

Pushing back the boundaries of machine translation for health

June 13 2018



Credit: Andrea Piacquadio from Pexels

EU researchers have brought us a step closer to fully-automated machine translation with a neural-based system capable of translating texts on public health from English into Czech, German, Polish and Romanian.

Online information is often only available in a few languages as organisations cannot afford to translate it into more. But researchers from the EU-funded Health in My Language, or HimL project, have brought the prospect of fully automated [machine translation](#) a step closer, by working with Scottish and international [public health](#) organisations to produce a system adapted for the [health](#) domain.

"Immigrant communities may have limited command of the local language – they need information about local health services but it is not available in their language," says Barry Haddow, project co-ordinator and senior researcher in informatics at the University of Edinburgh. "Information about best practices in health care, resulting from recent research, is mainly disseminated in English but consumers would like to access new meta-analyses in their own language."

Deep learning

The HimL team researched quality improvements in machine translation and incorporated these into a new system able to work from English into Czech, German, Polish and Romanian. It started using a syntactic or phrase-based approach, but quickly moved to neural machine translation (NMT), an approach based on [deep learning](#) which emerged during the life of the project.

New versions were released each year for use by project partners NHS 24, the Scottish [national health service](#), and Cochrane, an NGO that facilitates access to the latest research on health matters. The results were carefully evaluated using user surveys and application-focused testing.

The improvements were made in three main areas; domain adaptation or tuning the translation to the specific terminology of public health; semantics or ensuring accuracy of translation; morphology or making

sure morphological variants are correctly produced.

"English doesn't have a lot of morphology, but a lot of languages in Europe, such as Czech and Polish, do – they have different verb and nouns forms according to use and, if you get it wrong, this can change the meaning of the text," says Dr. Haddow.

Users were asked to rank the results produced by HimL compared to a well-known online system. "Our systems were able to offer better results in all language pairs," says Dr. Haddow, "although the extremely high quality required by NHS 24 and Cochrane users means that we are not yet able to automate translation completely."

Less human intervention

The team also looked at how well the HimL systems performed when combined with post-editing – this approach uses machine translation to produce a rough first version, then gets a human translator to edit the result. "Cochrane showed that post-editing using the HimL system in the MateCat tool was 30-40 % faster than translation from scratch for all languages except for Polish," says Dr. Haddow. "We were able to reduce the amount of [human intervention](#) by between 30–50% to produce as good a [translation](#) as we would have achieved with the fully human approach."

Other outputs include the [UFAL medical corpus](#), a standard data set for training systems to deal with medical texts. It covers eight European [language](#) pairs, including the HimL ones.

Analysing the output of NMT showed that problems present in earlier systems have now been largely overcome, but that these systems are still prone to omitting important information or adding incorrect information. "To counter this we use a technique called "reconstruction",

where the source should be reconstructable from the output," says Dr. Haddow, "we have also shown how to improve NMT using high quality dictionaries and how to incorporate semantic and syntactic information from external tools."

Provided by CORDIS

Citation: Pushing back the boundaries of machine translation for health (2018, June 13) retrieved 9 April 2024 from <https://medicalxpress.com/news/2018-06-boundaries-machine-health.html>

<p>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.</p>
--