

The HimL project aims to make public health information available in a wider variety of languages. We will do this using fully automatic machine translation, combining the statistical paradigm with deep linguistic techniques. The translation system will be adapted to the domain and integrated into the live systems run by [NHS 24](#) and [Cochrane](#).

WHAT ARE THE BIGGEST BENEFITS TO PROJECT PARTNERS?

The HimL project will carry out work that is vital to **Cochrane's** organisational strategy, and will increase accessibility of Cochrane's valuable medical information.

In the short-term, the availability of high quality fully automatic machine translation will:

- Allow Cochrane to disseminate its information to the language communities in the project.
- Increase impact and capacity in Eastern Europe, which is a region that is currently underrepresented and therefore a priority within Cochrane's strategic plan.

In the longer-term, success of the HimL Innovation Action will:

- Enable Cochrane to determine how to begin producing information in many more languages than it currently serves.
- Help achieve its vision of a world of improved health where decisions about health and health care are informed by high-quality, relevant and up-to-date synthesized research evidence.

On behalf of **NHS Scotland**, NHS 24 provides health and care information across a number of digital channels to the population of Scotland. Together with the other national health services in the United Kingdom, NHS Scotland is one of the most recognised and respected health care providers in the world.

The work of HimL fits with their remit to **expand their multi-lingual content** offering meeting the communication and language needs of their audience. NHS 24 is committed to exploring all opportunities in making NHS inform, Scotland's national health information service accessible for all and are equally committed to deploying, marketing and evaluating the EU HimL machine translation.

THREE IMPORTANT PIECES OF WORK DONE IN THE FIRST 6 MONTHS:

1. Creating test sets

- We have focussed significant resources on extracting representative texts from our use case partners.
- We then professionally translated these texts into our 4 target languages: German, Czech, Polish and Romanian.

Newsletter 2/2015

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 644402.



- We collected 30,000 words from both the NHS24 website and from Cochrane review summaries.

The test set is crucial for both tuning our translation model settings, and for testing the accuracy of the translations using automatic metrics. Although we will perform extensive human evaluation, automatic metrics are necessary for us to understand how incremental changes in our system have affected performance as they are cheap, fast and reproducible. The test set will be used to help verify advances we make in domain adaptation, semantically aware MT and in dealing with target languages like Czech which have extremely rich morphology.

2. The year 1 systems

We have created the Year 1 translation systems for HimL, and in September these will be deployed to translate content for our user partners, NHS 24 and Cochrane.

- We have translation systems to translate from English into each of the four HimL target languages, i.e. Czech, German, Polish and Romanian.
- For the first two languages the systems are based on the University of Edinburgh's strong Moses-based submissions to the Workshop in Machine Translation 2014 (WMT2014) medical translation task, incorporating extra freely available data resources.
- For Polish and Romanian, we built systems tuned to the medical domain using large freely available corpora, and the University of Edinburgh's typically high performing WMT setup.
- In the coming months we will evaluate these deployed systems, and seek to improve them by analysing their performance and applying further state-of-the-art domain adaptation techniques.

3. The development of the research which will underpin system improvements

The academic partners in HimL have been published a total of 16 HimL-related papers in 2015 describing the research that will form part of the HimL translation systems. The publications have appeared in conferences such as Empirical Methods in Natural Language Processing, the Workshop in Statistical Machine Translation and Syntax, Semantics and Structure in Machine Translation. Topics addressed by these publications include:

- The handling of negation, prepositions, and German verbs in SMT.
- Morphological analysis and lemmatisation of text.
- Translation system building for the WMT shared translation task.
- Improving phrase tables with deep syntax.

A full list of HimL publications can be found at <http://www.himl.eu>.

Newsletter 2/2015

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 644402.

