

The Icelandic Centre for Language Technology
(ICLT)
Annual Report
September 2008–August 2009

Sigrún Helgadóttir, Árni Magnússon Institute for Icelandic Studies
Hrafn Loftsson, Reykjavik University
Eiríkur Rögnvaldsson, University of Iceland

1 Introduction

The Icelandic Centre for Language Technology (ICLT) was established on June 15th, 2005. The Centre has its roots in the LangTec Project of the Ministry of Education, Science, and Culture, which lasted from 2000 to 2004. ICLT is run jointly by the *Institute of Linguistics* at the University of Iceland (UI), the *School of Computer Science* at Reykjavik University (RU) and the *Department of Lexicography* at The Árni Magnússon Institute for Icelandic Studies (AMI).

This annual report is meant to give the Language Technology (LT) community in Iceland and elsewhere, and our sponsors and funding agencies, an overview of the activities of ICLT for the operating year September 2008–August 2009.

ICLT serves as a platform for cooperation between the participating institutions. It serves its role by:

- serving as an information centre on Icelandic LT by running a website <http://www.tungutaekni.is>
- encouraging cooperation on LT projects between universities, institutions and private companies
- organising and coordinating university education in LT
- taking part in Nordic, European and international cooperation in the field of LT
- initiating and taking part in research projects in LT
- initiating and taking part in practical projects in LT
- keeping track of resources and products in the field of LT

- holding a bi-annual LT conference with the participation of researchers, companies and the public
- supporting the growth of Icelandic LT in all possible manners

2 Board members

Each member institution appoints one member of the board. The board members for the last operating year were:

- Professor Eiríkur Rögnvaldsson, appointed by the *Institute of Linguistics* at the UI (chairman)
- Dr. Hrafn Loftsson, appointed by the *School of Computer Science* at RU
- Sigrún Helgadóttir, MSc., appointed by the *Department of Lexicography* at the AMI

3 Current researchers and research areas

During the operating year, the following researchers from the member institutions have taken direct part in the activities of ICLT: Anna B. Nikulásdóttir, Eiríkur Rögnvaldsson and Matthew J. Whelpton (UI); Hannes H. Vilhjálmsson and Hrafn Loftsson (RU), Kristín Bjarnadóttir and Sigrún Helgadóttir (AMI). Furthermore, Anton K. Ingason, an undergraduate student in Icelandic language at the UI, worked closely with researchers in ICLT on many of the activities.

During the last few years, these individuals have carried out research in the following areas of LT: context-sensitive spell checking, corpus construction and annotation, lemmatisation, morphology, part-of-speech tagging, shallow parsing, speech recognition, and interactive virtual environments. As a result of these research projects, a number of BLARK (Basic Language Resource Kit) units have been created for the Icelandic language. Part of the BLARK has been made open source and is currently available at <http://sourceforge.net/projects/icenlp/>.

4 Funding

ICTL does not receive any direct funding for its activities. On the other hand, the participating institutions indirectly support the activities of ICLT in the form of facilities for seminars and students, overhead cost in research projects, etc.

During the operating year, the members of ICLT received the following grants:

- Grant of Excellence: *Viable Language Technology Beyond English – Icelandic as a Test Case*. Project manager: Eiríkur Rögnvaldsson (UI).

Co-applicants: Hrafn Loftsson (RU), Matthew J. Whelpton (UI), Kristín Bjarnadóttir (AMI), Anthony Kroch and Joel Wallenberg (University of Pennsylvania), Mikel L. Forcada (Universitat d’Alacant). Duration: 3 years. Grant: 43,5 million ISK. Sponsor: The Icelandic Research Fund.

- Project grant: *Syntactic parsing of Old Icelandic texts*. Project manager: Eiríkur Rögnvaldsson (UI). Duration: 8 months. Grant: 0,9 million ISK. Sponsor: University of Iceland Research Fund.
- Project Grant: *Tagging and correcting a new corpus*. Project manager: Hrafn Loftsson (RU). Co-applicants: Eiríkur Rögnvaldsson (UI), Sigrún Helgadóttir (AMI). Duration: 3 months. Grant: 0,3 million ISK. Sponsor: The Icelandic Student Innovation Fund

In addition, the following grants awarded to ICLT researchers before September 2008 were still running during the operating year:

