Human Language Technology Initiatives
EU research programmes FP6 and FP7

Mats Ljungqvist
European Commission, Luxembourg
Information Society and Media – unit E1

Message

- **Recognised importance of joint evaluation and shared data/tools/ annotation schemes**
- **FP6:**
  - significant effort in HLT (biggest ever) including work in MT and ML
- **FP7:**
  - HLT, ML, MT has a ”home”
  - Ambition to increase effort on MT, ML
  - However: EC need help in defining exactly what should be done and how
    - **Europe needs to set its own challenges!**

Evaluation

- **FP4-5 efforts in**
  - Evaluation, tools, data, standards
  - Through separate (small) projects
- **FP6 efforts**
  - Mainly though large projects (IP & NoE)
  - Increased weight in international evaluation campaigns (incl. NIST)
- **How to organise EU evaluation work?**
  - Cooperation with partners:
    - Organisations: ELRA, LDC, NIST,...
    - Projects: CHIL, AMI, CLEF, TC-STAR, GALE,....
  - Opportunity for FP7 NoE (or ERA-net, Art 169)?

Framework Programme 6

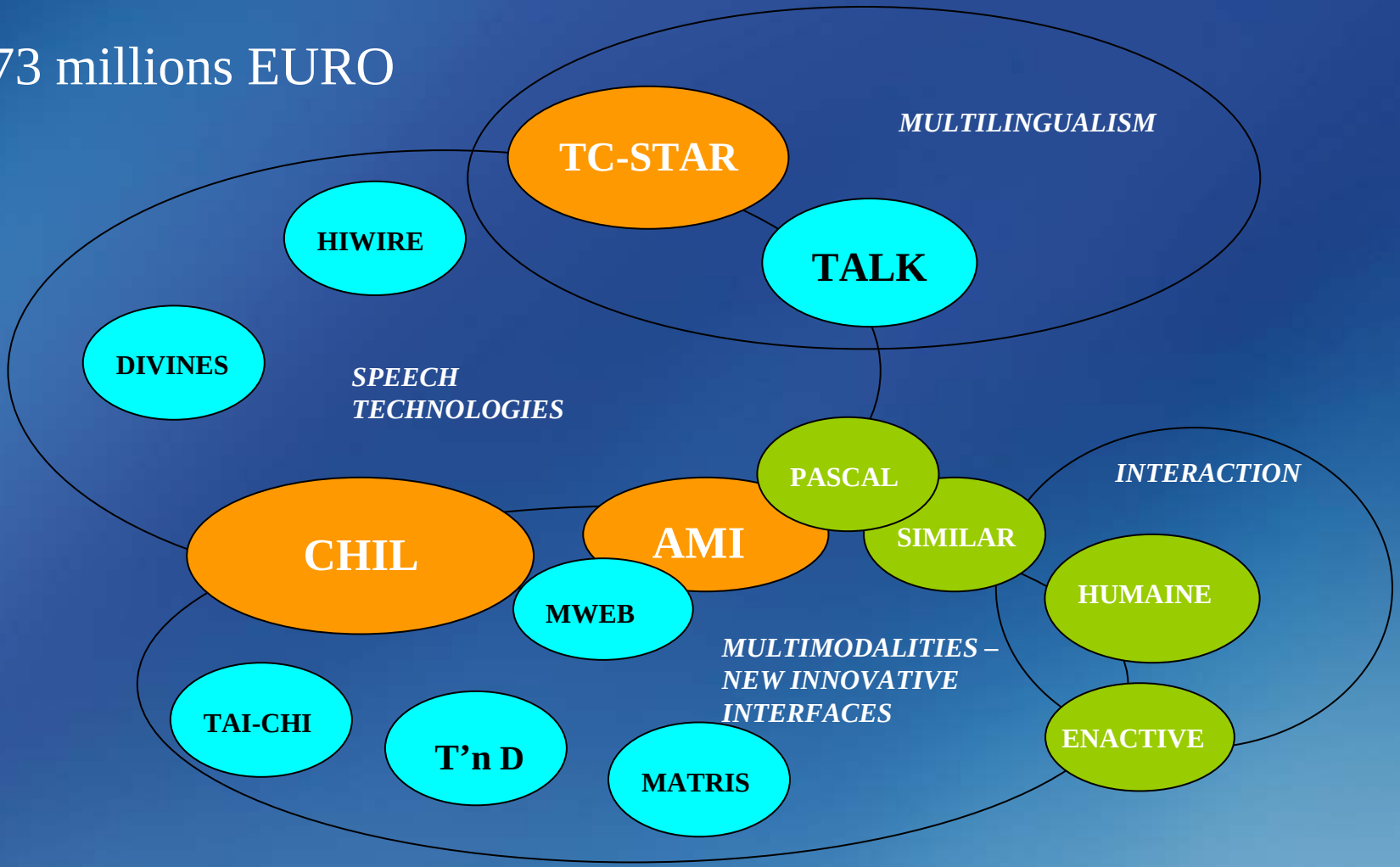
2002-2006

"The focus of IST in FP6 is on the future generation of technologies in which computers and networks will be integrated into the everyday environment, rendering accessible a multitude of services and applications through **easy-to-use human interfaces**"

- **2 Focus areas:**
  - Natural interaction between humans and the physical or virtual environment
  - Multilingual communication systems
- **Funding:**
  - Approx. 135 MEUR (EU-funding) 2002-2006
  - "pure HLT" R&D: 50-60%

FP6 portfolio (call 1)

73 millions EURO



IST call 5

Closed: 21 September 2005

Budget: 62 M€

Instruments: IP, STReP

New / Traditional Instruments: 60/40 funding

Number of proposals: 101

ESRs: mid-December 2005

Invitations to negotiations: February 2006

Project launch: 14-16 projects by mid-2006

Coverage Call 5

(areas covered by selected projects)

- **HLT, Multilingual, Machine Translation**
  - Statistical and hybrid Machine Translation
  - Language portability
