4 | 7 | 5 | 5 | 5 | 3 | 4 | 4 | 3 | 3 | 3 | 2 |
| | 11 | 11 | 10 | 10 | 10 | 4 | 4 | 71 | 5 | 3 | 3 | 3 | 4 | 4 | 3 | 3 | 3 | 3 | 2 | 2 | 3 | 3 | 2 |
| | 8 | 7 | 8 | 7 | 8 | 6 | 6 | 5 | 41 | 4 | 4 | 3 | 5 | 4 | 4 | 4 | 3 | 4 | 3 | 3 | 3 | 3 | 2 |
| | 8 | 6 | 6 | 5 | 6 | 5 | 4 | 3 | 4 | 34 | 5 | 4 | 4 | 4 | 4 | 4 | 3 | 3 | 3 | 3 | 3 | 3 | 2 |
| | 10 | 6 | 6 | 5 | 6 | 6 | 5 | 3 | 4 | 5 | 25 | 4 | 5 | 4 | 5 | 5 | 3 | 4 | 4 | 2 | 3 | 3 | 2 |
| | 9 | 6 | 6 | 4 | 4 | 4 | 4 | 3 | 3 | 4 | 4 | 24 | 4 | 4 | 4 | 4 | 4 | 3 | 3 | 2 | 3 | 3 | 2 |
| | 7 | 7 | 7 | 6 | 6 | 6 | 7 | 4 | 5 | 4 | 5 | 4 | 7 | 5 | 5 | 5 | 3 | 4 | 4 | 2 | 3 | 3 | 2 |
| | 7 | 5 | 5 | 5 | 5 | 5 | 5 | 4 | 4 | 4 | 4 | 4 | 5 | 9 | 4 | 4 | 3 | 3 | 3 | 2 | 3 | 3 | 2 |
| | 6 | 5 | 5 | 5 | 5 | 5 | 5 | 3 | 4 | 4 | 5 | 4 | 5 | 4 | 7 | 5 | 3 | 4 | 4 | 2 | 3 | 3 | 2 |
| | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 3 | 4 | 4 | 5 | 4 | 5 | 4 | 5 | 7 | 3 | 4 | 4 | 2 | 3 | 3 | 2 |
| | 5 | 4 | 4 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 4 | 3 | 3 | 3 | 3 | 13 | 3 | 2 | 2 | 3 | 3 | 2 |
| | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 3 | 4 | 3 | 4 | 3 | 4 | 3 | 4 | 4 | 3 | 4 | 3 | 2 | 3 | 3 | 2 |
| | 4 | 4 | 4 | 4 | 4 | 4 | 4 | 7 | 3 | 3 | 4 | 3 | 4 | 3 | 4 | 4 | 7 | 3 | 4 | 7 | 7 | 7 | 1 |
| | 4 | 3 | 2 | 2 | 3 | 2 | 3 | 2 | 3 | 3 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 17 | 2 | 2 | 1 |
| | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 2 | 2 | 3 | 3 | 2 |
| | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 3 | 2 | 2 | 3 | 3 | 2 |
| | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 2 | 1 | 1 | 2 | 2 | 2 |



Courtesy Euromatrix

MT Systems Matrix (BLEU)

