



*Ministère de la Jeunesse, de l'Éducation Nationale et de la
Recherche*

**Language Technologies & the ERA:
Let's join forces !**

Joseph Mariani

Director

« Information & Communication Technologies » Department
French Ministry of Research & New Technologies

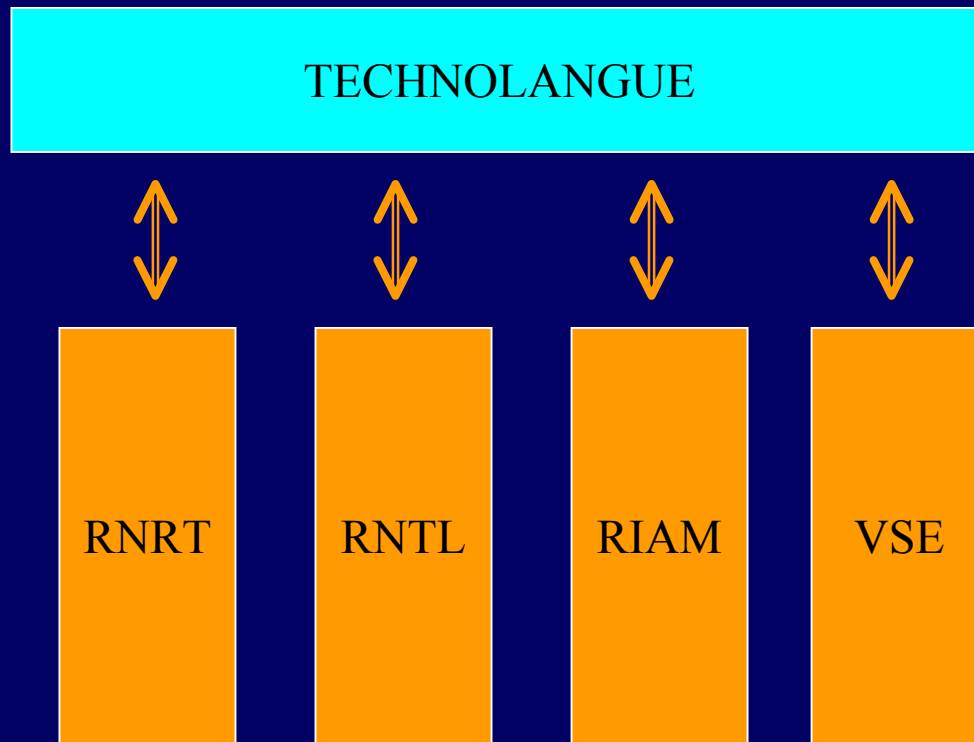
Language Technologies in Europe

- A national project in LT:
 - The TechnoLangue action in France
- A proposal for a future action:
 - ERA-Net in Language Technologies : LangNet

« TechnoLangue » action

- Report to the Prime Minister (November 2000)
- Meeting Min. Industry, Research, Culture: June 2001
- Action : Language technology survey and evaluation
- Basic Technological Research
- Articulate with related existing programs
 - Research & Innovation Technological Networks (RRIT)
 - 3 ICT RRIT: Telecommunications, Software Engineering, Audiovisual & multimedia (already funded by 3 ministries)
 - Ministry of Research action on Technological Survey (VSE)

« TechnoLangue » infrastructure



Infrastructural program to support innovative core technologies development, while existing R&D application projects stay with RRIT & VSE (120 M€ / year)

« TechnoLangue » Call

– Language resources

- Spoken/written data (corpus, dictionaries, terminological data...)
- Basic Language Processing Tools (Open Source software)

– Evaluation

- Technology (evaluation campaigns: comparison of systems)
- Applications (evaluation toolkits)

– Norms & standards

- Shared effort to improve French participation

– Technological survey

- In relationship with on-going actions (Euromap...)

« TechnoLangue » Call

- **Steering Committee:**
 - 15 participants – Chair : C. FLUHR (CEA)
 - Representatives from 3 RRIT + experts
 - 50% industry / 50% public research
 - Spoken / Written language
- **Calendar:**
 - Call For Proposals : April 15, Deadline : June 10, 2002
 - Final selection : September 5, 2002
 - Contracts granted by MRNT : December 2002
- **International cooperation**
 - Foreign entities may participate in the projects
 - Should bring their own funding

Results

- 52 proposals submitted
- 27 projects selected
- 82 participants
 - 32 industry
 - 39 public research
 - 11 other (Associations, CEA, French DoD...)
 - 11 foreign (Bell Labs (USA), NII (Japan), EPFL, LATL...)
- Budget : 13 M€ effort - 7.5 M€ funding
- Same on Vision Technology in 2004

