META≡NET

Speaking one's mother tongue, be it Latvian, Hungarian, or Portuguese must not become a social or economic disadvantage in the networked European society of the 21st century. Language Technology (LT) has the potential to become the key solution to this challenge if it is robust, cost-effective and available for all European languages and to all European citizens. In order to achieve these goals the pace of research and development has to be accelerated by a major, dedicated push. A push on this level and with the magnitude needed can only be put into motion through a joint action of all stakeholder groups involved. This highly diverse group consists of researchers, user and provider industries, technology integrators, language communities, administrators, politicians, journalists and society in general.

META-NET is a Network of Excellence forging the Multilingual Europe Technology Alliance (META) through a concerted effort to build a strong European community around LT. Its goal is to prepare the grounds for applications that enable automatic translation, multilingual information and knowledge management as well as content production across all European languages. This effort will enhance the development of intuitive language-based interfaces to technology ranging from household electronics, machinery and vehicles to computers and robots. The objective is to advance LT so that communication and cooperation across languages becomes possible and to secure users of any language equal access to information and knowledge.

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

We aim to further LT as a means towards realising the vision of a Europe united as one single digital market and information space. META-NET is supporting these goals by pursuing three lines of action:

1. META-VISION. This line of action is concerned with bringing together the fragmented European LT community and with collaboratively preparing, establishing and promoting a Strategic Research Agenda (SRA) for European LT. The SRA is meant to be a long-term instrument. It will cover the period until 2020 and will be an umbrella for industrial and academic research. The SRA will contain recommendations and suggestions for joint actions to be presented to the European Commission and national as well as regional bodies. The complex process of preparing the SRA includes Vision Groups that consist of external experts from industry, who provide ideas for innovative LT applications and scenarios for the future knowledge society. The META Technology Council consolidates their findings into the SRA.

2. META-SHARE. META-NET is building META-SHARE, a sustainable network of repositories of language data, tools and web services documented with high-quality metadata, aggregated in central inventories allowing for uniform search and access to resources. Data and tools can be open and with restricted access rights, free and for-a-fee. META-SHARE targets existing but also new and emerging language data, tools and systems required for building and evaluating new technologies, products and services. Reuse, combination, repurposing and re-engineering of language data and tools are crucial. META-SHARE will be an important component of an LT marketplace for researchers and developers, language professionals, as well as for industrial players including SMEs and big enterprises.

3. META-RESEARCH. This line of action is carrying out innovative Machine Translation (MT) research by building bridges to other fields and disciplines such as Machine Learning and the Semantic Web community. Central goals are bringing more semantics into MT, optimising the division of labour in hybrid MT, exploiting the context for Translation and preparing an empirical base for MT. Research includes, among others, the implementation of prototypes, the preparation of data sets for evaluation purposes and the organisation of internal workshops. This research is meant to advance significantly the state of the art in Machine Translation.

Activities and Results from the Second Year

The highlights of the second project year include key achievements in all three lines of action. As a prerequisite for reaching its ambitious goals, the **META-VISION** arm of the project has been building up a coherent and homogeneous LT community in Europe by bringing together representatives from various stakeholder groups (researchers, user industries, provider industries, administrators, politicians, integrators etc.) at various meetings, events and conferences organised by both META-NET and in cooperation with others from the community. Significant steps towards this goal have been taken through means such as, for example, successfully mobilising ca. 90 external participants for Vision Groups in the areas Translation and Localisation, Media and Information Services and Interactive Systems and the META Technology Council. META-NET has been engaged in intense dissemination and communication throughout the project's work to date. In 2011, these include (among others) presentations at the W3C Workshop "Content on the Multilingual Web" (April, Italy), FLaReNet Forum (May, Italy), Attensity Forum (May, Berlin), Media for All (June, UK), RANLP (September, Bulgaria), the EFNIL Conference (October, UK) – according to our estimates we already reached more than 3,000 LT professionals, stakeholders in related fields and language community representatives and informed them about META-NET's goals. In addition to the wide visibility provided by successful dissemination and mobilisation activities, META-NET organised or co-organised several conferences of its own as well as a number of workshops co-located with other conferences and events: co-organised TRALAGY 2011 (March, Paris), a workshop and challenge "Context in Machine Translation" at ICANN 2011 (June, Espoo), the EUROLAN 2011 Summer School (August, Romania), shared task ML4HMT 2011 at the workshop on Using Linguistic Information for Hybrid MT 2011 (September, Spain), a workshop on

"Language Resources, Technologies, and Standards in the Sharing Paradigm" at IJCNLP 2011 (November, Thailand). The main event was META-FORUM 2011 (June 27/28, Budapest, Hungary), which attracted over 300 participants. A satellite event, META-EXHIBITION, featured more than 40 exhibits in total, half of them industry exhibits, the other half posters and demos from collaborating projects. At META-FORUM 2011 delegates met to learn more about the work of the Vision Groups, to discuss the Vision Paper and SRA, and received early editions of the Language White Paper Series. The programme also featured presentations and discussions on META-SHARE and emerging research work in the field.

