META-NET: Towards a Strategic Research Agenda for Multilingual Europe

Georg Rehm
DFKI, Germany
georg.rehm@dfki.de

Multilingual Web Workshop – Limerick, Ireland
September 21, 2011

Co-funded by the 7th Framework Programme and the ICT Policy Support Programme of the European Commission through the contracts T4ME, CESAR, METANET4U, META-NORD (grant agreements no. 249119, 271022, 270893, 270899).
Outline

- Introduction
- The META-NET Language White Paper Series
- Towards a Strategic Research Agenda for Multilingual Europe
Challenge: Providing each language community with the most advanced technologies for communication and information so that maintaining their mother tongue does not turn into a disadvantage.

Research has made considerable progress in recent years.

But: the pace of progress is not fast enough to meet the challenge within the next 10-20 years.

All stakeholders – researchers, LT user and provider industries, language communities, funding programmes, policy makers – should team up for a major dedicated push.
META-NET is a network of excellence dedicated to fostering the technological foundations of the European multilingual information society.

**META-VISION:** Building a community with a shared vision and strategic research agenda

**META-SHARE:** Building an open resource exchange infrastructure

**META-RESEARCH:** Building bridges to neighbouring technology fields

http://www.meta-net.eu
Four Funded Projects

- Initial project: T4ME (FP7; 13 partners, 10 countries; 7 Mio. €)
- Three new support consortia (ICT-PSP) started in February 2011.
- All EU member states and several non-member states covered.
- META-NET in September 2011: 47 members from 31 countries.

http://www.meta-net.eu
http://www.meta-net.eu/members
- **META-NET** is a network of excellence.
- **META** is an open and growing strategic technology alliance: Multilingual Europe Technology Alliance.
  - Almost 300 members, including W3C, Google, Microsoft, GALA, research centres, LT companies etc.
  - META includes multiple stakeholders to prepare the ground for a large-scale concerted effort.
  - Main goal: to support the Strategic Research Agenda.
META-VISION

The META-NET Language White Paper Series

http://www.meta-net.eu
The Language White Papers

- LT support varies greatly from language to language.
- Inform about the current status and availability of LRs and LTs.
- Survey of the state of ca. 30 languages in the digital society.
- Target audience: politicians, journalists, decision makers, the public at large.
- Key messages: societal and technological problems, challenges, economic opportunities.

http://www.meta-net.eu
Structure of the White Papers

- Executive Summary
- Part 1: Introduction – A Risk for Our Languages and a Challenge for LT
- Part 2: Language in the European Information Society
- Part 3: LT Support for Language
- Part 4: About META-NET
- References

http://www.meta-net.eu
30 Languages Covered so far

- Basque
- Bulgarian*
- Catalan
- Czech*
- Danish*
- Dutch*
- English*
- Estonian*
- Finnish*
- French*
- Galician
- German*
- Greek*
- Hungarian*
- Icelandic
- Irish*
- Italian*
- Latvian*
- Lithuanian*
- Maltese*
- Norwegian
- Polish*
- Portuguese*
- Romanian*
- Serbian
- Slovak*
- Slovene*
- Spanish*
- Swedish*
- Croatian

* = Official EU language
Assessing LT Support

- Experts provided estimations, condensed several times, aggregated in a table assessing core technology areas and resources.

- Individual tables with tools and resources provide data for each language (existing tools, gaps etc.).

- Results for each application area and resource type were derived from two features (*quality, coverage*), resulting in a *big* table:

```
<table>
<thead>
<tr>
<th>Application Area</th>
<th>Resources Type</th>
<th>Language</th>
<th>Quality</th>
<th>Coverage</th>
</tr>
</thead>
<tbody>
<tr>
<td>Discourse Processing</td>
<td>Text Semantics</td>
<td>English</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Reference Corpora</td>
<td>Speech Synthesis</td>
<td>Basque</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Language Resources</td>
<td>Machine Translation</td>
<td>French</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Language Technology</td>
<td>Information Extraction</td>
<td>Italian</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Speech Corpora</td>
<td>Information Retrieval</td>
<td>Greek</td>
<td>2</td>
<td>2</td>
</tr>
<tr>
<td>Multimedia and Multimodal Data</td>
<td>Summarization</td>
<td>Hungarian</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Lexicons, Terminologies</td>
<td>Advanced Discourse Processing</td>
<td>Maltese</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>Morphological Analysis/Generation</td>
<td>Text Generation</td>
<td>Norwegian</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Multimodal and Multimedia Data</td>
<td>Event/Relation Extraction</td>
<td>Polish</td>
<td>3</td>
<td>3</td>
</tr>
<tr>
<td>Discourse Corpora</td>
<td>Opinion/Sentiment Recognition</td>
<td>Portuguese</td>
<td>4</td>
<td>4</td>
</tr>
<tr>
<td>Semantics Corpora</td>
<td>Rhetorical Structure/RST</td>
<td>Romanian</td>
<td>2</td>
<td>2</td>
</tr>
</tbody>
</table>
```

http://www.meta-net.eu
For journalists and politicians the big table is useless.

Solution is a cluster-based approach: (a) Speech; (b) Machine Translation; (c) Text Analysis; (d) Resources.

- **Cluster 1: excellent LT support**
  - Technologies are in widespread use showing human-quality performance.