Results

- 27 selected projects:
 - 9 on Language resources
 - BLARK (Cf BNC), Fr-En, G, Sp, It, Arabic dictionaries
 - Specialized (aerospace, automotive...), proper names dictionaries
 - Aligned corpus (7 novels 19th century literature in 4 languages)
 - Children voices and generic voices
 - 5 on Tools (Open source software)
 - Lemmatizer, Chunker, Guesser, Tagger, Parser, Speaker recogn., Topic & NE detector, summarizer, term. extractor, Search engine...
 - 2 on Standards (Spoken / Written)
 - 1 on Technological survey (technolange.net Portal)
 - 10 on Evaluation : 9 on technology eval., 1 on usage eval.

Technology Evaluation

- **Written language**
 - Syntactic parsing
 - Text alignment
 - Terminology extraction
 - Information retrieval : Q&A
 - Machine translation
- **Spoken Language**
 - Speech synthesis
 - Speaker recognition
 - Spoken dialog
 - Speech transcription / automatic indexing

FP6 and the ERA

- 6th Framework Program (2002-2006)
 - Usual projects (Streps)
 - New instruments
 - Networks of Excellence
 - Integrated Projects
 - Construction of the European Research Area (ERA)
 - Article 169
 - Proposal comes from Member States
 - Co-decision Council + Parliament
 - «European developing countries clinical trials» program labelled in 2002 (3*200 = 600 M€)
 - ERA-Net projects (may prepare Article 169...)

ERA-Net

- Cooperation & coordination of national or regional research and innovation activities (i.e. programs)
 - Exchange of information and best practices (projects, priorities, evaluation process, management...)
 - Strategic activities (complementarities, multinational scheme, barriers identification)
 - Implementation joint activities (clustering, joint evaluation, use of testbeds)
 - Transnational research activities (joint programs, calls, funding...)

ERA-Net

- 148 M€ total FP6 budget (2002-2006)
- EC funds the coordination costs, no R&D
- First deadline : June 3, 2003. Next: March 2, 2004
- Eligibility: Legal entities from at least 3 MS / AS
 - Public bodies financing/managing research at national/regional level
 - In addition, other organizations doing the same (charities)
 - European bodies (EEIG...(may be sole as a group of public bodies))
- **Coordinating Actions (CA)**
 - 3 M€ max. over 5 years max. - 100% costs (incl. subcontracts)
- **Specific Support Action (SSA)**
 - Preparation of CA: 200 K€ max. over 1 year max.

Language Technologies in EU

- **Language as a specific issue for Europe**
 - Economical, cultural and political challenge
 - Preference for native language (Germany (75%)...)
 - Need for communication across countries
 - 15 countries / 11 languages
 - 1170 translators at the EC - 1.3 Mpages translated in 2001
 - EU enlargement : 10 more countries in May 2004
 - Enormous cost for the EC : even more enormous
- **Need for Language Technologies**
- **Huge effort, too large for the EC alone : should be shared with EU member states (subsidiarity)**

Language Technologies Program

- **LT well fitted with European Research Area (ERA)**
 - The EC supports
 - the coordination: management, standards, technology evaluation, communication...
 - the development cost of generic language technologies:
 - Speech recognition, synthesis, understanding, spoken dialog, language tagging, parsing, analysis, generation, text retrieval, document understanding, machine translation...
 - Each member state supports the cost for covering its language(s):
 - Language Resources (essential) : (annotated) corpus (spoken / written), lexicon (including pronunciations), dictionaries...
 - Language specific technology development/adaptation

LangNet

- **Infrastructural nature**
 - Language Resources
 - Language Technology evaluation
 - Norms and standards
 - Survey
- **Extension to a larger program including EC FP6 projects**
 - Contact with EC
 - Article 169 on Technology support to Multilingual Europe ?

LangNet

- **Identify EU Countries having similar programs**
 - France (TechnoLangue)
 - Germany, Italy, Trentino, Norway, Finland, Denmark, The Netherlands / Belgium (Dutch Union) (contacts with officials)
 - Spain, Austria, Greece, UK, Portugal, Sweden (first contacts)
 - Czech republic (first contact)
 - Switzerland... (national programs)
- **Extendable to other partners**
 - AS (Slovenia, Cyprus, Romania, Bulgaria, Poland, Hungary...)
 - USA, Japan, South Africa (contacts)...
- **Language as a grand strategic challenge for Europe**