

# Multilingual Europe: A challenge for language tech

*Georg Rehm & Hans Uszkoreit*

Speaking one's mother tongue, be it Latvian, Hungarian or Portuguese, must not become a social or economic disadvantage in the networked European information society of the twenty-first century. Many European languages run the risk of becoming victims of the digital age, as they are under-represented and under-resourced. Huge regional market opportunities remain untapped because of language barriers.

Language technology has the potential to become the solution to this crucial challenge if it is robust, cost-effective and available for all European languages and to all European citizens. However, in order to achieve these goals, the pace of research and development has to be accelerated by means of a major, dedicated push.

The key to protecting and furthering the highly heterogeneous group of more than 60 European languages is language technology, though the degree to which it is used in Europe varies enormously from language to language. Research in this area has made considerable progress in the last few years, but unfortunately, the current pace of technological progress is much too slow to arrive at substantial software products that are able to move communication in a multilingual environment significantly forward within the next 10 to 20 years. Those basic technologies that are already widely used are usually monolingual and only available for a handful of languages. Services that are available online, such as Google Translate or Bing Translator, are helpful when it comes to getting a rough idea of what a document in a foreign language is about, but these applications are fraught with difficulties. Applications such as language and voice-based user interfaces or dialogue systems are used only in specialized domains and exhibit limited performance. For potentially life-saving technologies, such as rescue operations technology and robotics in the health-care sector, the importance of accurate translations cannot be overemphasized.



*Georg Rehm works as a senior consultant at DFKI in Berlin.*

## Introducing META-NET

META-NET is a European network of excellence consisting of 44 language technology research centers based in 31 countries, forging the Multilingual Europe Technology Alliance (META) through a concerted effort to build a strong and powerful European community for and around language technology. Its goal is to prepare the grounds for multilingual applications that enable automatic translation, information and knowledge management, including localization, as well as content production and applications in related areas across all European languages. META-NET, which started work on February 1, 2010, aims to advance research in language technology as a means towards realizing the vision of a Europe united in a single digital market and information space and is supporting these goals by pursuing three lines of action: META-VISION, META-SHARE and META-RESEARCH.

META-VISION is concerned with a goal that is not only important but strategically indispensable for the overall success of the initiative: building up a coherent and homogeneous community by bringing together representatives from the highly fragmented and heterogeneous stakeholder groups. According to our estimates, we have already reached more than 2,500 language technology professionals and informed them about META-NET's goals. For example, at Translingual Europe 2010 (June 7, Berlin) researchers discussed current problems and visions with representatives of the provider industries (such as Microsoft, Asia Online and ProMT) and language technology as well as machine translation (MT) users (European Patent Office, Symantec, EC DGT). The main META-NET event within its first year was META-FORUM 2010 (November 17-18, Brussels, Belgium). At META-FORUM the initial results of the vision building process were showcased to more than 250 participants.

The second important goal of META-VISION is collaboratively – within and by the community – to prepare, establish



*Hans Uszkoreit is a professor of computational linguistics at Saarland University.*

and also promote a strategic research agenda, intended to be a long-term instrument that will serve as an umbrella for both industrial and academic research and development in the period leading up to 2020. It will contain high-level recommendations and suggestions for joint actions to be presented to the European Commission and national as well as regional bodies and funding agencies.

META-NET is also building META-SHARE, a sustainable peer-to-peer network of repositories of language data, tools and web services documented with high-quality metadata and aggregated in inventories allowing for uniform search and access to resources. Data and tools can be both 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. In this respect, reuse, combination, repurposing and re-engineering of language data and tools play a crucial role.

A fully functional prototype of META-SHARE was presented at the first annual conference, META-FORUM 2010. Among the important relevant components of the infrastructure is a universal metadata scheme for the description of language resources and language technologies. This metadata scheme is currently being prepared and discussed by a working group that consists of experts from within the initiative and several other European experts. At the same time, the landscape of language resources licensing is being explored thoroughly and, with the help of legal experts, a first set of licensing templates has been adopted and prepared. META-SHARE favors and aligns itself with the growing open data and open source tools movement, especially the Creative Commons Initiative.

META-RESEARCH focuses on bringing more semantics into MT, optimizing the division of labor in hybrid MT, preparing an empirical base for MT and exploiting the context when computing an automatic translation. To this end, META-NET is carrying out research by building bridges to other fields and disciplines such as machine learning and the Semantic Web community. META-RESEARCH is concerned with collecting data, preparing data sets and language resources for evaluation purposes, compiling inventories of tools and methods,

and organizing workshops and advanced training events for its staff members. Among its current major outcomes are the clear identification of issues in MT, in which semantics has shown potential to positively impact the state of the art, recommendations for approaching the problem of integrating semantic information in MT, and a list of tools and resources that could be employed for this purpose. A new language resource for MT, the Annotated Hybrid Sample MT Corpus, is currently being finalized. It provides data for the language pairs English-German, English-Spanish and English-Czech. A third important outcome is software for the collection of multilingual hidden-web corpora. The tool clusters news articles from different languages discussing the same topic or event and clusters pages identified as being translations of each other. The research that is carried out in this line of action is meant to advance significantly the state of the art in MT.

## Extension and impact

The initiative has a founding consortium that consists of 13 partners in ten countries. However, META-NET operates on a European level. This is why, in November 2010, the network was extended. The enlarged network consists of 44 partners in 31 countries (see the table) and was presented to the public for the very first time at META-FORUM 2010. Most of the new members participate in three EU-funded projects that have the mission of supporting the META-NET objectives by systematically collecting language resources and language technologies, curating and describing them with metadata records and making them

available through META-SHARE, mobilizing the communities in their respective countries and organizing general awareness raising activities. The three projects commenced their work on February 1, 2011 – exactly one year after the start of META-NET.

The goal of META-VISION is to provide a long-term plan for our mission of realizing a truly multilingual Europe. With a large and diverse community behind our goals, META can achieve the critical mass needed to really make a difference to how language technology can enable and secure multilingualism in Europe's future. To researchers, technologists, professionals and administrators developing, providing or using language technologies, META offers a unique opportunity to stay informed, contribute ideas or advocate on behalf of specific languages, while participating in expert discussions, working groups and planning activities that will shape Europe's linguistic future. Research and technology development projects are invited to join META-SHARE to access a pool of language resources and technologies while helping to plan and validate META services. Commercial enterprises are welcome to contribute their visions for products and services, to participate in META's planning process and to use META to grow profitable partnerships. Schools and educators, journalists and the media, politicians, public institutions and organizations are encouraged to participate in open discussions on the vision of and way towards a truly multilingual information society. Your voice is important, just like the language in which you express yourself.

To learn more, or find out how you can participate, visit [www.meta-net.eu](http://www.meta-net.eu) **M**



**Innovation driven LSP**

**Full MT & PE cycle**

**Multilingual Content Solutions**

**Customized Statistical MT Solutions**

**Translation Workflow Optimization**

*in association with*  
**B.I Japan**

Trade Center, Profesor Beltrán Báuena 4, 46009 Valencia - Spain  
Tel: +34 96 338 5771 - Fax: +34 96 338 5772

**Valencia - Tokyo - Shanghai**