

Researchers want to achieve machine translation of the 24 languages of the EU

July 7 2016



Professor of Translation-Oriented Language Technologies at Saarland University and a Scientific Director at the German Research Center for Artificial Intelligence. Credit: DFKI

The aim of their collaboration is to achieve machine-based translation

between the languages of the European Union so that comprehensible texts are achieved for as many language combinations as possible. Two of the EU-funded research projects are being led by the Saarbrücken computer linguist Josef van Genabith.

Anyone who wants to learn Finnish has to be prepared to deal with a complex grammar that includes fifteen different cases. The grammatical cases are marked in part by appending syllables to nouns resulting in a dizzying array of word forms and expressive possibilities. "Teaching a computer to understand all these grammatical nuances and to translate them correctly into another language is exceptionally difficult," says Josef van Genabith, Professor of Translation-Oriented Language Technologies at Saarland University and a Scientific Director at the German Research Center for Artificial Intelligence (DFKI). His team is therefore following a different path. The computers are not fed with grammar rules and linguistic details, but are taught to recognize patterns in huge text repositories and to learn from them. In the computer linguistics community, this approach is referred to as "deep learning". The method recently made headline news when Google used the technique to beat one of the world's top Go players.

"This machine learning strategy has nothing to do with natural intelligence, but it does have similarities with the processes that occur in the human brain when we control the muscles in our bodies. Children have to learn to pick up their feet when walking in the woods so as not to trip over roots or stones. In adults, this sort of mental process runs automatically in the background, as the brain has learnt how their feet have to be placed," explains van Genabith. Computers could also be trained to learn continuously in this way and to apply the knowledge so acquired. In the case of automatic translation, the focus is not on the structures that a student would learn from a grammar book, but on the patterns that the computer recognizes and acquires.

QT21 is a consortium of fourteen leading research institutions for [machine translation](#) in Europe and Hong Kong that includes universities, research institutes, such as DFKI, and numerous companies. "Our common goal is to exploit machine learning to significantly improve [automatic translation](#), particularly of more complex languages such as Latvian or Czech," says van Genabith, who heads the project, which was rolled out a year ago. The European Union has approved a total of 3.9 million euros for the three-year project, of which around one million has been allocated to Saarbrücken.

The European Language Resources Coordination (ELRC) is a second project lead by DFKI and Josef van Genabith in which a European consortium has been contracted by the European Commission to collect suitable language data sets that will enable the European Commission's automated translation platform (CEF AT) to be adapted and optimized for the daily requirements of public administrators in all EU Member States as well as Iceland and Norway. ELRC is one of the most comprehensive collections of language data worldwide.

"We are currently identifying all possible Language Resources covering related subject areas, such as texts and their translations from European government ministries in the areas of finance, economics, interior affairs and foreign affairs. These data sets help the European Commission to train the translation software and to adjust it to meet the requirements of public administrators and European citizens" explains van Genabith. The two-year project will receive 1.7 million euros in financial support from the EU Commission.

Improved translation software should also benefit trade within the European internal market. But that doesn't mean that translators are going to run out of work; quite the contrary, in fact, according to van Genabith: "Computers can translate huge quantities of text far faster than a human. But the translations won't be perfect, so that, depending

on requirements, translators will still need to post-edit the texts."

In a webinar, Josef van Genabith explained how statistically driven machine translation works. It can be viewed free of charge at:

<http://www.gala-global.org/ondemand/how-does-modern-machine-translation-work-story-pictures-not-math>

Research results in machine translation will be presented at the annual meeting of the Association for Computational Linguistics (ACL) which will take place August 7- 12, 2016 at the Humboldt University in Berlin (acl2016.org). The scientists will present the results of an international competition on machine translation in a workshop at the conference:

<http://www.statmt.org/wmt16>

More information: www.qt21.eu/

Provided by Saarland University

Citation: Researchers want to achieve machine translation of the 24 languages of the EU (2016, July 7) retrieved 26 April 2024 from <https://phys.org/news/2016-07-machine-languages-eu.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.