

The State of Icelandic LT



Eiríkur Rögnvaldsson
University of Iceland, Reykjavík

Introduction

At the turn of the century, Icelandic language technology (henceforth LT) was virtually non-existent. There was a relatively good spell checker, a not-so-good speech synthesizer, and that was all. There were no programs or even individual courses on language technology or computational linguistics at any Icelandic university, there was no ongoing research in these areas, and no Icelandic software companies were working on language technology.

In 1998, the Minister of Education, Science and Culture appointed a group of experts to investigate the situation in language technology in Iceland and come up with proposals for strengthening the status of Icelandic language technology. The group handed its report to the Minister in April 1999 and in 2000, the Government launched a special Language Technology Program, with the aim of supporting institutions and companies in creating basic resources for Icelandic language technology work. This initiative resulted in several projects which have had profound influence on the field.

Icelandic LT Work 2000-2010

The main direct products of the government-funded LT Program are the following:

- full-form morphological database of Modern Icelandic inflections
- balanced morphosyntactically tagged corpus of 25 million words
- training model for data-driven POS taggers
- text-to-speech system
- speech recognizer
- improved spell checker

Most of these products were developed in cooperation between research institutes and commercial companies. After the LT Program ended six years ago, LT researchers from three research institutes (University of Iceland, Reykjavík University, and the Árni Magnússon Institute for Icelandic Studies), who had been involved in most of the projects funded by the LT Program, decided to join forces in a consortium called the Icelandic Centre for Language Technology (ICLT), in order to fol-

low up on the tasks of the Program. The main roles of the ICLT are to

- serve as an information centre on Icelandic LT by running a website (<http://iclt.is>)
- encourage cooperation on LT projects between universities, institutions and commercial companies
- organize and coordinate university education in LT
- participate in Nordic, European and international cooperation within LT
- initiate and participate in R&D projects in LT
- keep track on resources and products in the field of Icelandic LT
- hold LT conferences with the participation of researchers, companies and the public
- support the growth of Icelandic LT in all possible manners

Over the past six years, the ICLT researchers have initiated several new projects which have

- treebank with a historical dimension

These resources were chosen because they were considered central to current LT work and prerequisites for further research and development in Icelandic LT.


The Prospects of Icelandic LT

Twelve years ago, the LT expert group estimated that it would cost around one billion Icelandic krónas (which then amounted to about ten million Euros) to make Icelandic language technology self-sustained. After that, the free market should be able to take over, since it would have access to public resources that would have been created by the Language Technology Program, and that would be made available on an equal basis to everyone who was going to use these resources in their commercial products.

However, the total budget of the government-funded LT program over its lifespan (2000-2004) was only 133 million Icelandic krónas – that is, 1/8 of the sum that the expert group estimated would be needed. It should therefore come as no surprise that we still have a long way to go. There are only 320,000 people speaking Icelandic, and that is not enough to sustain costly development of new products. At present, no commercial companies are working in the LT area because they don't see it as profitable. It is thus extremely important to continue public support for Icelandic language technology for some time, but given the current financial situation, it does not seem likely that such support will come from the state budget in the near future.

For a small language community and a small research environment like the Icelandic one, it is vital to

cooperate, not only on the national level but also internationally. Since 2000, Icelandic researchers and policy makers have taken an active part in Nordic cooperation on language technology. This has been of major importance in establishing the field in Iceland. The Nordic Language Technology Research Programme 2000-2004 was instrumental in this respect.

Iceland has just recently entered the CLARIN consortium, and together with the other Nordic and Baltic countries, Iceland also takes part in the META-NORD project which starts February 1st this year. We sincerely hope that our participation in these projects will help us to develop, standardize and make available several important LT resources and thus contribute to the growth of Icelandic language technology. 



The front page of the IceNLP, a NLP toolkit for Icelandic

been partly supported by the Icelandic Research Fund and the Icelandic Technical Development Fund. The most important products of these projects are:

- linguistic rule-based tagger, IceTagger
 - shallow parser, IceParser
 - mixed-method lemmatizer, Lemmald
 - (prototype of a) context-sensitive spell checker
- In 2009, the ICLT received a relatively large three year Grant of Excellence from the Icelandic Research Fund for the project "Viable Language Technology beyond English – Icelandic as a test case". Within that project, three types of LT resources are being developed:
- database of semantic relations (a pilot WordNet)
 - prototype of a shallow-transfer machine translation system