

Most European languages in danger of digital extinction

September 25 2012

Scientists from The University of Manchester were part of a European team of researchers who concluded that digital assistance for 21 of the 30 languages investigated is 'non-existent' or 'weak' at best.

The report coincides with the European Day of Languages [September 26], a day which recognises the importance of fostering and developing the rich linguistic and [cultural heritage](#) of our continent.

Languages spoken by a small number of people could be at risk because they do not have technological support, the report by META-NET, a European network of excellence that consists of 60 research centres in 34 countries, including the University of Manchester's National Centre for Text Mining (NaCTeM), found.

Icelandic, Latvian, Lithuanian and Maltese are at the highest risk of disappearing, while other languages such as Bulgarian, Greek, Hungarian and Polish are also at risk. The researchers claim that 21 out of 30 European languages could become extinct in the digital world.

The study, prepared by more than 200 experts and documented in 30 volumes of the META-NET White Paper Series, assessed [language](#) technology support for each language in four different areas: automatic translation, speech interaction, text analysis and the availability of language resources.

Several languages, for example Icelandic, Latvian, Lithuanian and

Maltese, receive this lowest score in all four areas.

While English has the best language technology support amongst all European languages, it can still not be considered as "excellent support", but rather only "good support", the University of Manchester researchers found in the White Paper for English.

Dutch, French, German, Italian and Spanish are considered to have "moderate support". Languages such as Basque, Bulgarian, Catalan, Greek, Hungarian and Polish exhibit "fragmentary support", which also places them in the set of high-risk languages.

Language technology software is used to process spoken or written [human language](#). Well-known examples include spelling and grammar checkers, interactive personal assistants on smartphones (such as Siri on the iPhone), dialogue systems that work over the phone, [automatic translation](#) systems and web search engines.

The lack of available software for the high-risk languages means that without drastic action, they will be unable to survive in today's digital world.

Language technology systems primarily rely on statistical methods that require incredibly large amounts of written or spoken data - difficult to acquire for languages with relatively few speakers.

Furthermore, statistical language technology systems have inherent limits on their quality, as can be seen, for example, in the often amusing incorrect translations produced by online machine translation systems.

The report concludes that a coordinated, large-scale effort has to be made in Europe to create the missing technologies and transfer this technology to the languages faced with digital extinction .

Professor Sophia Ananiadou, director of NaCTeM, said: "In the UK, most of us use software that incorporates language technology without even realising it.

"Language technology already makes our lives easier and has huge potential to help us in many different ways. As digital information and communication is becoming increasingly dominant, it is vital that sophisticated language technology support is available for a wider range of languages, otherwise collaboration with our European neighbours will become more difficult."

Professor Hans Uszkoreit, coordinator of META-NET, said: "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.

Provided by University of Manchester

Citation: Most European languages in danger of digital extinction (2012, September 25) retrieved 25 April 2024 from

<https://phys.org/news/2012-09-european-languages-danger-digital-extinction.html>

This document is subject to copyright. Apart from any fair dealing for the purpose of private study or research, no part may be reproduced without the written permission. The content is provided for information purposes only.