

META-NET

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.**



Objectives

META-NET is a network of excellence dedicated to fostering the technological foundations of the European multilingual information society.

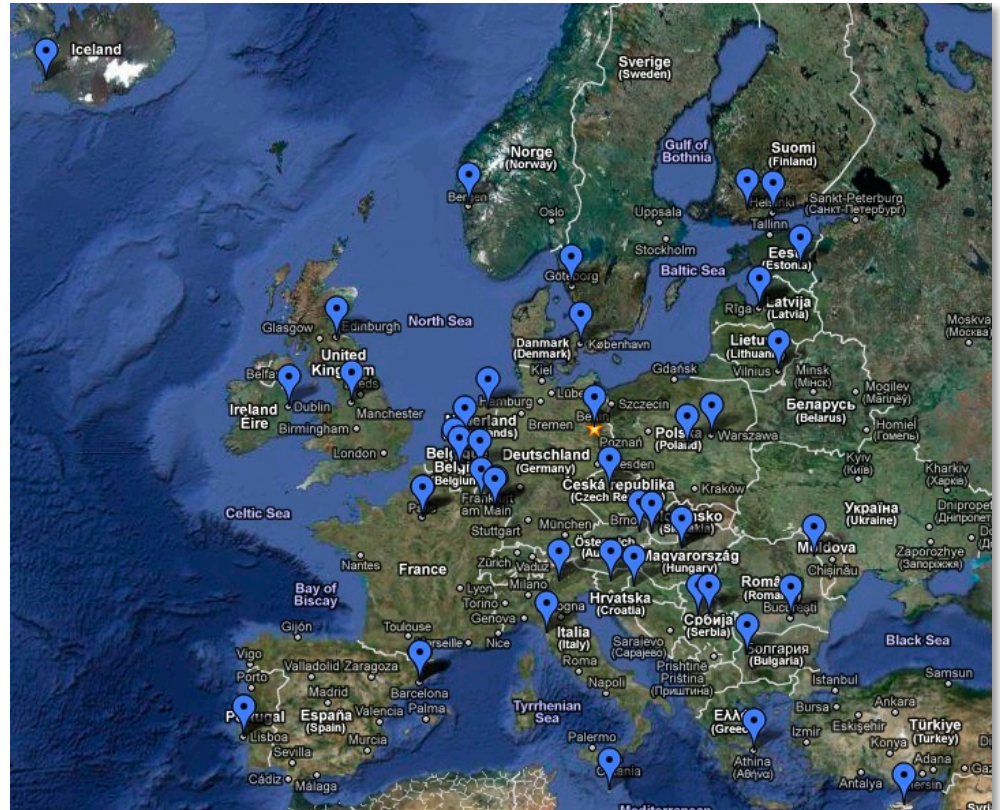
META-VISION: Building a community with a shared vision and strategic research agenda

META-SHARE: Building an open resource exchange infrastructure

META-RESEARCH: Building bridges to neighbouring technology fields

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.



- ❑ **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.
 - Join us! <http://www.meta-net.eu/join>



META-VISION

The META-NET Language White Paper Series

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.

META-NET White Paper Series

Languages in the European Information Society

– Latvian –

Early Release Edition

META-FORUM 2011

27-28 June 2011

Budapest, Hungary



The development of this white paper has been funded by the Seventh Framework Programme and the ICT Policy Support Programme of the European Commission under contracts T4ME (Grant Agreement 249119), CESAR (Grant Agreement 271022), METANET4U (Grant Agreement 270893) and META-NORD (Grant Agreement 270899).

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



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:

Language Technology (Tools, Technologies, Applications)	Basque	Bulgarian	Catalan	Croatian	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
Tokenization, Morphology (tokenization, POS tagging)	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
Parsing (shallow or deep syntactic analysis)	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
Sentence Semantics (WSD, argument structure, semantic roles)	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
Text Semantics (coreference resolution, context, pragmatics)	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Advanced Discourse Processing (text structure, coherence, information retrieval/text indexing, multimedia IR, crosslingual)	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Information Extraction (named entity recognition)	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
Language Generation (text/answer generation, report generation)	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
Summarization, Question Answering, advanced Information	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
Machine Translation	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
Speech Recognition	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Speech Synthesis	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
Dialogue Management (dialogue capabilities and user language resources (Resources, Data, Knowledge Bases))	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
Reference Corpora	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
Syntax-Corpora (treebanks, dependency banks)	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
Semantics-Corpora	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Discourse-Corpora	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
Parallel Corpora, Translation Memories	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
Speech-Corpora (raw speech data, labelled/unannotated speech)	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
Multimedia and multimodal data	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
Language Models	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2
Lexicons, Terminologies	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5	5
Grammars	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3	3
Thesauri, WordNets	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4	4
Ontological Resources for World Knowledge (e.g. upper)	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2	2

- ❑ 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

Cluster: Speech

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

Cluster: Machine Translation

Cluster 1: <i>excellent support</i>	Cluster 2: <i>good support</i>	Cluster 3: <i>medium support</i>	Cluster 4: <i>low/no support</i>
	English	Spanish, Catalan, German, Italian, French, Czech	<i>all other languages</i>

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

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**



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

From Visions to the SRA

- **META Technology Council** prepares two documents:
 - **“The Future European Multilingual Information Society”
Vision Paper for a Strategic Research Agenda**
<http://www.meta-net.eu/vision/reports/meta-net-vision-paper.pdf>
 - **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.*



- ❑ **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



- ❑ 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)

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!**

Q/A

META  NET

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.

office@meta-net.eu

<http://www.meta-net.eu>

<http://www.facebook.com/META.Alliance>