Translating between all EU-27 languages ³



| | Target Language | | | | | | | | | | | | | | | | | | | | | |
|----|-----------------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|------|
| | en | bg | de | cs | da | el | es | et | fi | fr | hu | it | lt | lv | mt | nl | pl | pt | ro | sk | sl | sv |
| en | - | 40.5 | 46.8 | 52.6 | 50.0 | 41.0 | 55.2 | 34.8 | 38.6 | 50.1 | 37.2 | 50.4 | 39.6 | 43.4 | 39.8 | 52.3 | 49.2 | 55.0 | 49.0 | 44.7 | 50.7 | 52.0 |
| bg | 61.3 | - | 38.7 | 39.4 | 39.6 | 34.5 | 46.9 | 25.5 | 26.7 | 42.4 | 22.0 | 43.5 | 29.3 | 29.1 | 25.9 | 44.9 | 35.1 | 45.9 | 36.8 | 34.1 | 34.1 | 39.9 |
| de | 53.6 | 26.3 | - | 35.4 | 43.1 | 32.8 | 47.1 | 26.7 | 29.5 | 39.4 | 27.6 | 42.7 | 27.6 | 30.3 | 19.8 | 50.2 | 30.2 | 44.1 | 30.7 | 29.4 | 31.4 | 41.2 |
| cs | 58.4 | 32.0 | 42.6 | - | 43.6 | 34.6 | 48.9 | 30.7 | 30.5 | 41.6 | 27.4 | 44.3 | 34.5 | 35.8 | 26.3 | 46.5 | 39.2 | 45.7 | 36.5 | 43.6 | 41.3 | 42.9 |
| da | 57.6 | 28.7 | 44.1 | 35.7 | - | 34.3 | 47.5 | 27.8 | 31.6 | 41.3 | 24.2 | 43.8 | 29.7 | 32.9 | 21.1 | 48.5 | 34.3 | 45.4 | 33.9 | 33.0 | 36.2 | 47.2 |
| el | 59.5 | 32.4 | 43.1 | 37.7 | 44.5 | - | 54.0 | 26.5 | 29.0 | 48.3 | 23.7 | 49.6 | 29.0 | 32.6 | 23.8 | 48.9 | 34.2 | 52.5 | 37.2 | 33.1 | 36.3 | 43.3 |
| es | 60.0 | 31.1 | 42.7 | 37.5 | 44.4 | 39.4 | - | 25.4 | 28.5 | 51.3 | 24.0 | 51.7 | 26.8 | 30.5 | 24.6 | 48.8 | 33.9 | 57.3 | 38.1 | 31.7 | 33.9 | 43.7 |
| et | 52.0 | 24.6 | 37.3 | 35.2 | 37.8 | 28.2 | 40.4 | - | 37.7 | 33.4 | 30.9 | 37.0 | 35.0 | 36.9 | 20.5 | 41.3 | 32.0 | 37.8 | 28.0 | 30.6 | 32.9 | 37.3 |
| fi | 49.3 | 23.2 | 36.0 | 32.0 | 37.9 | 27.2 | 39.7 | 34.9 | - | 29.5 | 27.2 | 36.6 | 30.5 | 32.5 | 19.4 | 40.6 | 28.8 | 37.5 | 26.5 | 27.3 | 28.2 | 37.6 |
| fr | 64.0 | 34.5 | 45.1 | 39.5 | 47.4 | 42.8 | 60.9 | 26.7 | 30.0 | - | 25.5 | 56.1 | 28.3 | 31.9 | 25.3 | 51.6 | 35.7 | 61.0 | 43.8 | 33.1 | 35.6 | 45.8 |
| hu | 48.0 | 24.7 | 34.3 | 30.0 | 33.0 | 25.5 | 34.1 | 29.6 | 29.4 | 30.7 | - | 33.5 | 29.6 | 31.9 | 18.1 | 36.1 | 29.8 | 34.2 | 25.7 | 25.6 | 28.2 | 30.5 |
| it | 61.0 | 32.1 | 44.3 | 38.9 | 45.8 | 40.6 | 26.9 | 25.0 | 29.7 | 52.7 | 24.2 | - | 29.4 | 32.6 | 24.6 | 50.5 | 35.2 | 56.5 | 39.3 | 32.5 | 34.7 | 44.3 |
| lt | 51.8 | 27.6 | 33.9 | 37.0 | 36.8 | 26.5 | 21.1 | 34.2 | 32.0 | 34.4 | 28.5 | 36.8 | - | 40.1 | 22.2 | 38.1 | 31.6 | 31.6 | 29.3 | 31.8 | 35.3 | 35.3 |