  - Conversational interfaces, intelligent agents
  - Speech technology and dialogue
- **New modalities, Multimodality**
  - Haptics and tactile interfaces
  - Holographic displays, 3D tracking
  - Emotional aspects
  - Collaborative systems, meeting support
  - Open software platforms & user centred design
  - Interfaces for demanding applications, e.g.
    - Automotive, mobile and home environments
    - Surgery, artistic creation, security

Framework Programme 7

2007-2013
?

- FP7 specific programme on Cooperation:
  - "Technology pillar": Simulation, Visualisation, Interaction and Mixed Realities
  - "Integration of technologies" and "Applications research"

Includes:

- **Increased and more visible effort in multilinguality and machine translation**
- Intuitive user interfaces
- Budget still to be defined

FP7

- Commission proposal for specific programmes published
  - europa.cec.eu.int (cf. programme "cooperation")

- **ICT technology pillar** on simulation, visualisation, interaction and mixed realities:
  - *"Natural intuitive and easy-to-use interfaces and new ways to interact with technology, machines, devices and artefacts"*
  - *"Multilingual and automatic machine translation system"*
- **Integration of technologies**
  - *"Personal environments: integration of multimodal interfaces, …., personal communication and computing devices"*
  - *"Home environments: …, access to information,…, Management of knowledge"*
- Also in **applications research** and **ICT supporting business and industry**

FP7

- **What does it mean?**
  <u>You decide!</u>
  - **<u>You</u>** need to feed us ideas, roadmaps, justifications, target application areas, good R&D
    - Ad hoc expert group on multilinguality and MT
      - June 2005 meeting in Luxembourg
    - Dedicated FP7 workshop with repr. of current MMI&ML projects: 30-31 January 2006
    - Possible joint EU – GALE workshop in February (or April)
    - Suggestions, “white papers” welcome anytime
    - Open web-consultations likely to be organised

