

National Language Technology Programme for Icelandic

Anna Björk Nikulásdóttir¹, Jóhanna Vigdís Guðmundsdóttir² and Jón Guðnason³

¹Consortium for Icelandic LT (SÍM) - anna@grammatek.com

²Almannarómur - jvg@almannaromur.is

³Reykjavik University - jg@ru.is



Language Technology Programme

The Icelandic Government has set a goal to ensure that Icelandic can be used for communication and interaction in the digital world.

On the 1st of October 2019, work on a five-year national project plan for Icelandic language technology (LT) started.

The aim of the programme is to deliver open and usable software and resources for further R&D, including commercial LT products.

From the beginning, the programme emphasises cooperation between academia and industries.

Core Projects – Language Resources



Text corpora

Parallel text corpora

Lexicographic resources



Speech corpora – crowdsourced

Speech corpora – high quality recordings

Core Projects – NLP tools

Tokeniser

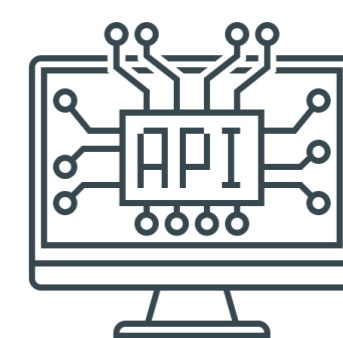
Part-of-speech taggers

Parsers

Lemmatiser

Named Entity Recogniser

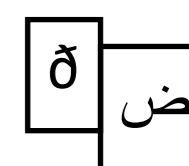
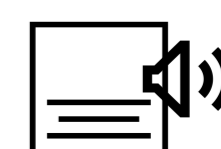
Semantic Analysis



Core Projects – Infrastructure Software

Speech recognition – Speech synthesis – Machine Translation – Spell and grammar checking

For all infrastructure projects, the aim is to publish open environments for adaption and further research and development, as well as to deliver software packages that can be integrated into other products. The delivered software will be designed for defined domains, with instructions for adaption to other domains.



Standards and Deliverables

Aim is to publish the majority of the language resources under the CC BY 4.0 license.

All delivered software will be published under the Apache 2.0 license, or a comparable license.

Data and software development guidelines and quality standards are coordinated accross all projects.

All deliverables will be stored and distributed by CLARIN.



Project Organization

The self-owned foundation *Almannarómur* conducts the programme on behalf of the Icelandic Ministry of Science, Education and Culture.

The research and development work is carried out by teams built across a consortium of universities, institutions, and private companies, *SÍM*.

Members of *SÍM*: Reykjavik University, University of Iceland, The Árni Magnússon Institute for Icelandic Studies, The Icelandic National Broadcasting Service (RÚV), Creditinfo, The Association of the Visually Impaired, Grammatek, Miðeind, and Tiro.