- **Cluster 2: good support**
  - Technologies exist; reasonable quality and performance

- **Cluster 3: medium support**
  - Research prototypes, quality and performance varies

- **Cluster 4: low to almost no support**
  - Drawing board or rudimentary prototypes; very limited quality and performance

http://www.meta-net.eu
## Cluster: Speech

<table>
<thead>
<tr>
<th>Cluster 1: excellent support</th>
<th>Cluster 2: good support</th>
<th>Cluster 3: medium support</th>
<th>Cluster 4: low/no support</th>
</tr>
</thead>
<tbody>
<tr>
<td>English, French, German, Spanish, Italian, Dutch, Czech, Danish, Portuguese, Finnish</td>
<td>Basque, Bulgarian Catalan, Croatian Estonian, Galician Greek, Hungarian Polish, Serbian Slovene, Swedish</td>
<td>Icelandic, Irish Latvian, Lithuanian, Maltese, Norwegian Romanian, Slovak</td>
<td></td>
</tr>
</tbody>
</table>
Cluster: Machine Translation

<table>
<thead>
<tr>
<th>Cluster 1: excellent support</th>
<th>Cluster 2: good support</th>
<th>Cluster 3: medium support</th>
<th>Cluster 4: low/no support</th>
</tr>
</thead>
<tbody>
<tr>
<td>English</td>
<td>Spanish, Catalan, German, Italian, French, Czech</td>
<td>all other languages</td>
<td></td>
</tr>
</tbody>
</table>
A Few General Observations

- Speech processing and synthesis are more mature than processing of written text.
- Most (very) large companies have stopped working in LT, leaving the field to SMEs, which can hardly attack an international market.
- Draft versions available at http://www.meta-net.eu/whitepapers
- Comments? Suggestions for extensions?
META-VISION

Towards a Strategic Research Agenda for Multilingual Europe

http://www.meta-net.eu
Shared Vision and SRA

- Mobilize researchers, decision makers, users and providers of LT, R&D programmes for cooperation and collaboration in META.
- Large-scale joint action to
  - Building a community around Language Technology in Europe (META)
  - Creating a shared vision
  - Preparing a Strategic Research Agenda for Multilingual Europe

http://www.meta-net.eu
From Visions to the SRA

- Three **Vision Groups** bring together researchers, developers, integrators and (corporate or professional) users of LT-based products, services and applications (ca. 25 members each).

- **Vision Groups** collect domain-specific visions, prepare individual reports and provide them to the META Technology Council.

- **Vision Group**
  - **Translation and Localisation**
    - July 23, 2010, Berlin, Germany
    - September 28, 2010, Brussels, Belgium
    - April 7/8, 2011, Prague, Czech Republic

- **Vision Group**
  - **Media and Information Services**
    - September 10, 2010, Paris, France
    - October 15, 2010, Barcelona, Spain
    - April 1, 2011, Vienna, Austria

- **Vision Group**
  - **Interactive Systems**
    - September 10, 2010, Paris, France
    - October 5, 2010, Prague, Czech Republic
    - March 28, 2011, Utrecht, The Netherlands

http://www.meta-net.eu
META Technology Council prepares two documents:


- Strategic Research Agenda for Multilingual Europe.
  - To be presented to national and international politicians, administrators and funding agencies.
  - Will cover a timeframe from now to ca. 2025
  - Work in progress.
Visions for Multilingual Europe

- Language-Transparent Web and Media
  - The web is multilingual and multimedia.
  - Multimedia multi-language subtitling.

- Natural and Inclusive Interaction
  - Digital communication does not have any borders.
  - Cross-lingual meeting assistants that support speech-to-speech translation.

- Efficient Information Management
  - Information is growing without limits.
  - Federated multilingual audio-visual search.
The Planning Process

2010

Within META-NET (META-VISION)

Communication within META-NET (META-VISION)

Communication in the wider LT community

Communication to policy makers, funding bodies, public

and among other stakeholders

Visions

2011

2012

today

http://www.meta-net.eu
Towards the SRA

- Many suggestions by the Vision Group members.
- Additional input in meetings, workshops, discussions etc.
- We screened the Strategic Research Agendas of other initiatives.
- We discussed procedures, input and structure of the SRA in two meetings of the META Technology Council.
  - Brussels, Belgium, November 16, 2010
  - Venice, Italy, May 25, 2011
  - Berlin, Germany, September 30, 2011 (upcoming)

http://www.meta-net.eu
Letter from the META-NET Partners
META: Multilingual Europe Technology Alliance
1. Executive Summary
2. Multilingual Europe: Facts, Challenges, Opportunities
3. ICT: Current State, Major Trends and Predictions
4. Language Technology: State, Limitations, Potential
5. Language Technology for Multilingual Europe: The Grand Vision
6. Language Technology for Multilingual Europe: Priorities, Plans, Roadmap
7. References
8. List of Contributors
Get Involved!

- Provide feedback and new ideas in the discussion forum at http://www.meta-net.eu/forum
- The Multilingual Europe Technology Alliance needs as many members as possible. Join us! http://www.meta-net.eu/join
- Approach your friends and colleagues from other organisations and ask them also to join META!
- Get involved and have your say!
- Tomorrow: Breakout session about SRA! Let’s hear your suggestions for key slogans, solution visions – and more!
Thank you very much!

Joint work with Aljoscha Burchardt, Kathrin Eichler, Felix Sasaki, Hans Uszkoreit, the ca. 70 members of the Vision Groups, the 30 members of the META Technology Council and the ca. 140 authors of and contributors to the META-NET Language White Papers.