| lv | 54.0 | 29.1 | 35.0 | 37.8 | 38.5 | 29.7 | 8.0 | 34.2 | 32.4 | 35.6 | 29.3 | 38.9 | 38.4 | - | 23.3 | 41.5 | 34.4 | 39.6 | 31.0 | 33.3 | 37.1 | 38.0 |
| mt | 72.1 | 32.2 | 37.2 | 37.9 | 38.9 | 33.7 | 48.7 | 26.9 | 25.8 | 42.4 | 22.4 | 43.7 | 30.2 | 33.2 | - | 44.0 | 37.1 | 45.9 | 38.9 | 35.8 | 40.0 | 41.6 |
| nl | 56.9 | 29.3 | 46.9 | 37.0 | 45.4 | 35.3 | 49.7 | 27.5 | 29.8 | 43.4 | 25.3 | 44.5 | 28.6 | 31.7 | 22.0 | - | 32.0 | 47.7 | 33.0 | 30.1 | 34.6 | 43.6 |
| pl | 60.8 | 31.5 | 40.2 | 44.2 | 42.1 | 34.2 | 46.2 | 29.2 | 29.0 | 40.0 | 24.5 | 43.2 | 33.2 | 35.6 | 27.9 | 44.8 | - | 44.1 | 38.2 | 38.2 | 39.8 | 42.1 |
| pt | 60.7 | 31.4 | 42.9 | 38.4 | 42.8 | 40.2 | 60.7 | 26.4 | 29.2 | 53.2 | 23.8 | 52.8 | 28.0 | 31.5 | 24.8 | 49.3 | 34.5 | - | 39.4 | 32.1 | 34.4 | 43.9 |
| ro | 60.8 | 33.1 | 38.5 | 37.8 | 40.3 | 35.6 | 50.4 | 24.6 | 26.2 | 46.5 | 25.0 | 44.8 | 28.4 | 29.9 | 28.7 | 43.0 | 35.8 | 48.5 | - | 31.5 | 35.1 | 39.4 |
| sk | 60.8 | 32.6 | 39.4 | 48.1 | 41.0 | 33.3 | 46.2 | 29.8 | 28.4 | 39.4 | 27.4 | 41.8 | 33.8 | 36.7 | 28.5 | 44.4 | 39.0 | 43.3 | 35.3 | - | 42.6 | 41.8 |
| sl | 61.0 | 33.1 | 37.9 | 43.5 | 42.6 | 34.0 | 47.0 | 31.1 | 28.8 | 38.2 | 25.7 | 42.3 | 34.6 | 37.3 | 30.0 | 45.9 | 38.2 | 44.1 | 35.8 | 38.9 | - | 42.7 |
| sv | 58.5 | 26.9 | 41.0 | 35.6 | 46.6 | 33.3 | 46.6 | 27.4 | 30.9 | 38.9 | 22.7 | 42.0 | 28.2 | 31.0 | 23.7 | 45.6 | 32.2 | 44.2 | 32.7 | 31.3 | 33.5 | - |

(using the Acquis corpus)

[from Koehn et al., 2009]

Conclusion

- Need for LT to allow for multilingualism in Europe
- Language Technologies presently considered as any other scientific area, not as a political issue for the EU
 - Not all 23 languages are covered in terms of LT/LR
 - PR “At Least 21 European Languages in Danger of Digital Extinction “
 - Many generally short term / medium scale / national / stand alone / conflicting projects
 - No sustainable long term action
- ELRA as the counterpart of LDC
- No permanent entity comparable to NIST
 - CELCT, ELRA, LNE...
- Need for a shared effort between EC and MS on LT

Perspectives

- Distributed LRE European infrastructure
 - META-SHARE
 - Extend from distribution to col-laboration
 - Specification, Production, Annotation, Evaluation...
- International coordination on LRE
 - Avoid redundancy
 - Assign unique LR Identifier (LRID / ISLRN)