In 2011 **META-SHARE** has developed on all fronts, including crystallising such finer points as: the language resources metadata model which describes the resources shared, the functionalities of the platform itself, the legal framework, and the basic operational procedures of the infrastructure. Versions 1.0 and 1.1 of the META-SHARE software have been implemented and rolled out to the META-NET Network, setting up 13 repositories and populating them with language related datasets and processing tools. These developments follow a revised work-plan to accommodate both the findings at the first review and the needs of the extended META-NET network.

Three Working Groups (WGs), on metadata, IPR and legal issues, and the technical infrastructure have been established pulling together experts and relevant stakeholders from the wider community. The WGs have formulated initial proposals and specifications for each respective area and discussed and validated them in a series of meetings and workshops. A workshop dedicated to presenting, discussing and offering a hands-on experience to META-SHARE was organised in October 2011. Work on the metadata model that was initiated in the first year, has been completed, now catering for all resource types, different types of datasets and tools. Mapping tools from existing schemas to the one of the platform have been developed, focusing first on creating a mapping from the ELRA catalogue to META-SHARE. META-SHARE v1.0 was demonstrated at META-FORUM 2011. This version together with detailed documentation and a user manual was released in September 2011 and tested by the META-NET network. A release version 1.1 was later added to accommodate the requirements of the collaborating projects, in view of their first population round in November 2011. By the end of November 2011, 13 distinct META-SHARE repositories had been installed, with a total of 1277 resources. META-SHARE has been presented at FLaReNet Forum, IJCNLP 2011, LTC 2011, EFNIL 2011, a wide range of workshops, national and regional events, and naturally META-FORUM 2011.

META-RESEARCH made further progress in four key areas, "Semantics in MT", "Hybrid MT", "Context in MT", and "Novel use of large data resources". This work has produced several high quality publications (with several more submitted to conferences in 2012), the most visible ones being those presented at WMT 2011 in Edinburgh. Moreover, META-RESEARCH has completed several crucial datasets in the area of semantically annotated parallel and multilingual resources:

the parallel Czech-English treebank and large automatically annotated parallel corpus (100M words). These datasets have been released under the Creative Commons license and are available through META-SHARE.

Work continued to develop hybrid methods in translation. Intensive effort has been devoted to the preparation of the ML4HMT Challenge. This was a first for this area of research in bringing such a challenge to the Machine Learning community. Work has also continued on the preparation of the Context in Machine Translation Challenge, to be organised jointly with the ICANN 2012 conference. The data is ready for the challenge; the task is to use contextual knowledge in selecting best translation from an n-best list of candidates. Furthermore, work on hierarchical phrase-based translation models enriched with syntactic and dependency features has continued. Results have been presented in December 2011 at the IWSLT workshop.

Expected Final Results, Impact and Use

We expect META-NET to have a significant, unprecedented long-term impact on the European LT landscape. The three lines of action all aim at the same goal, albeit on different levels: to provide technological antidotes for the language barriers Europe has been facing for quite some time, in the form of robust, precise, high-performance, multilingual LT, to foster the development of novel applications, to assemble and to strengthen the European LT community and to raise awareness about the enormous potential of LT for the European information society and the society at large.

In META-RESEARCH, innovative MT approaches are being developed in close collaboration with neighbouring fields, establishing and strengthening ties to other research areas. META-SHARE is taking care of closing a gap that has been wide open for decades. The successful implementation of a distributed peer-to-peer network for the distribution and sharing of language resources and technologies will enable researchers, SMEs and large industrial players quickly to build systems that employ monolingual, crosslingual or multilingual LT. META-VISION helps to provide a long-term plan for our mission of realising a truly multilingual Europe, which is based on robust LT. An important objective of META-NET is to assemble a coherent, convincing, powerful community around LT, i.e., researchers, industrial players, journalists, politicians, administrators, language communities and several others. This community building effort along with the process of assembling Vision Groups as well as the META Technology Council is progressing very successfully. One of the key deliverables of META-VISION, the Strategic Research Agenda, was discussed at META-FORUM 2011. The final SRA will be presented publicly at META-FOURM 2012. It will contain visions, plans and roadmaps for the path to a truly multilingual Europe, realised through high-performance, robust and precise LT.

Coordination & Contact

Prof. Dr. Hans Uszkoreit,
Dr. Georg Rehm
DFKI GmbH, LT Lab
Alt-Moabit 91c
10559 Berlin, Germany
georg.rehm@dfki.de

Multilingual Europe Technology Alliance