HimL: Health in My Language

Barry HADDOW, Alex FRASER

Ludwig Maximilian University of Munich

`fraser@cis.lmu.de`

Abstract. HimL (www.himl.eu) is a three-year EU H2020 innovation action, which started in February 2015. Its aim is to increase the availability of public health information via automatic translation. Targeting languages of Central and Eastern Europe (Czech, German, Polish and Romanian) we aim to produce translations which are adapted to the health domain, semantically accurate and morphologically correct. The project is coordinated by Barry Haddow (University of Edinburgh) and includes two additional academic partners (Charles University and LMU Munich), one integration partner (Lingea) and two user partners (NHS 24 and Cochrane).

Description

In HimL we aim to deploy and evaluate machine translation systems for the public health domain, addressing domain adaptation, semantic accuracy and target morphology. The systems are used to tramslate content for NHS 24 (Scotland's national telehealth organisation) and Cochrane (an international NGO that produces systematic reviews of healthcare topics). The project has now been running for over a year, and we have already developed the first release of our translation systems and used them to translate the user partner websites. To build these systems, we have collected a large and diverse training set and are analysing the performance of existing domain adaptation techniques in combining these resources, as well as investigating the use of neural models in domain adaptation. We have been developing our corrective approaches to morphology handling, using machine learning to provide language independence, as well as extending the two-step approach to morphology to handle a wider range of phenomena, and new language pairs. For improved semantic accuracy we have experimented with using semantic roles to make sure important information is not lost, as well as developing methods to remove semantically incorrect translations from the model, and analysing the problems that arise in the translation of negation. The goal of semantic accuracy is supported by our development of new human and automatic semantic evaluation measures based on the UCCA (universal conceptual cognitive annotation) framework.

Acknowledgements