S&T Forum on Multilingualism

- Joseph Mariani, Ministry of Research, France
- Eva Hajicova, Charles University, Czech Republic
- Gabor Pròszeky, Morphology, Hungary
- Piek Vossen, Irion Technologies, The Netherlands
- Nicoletta Calzolari, CNR-LC, Italy
- Jörg Schütz, IAI, Germany (absent)
- Stelios Piperidis, ILSP, Greece
- Bente Maegaard, CST, Denmark
- Daniel Grasmick, SAP AG, Germany
- Bernd Reuse, BMBF, Germany
- Hans Uszkoreit, DFKI, Germany
- Alex Waibel, University Karlsruhe, Germany
- Gianni Lazzari, ITC-IRST, Italy
- Rose Lockwood, consultant, UK
- Martin Kay, Stanford University, USA
- Hervé Blanchon, GETA, France

Luxembourg, 6 June 2005

1-2 December 2005, Malta

Priorities for the future

- **Weak integration into real applications**
  - need to demonstrate added value - focus on less ambitious tasks with high potential impact – get useful results fast!
  - parallel evolution of technologies and resource
    - implementations often do not exploit new types of resources
    - claims for resources often not empirically evaluated
  - focus on language and media independent analysis and synthesis – text summarisation, retrieval in multilingual corpora, processing multi-modal data (speech, video, etc.), automatic encyclopaedia for resources in different source languages answer questions, help decision making
  - focus on basic system features – performance and robustness, systems that learn and adapt, human-factors and usability, portability across languages, media and delivery

Ways and devices that make language "disappear"

Priorities for the future

- **Language infrastructure(s)**
  - basic resources (spoken, written) for all languages
  - make large-scale resources available on the Internet using the “Open Source” model
  - focus on interoperability, reusability and tools for rapid resource creation (acquisition, annotation, porting across domains, languages, etc.)
  - new types of resources (metadata, multilingual, facts and common-sense knowledge, example-based context-sensitive)
  - integrate lexicons, terminologies and ontologies into knowledge resources
  - share and integrate annotation environments
  - foster international consensus

Priorities for the future

- **Speech**
  - open (domain-unlimited) speech translation systems
  - fully automatic speech recognition of spontaneous, conversational speech with error rates <10%
  - fusion with other modalities
- **Evaluation / performance**
  - shift to evaluating quality
  - end-to-end evaluation in real world situations
  - more coordination of distributed effort
  - develop a strong persistent evaluation infrastructure in Europe
  - Europe should set its own challenges

Priorities for the future

- **Machine translation**
  - translation does not have to perfect, or even very good, to be useful, how to build confidence with users?
  - focus on helping the translator, and exploit new ways for man-machine cooperation in the dialogue between the translator and machine
  - focus on reliability (robustness) and performance (quality for “publishing” information not just on quantity for “gisting”), predicting the quality of a translation
  - major shift to hybrid systems integrating statistical modelling, semantic knowledge, and machine learning, but avoid technological convergence, and foster competing solutions
  - demonstrate potential solutions by showcasing results in multilingual systems working in specific domains
  - help translation document production (improve “translatability” of source texts, terminology look-up, automatic language correction tools)
  - new evaluation protocols and metrics, e.g. for quality translations, for machine learning algorithms

Translation is hard, evaluating translations is even harder

And so ...

- **Application drivers** – media industry requirements such as sub-titling, multilingual information systems (showcases), cross-language question answering systems, location-based services, emergence of intelligent communicative agents, “machine” that follow-up discussions between humans (e.g. meeting transcription, air traffic control, etc.), personal digital memories
- Need to continue to **integrate work** on basic research, technology development (with performance evaluation) and application building and testing with industry, and foster an “Open Source” movement
- At the **European level** focus on infrastructure, coordination and R&D for multilingual systems – promote standards and portability, share systematically data, tools, computing, information, resources, as well as requirements and specifications for high-quality resources – support evaluation and performance assessment – build competences, develop service functions, demonstrate the outcome of research and the potential of the technologies – and support work on non-European languages

We need a place to meet and exchange experience

Further information

General FP6:

http://europa.eu.int/comm/research/fp6/

http://www.cordis.lu/

IST:

http://www.cordis.lu/ist

Mats Ljungqvist

email: mats.ljungqvist@cec.eu.int

tel: +352 4301 38078