

At Least 21 European Languages in Danger of Digital Extinction

Study by Europe's Leading Language Technology Experts Warns Most European Languages Unlikely to Survive in the Digital Age

Most European languages face digital extinction, a new study by Europe's leading Language Technology experts finds. Assessing the level of support through language technology for 30 of the approximately 80 European languages, the experts conclude that digital support for 21 of the 30 languages investigated is "non-existent" or "weak" at best. The study was carried out by META-NET, a European network of excellence that consists of 60 research centres in 34 countries.

The study, prepared by more than 200 experts and documented in 30 volumes of the META-NET White Paper Series (available both online and in print), assessed language technology support for each language in four different areas: automatic translation, speech interaction, text analysis and the availability of language resources. A total of 21 of the 30 languages (70%) were placed in the lowest category, "support is weak or non-existent" for at least one area by the experts. Several languages, for example, Icelandic, Latvian, Lithuanian and Maltese, receive this lowest score in all four areas. On the other end of the spectrum, while no language was considered to have "excellent support", only English was assessed as having "good support", followed by languages such as Dutch, French, German, Italian and Spanish with "moderate support". Languages such as Basque, Bulgarian, Catalan, Greek, Hungarian and Polish exhibit "fragmentary support", placing them also in the set of high-risk languages.

"The results of our study are most alarming. The majority of European languages are severely under-resourced and some are almost completely neglected. In this sense, many of our languages are not yet future-proof.", says Prof. Hans Uszkoreit, coordinator of META-NET, scientific director at DFKI (German Research Center for Artificial Intelligence) and co-editor of the study. The study's other co-editor, Dr. Georg Rehm (DFKI), adds: "There are dramatic differences in language technology support between the various European languages and technology areas. The gap between 'big' and 'small' languages still keeps widening. We have to make sure that we equip all smaller and under-resourced languages with the needed base technologies, otherwise these languages are doomed to digital extinction."

The field of language technology produces software that can process spoken or written human language. Well-known examples of language technology software include spell and grammar checkers, interactive personal assistants on smartphones (such as Siri on the iPhone), dialogue systems that work over the phone, automatic translation systems, web search engines, and synthetic voices used in car navigation systems. Today language technology systems primarily rely on statistical methods that require incredibly large amounts of written or spoken data. Especially for languages with relatively few speakers it is difficult to acquire the needed mass of data. Furthermore, statistical language technology systems have inherent limits in their quality, as can be seen, for example, in the often amusing incorrect translations produced by online machine translation systems.

Europe has succeeded in removing almost all borders between its countries. One border still exists, however, and it seems to be impenetrable: the invisible border of language barriers is one that hinders the free flow of knowledge and information. It also harms the long-term goal of establishing a single digital market because it hinders the free flow of goods, products, and services. While language technology has the potential to get rid of language barriers through modern machine translation systems, the results of the META-NET study clearly show that many of European languages are not yet ready. There are significant gaps in technology due to the English-language focus of most R&D, a lack of commitment and financial resources, and also a lack of a clear research and technology vision.

A coordinated, large-scale effort has to be made in Europe to create the missing technologies as well as transfer technology to the majority of languages. There are strong reasons for approaching this immense challenge in a community effort involving the European Union, its member states and associated countries, as well as industry. These reasons include the high per-capita financial burden for smaller language communities; the needed transfer of technologies between languages; the lack of interoperability of resources, tools, and services; and the fact that linguistic borders often do not coincide with political borders. Europe must take action to prepare its languages for the digital age. They are a precious component of our cultural heritage and, as such, they deserve future-proofing. On September 26 the Council of Europe marks the European Day of Languages, a day which recognises the importance of fostering and developing the rich linguistic and cultural heritage of our continent. META-NET's work is a stark reminder of the challenges and possibilities facing our linguistic heritage in the information age.

Language Technology: Background

Language technology already supports us in everyday tasks, such as writing e-mails or buying tickets. We benefit from language technology when searching for and translating web pages, using a word processor's spell and grammar checking features, operating our car's entertainment system or our mobile phone with spoken commands, getting recommendations in an online book-store, or following the instructions spoken by a mobile navigation app. In the near future, we will be able to talk to computer programs as well as machines and appliances, including the long-awaited service robots that will soon enter our homes and work places. Wherever we are, when we need information, we will simply ask for it, and, when we need help, we will demand it out loud. Removing the communication barrier between people and technology will change our world.

Language technology is generally acknowledged today as one of the key growth areas in information technology. Large international corporations such as Google, Microsoft, IBM, and Nuance have invested substantially in this area. In Europe, hundreds of small and medium enterprises have specialized in certain language technology applications or services. Language technology allows people to collaborate, learn, do business, and share knowledge across language borders and independently of their computer skills.

The META-NET White Paper Series

The META-NET White Paper series "Languages in the European Information Society" reports on the state of 30 European languages with respect to Language Technology and explains the most urgent risks and chances. The series covers all official EU Member State languages and several other languages spoken in Europe. While there have been a number of valuable and comprehensive scientific studies on certain aspects of languages and technology, until now there has been no generally understandable compendium that presents the main findings and challenges for each language with regard to a technology-supported multilingual Europe. The META-NET White Paper Series fills this gap. META-NET can now show why most languages face serious problems and pinpoint the most threatening gaps. In total, more than 200 authors and contributors helped preparing the Language White Papers.

The white papers were written for the following European languages: Basque, Bulgarian, Catalan, Croatian, Czech, Danish, Dutch, English, Estonian, Finnish, French, Galician, German, Greek, Hungarian, Icelandic, Irish, Italian, Latvian, Lithuanian, Maltese, Norwegian (bokmål and nynorsk), Polish, Portuguese, Romanian, Serbian, Slovak, Slovene, Spanish, and Swedish. Each Language White Paper is written in the language it reports upon and includes a complete English translation.

About META-NET and META

META-NET, a Network of Excellence consisting of 60 research centres from 34 countries, is dedicated to building the technological foundations of a multilingual European information society. META-NET is co-funded by the European Commission through a total of four projects.

META-NET is forging META, the Multilingual Europe Technology Alliance. More than 600 organisations from 55 countries, including research centres, universities, small and medium companies as well as several big enterprises, have already joined this open technology alliance.

Contact:

Prof. Dr. Hans Uszkoreit
Dr. Georg Rehm
META-NET Office c/o DFKI Berlin
Alt-Moabit 91c
10559 Berlin, Germany
Phone: +49 (0)30 23895-1833
Email: office@meta-net.eu

Background Information:

META-NET: <http://www.meta-net.eu>
White Papers: <http://www.meta-net.eu/whitepapers>