

## **Machine Learning Assisted Approach to Bilingual Dictionary Compilation**

*Steinþór Steingrímsson & Þórdís Úlfarsdóttir*

*The Árni Magnússon Institute for Icelandic Studies*

Compiling dictionaries is traditionally labour intensive and expensive. Automatic methods to aid in that process can thus be of great practical value. The growing availability of open, high-quality lexical resources and semantic data, monolingual as well as multilingual, together with more adequate machine translation systems, opens up a wide range of possibilities for new methods and approaches in this field.

For word translations in a new Icelandic-English dictionary, automatic bilingual lexicon induction is used to create candidate lists. We show how the dictionary building process can be expedited by manually filtering that list to identify good translational equivalents.

The new dictionary is based on the ISLEX dictionaries, a multilingual Icelandic-Nordic dictionaries containing over 50 thousand entries for Icelandic as a source language and Danish, Faroese, Finnish, Norwegian Bokmål, Nynorsk and Swedish as target languages.

We experiment with a variety of approaches, using pivoting through intermediary dictionaries, word alignment on parallel and comparable corpora, machine translation and a combination of these to suggest the most likely equivalency candidates for each word entry, and to filter out unlikely ones. Furthermore, we describe how we use a combination of different machine translation systems and models, by taking advantage of the available translations of examples and definitions in ISLEX into the Nordic languages, to aid in the process of translating definitions and examples for the dictionary.

Keywords: Dictionaries; NLP; Bilingual Lexicon Induction; Machine Translation; Icelandic