- Project grant: *A tagged Icelandic corpus*. Project manager: Sigrún Helgadóttir (AMI). Duration: 4 years. Grant: 18.5 million ISK. Sponsor: The Ministry of Education, Science and Culture.
- Project grant: *Context-sensitive spell checking*. Project manager: Eiríkur Rögnvaldsson (UI). Collaborators: Hrafn Loftsson (RU) and Sigrún Helgadóttir (AMI). Duration: 2 years. Grant: 3.1 million ISK. Sponsor: The Icelandic Research Fund.
- Project grant: *Improved tagging accuracy of Icelandic Text*. Project manager: Hrafn Loftsson (RU). Collaborators: Eiríkur Rögnvaldsson (UI) and Sigrún Helgadóttir (AMI). Duration: 2 years. Grant: 3.1 million ISK. Sponsor: The Icelandic Research Fund.

5 Activities

5.1 Service

Eiríkur Rögnvaldsson is active in the international research community in the field of LT. The following lists international service activities contributed by him during the operating year:

- Vice-President, NEALT (North-European Association for Language Technology)
- Member, Language Technology Working Group (Nordic Language Council)
- Board member, NGS LT (Nordic Graduate School of Language Technology)

The following lists participation of the members of ICLT in program committees or reviewing for conferences/journals/grant agencies:

- Eiríkur Rögnvaldsson: Reviewer for the *NoDaLiDa conference*, Odense, Denmark. Spring 2009.
- Hrafn Loftsson: Reviewer for the *National Science Foundation*, USA. Spring 2009.

5.2 Membership in international organisations

ICLT is a member of CLARIN (Common Language Resource and Technology Infrastructure; <http://www.clarin.eu/>) and ELRA (European Language Resources Association; <http://www.elra.info/>), and a supporting member of NEALT (Northern European Association for Language Technology; <http://omilia.uio.no/nealt/>).

Both RU and UI are members of NGSLT.

5.3 International collaboration

During the operating year, ICLT cooperated in the following international projects:

- “Viable Language Technology Beyond English – Icelandic as a Test Case”. The main objective is to develop scientific LT methods that are suited for less-resourced languages (<http://iceblark.wordpress.com/>). International collaborators are from the University of Pennsylvania, USA, and Universitat d’Alacant, Spain.

5.4 ICLT seminar series

During the operating year, ICLT continued its LT seminar series. Table 1 shows the talks given in the series.

Date	Venue	Title	Lecturer
24/11/2008	UI	The Linguistic Utility of Tagging Icelandic	Joel Wallenberg University of Pennsylvania
02/12/2008	RU	CombiTagger: A system for developing combined taggers	Verena Henrich and Timo Reuter University of Applied Sciences, Darmstadt
20/01/2009	RU	Correcting a PoS-tagged corpus using three complementary methods	Hrafn Loftsson, Reykjavik University

Table 1: Talks given during the operating year

5.5 Conference on Social Sciences

On March 13th 2009, ICLT took part in “Hugvísindafing”, an annual conference on social sciences at UI. ICLT was responsible for the seminar “Máltækni í mó-tun” [Formation of Language Technology]. Members of ICLT gave the following talks:

- *Íslensk máltækni – fortíð og framtíð [Icelandic LT – past and future]*. Eiríkur Rögnvaldsson, UI.
- *Íslensk beygingarlýsing og hagnýting hennar [The Database of Modern Icelandic Inflections and its utilisation]*. Kristín Bjarnadóttir, AMI.
- *Mörkun texta og markaðar málheildir [Tagging text and tagged corpora]*. Sigrún Helgadóttir, AMI.
- *“Eyríki finnst málfræði skemmtileg” – vélræn leiðrétting samhengisháðra ritvillna [Automatic correction of context-sensitive spelling errors]*. Anton K. Ingason, UI.
- *Páttun og trjábankar [Parsing and treebanks]*. Hrafn Loftsson, RU, Eiríkur Rögnvaldsson and Anton K. Ingason, UI.
- *Vélrænar grófbýðingar [Shallow-transfer machine translations]*. Hrafn Loftsson, RU.
- *Merkingarbrunnur fyrir íslenska máltækni [A semantic network for Icelandic LT]*. Matthew J. Whelpton and Anna B. Nikulásdóttir, UI.

5.6 Masters program in LT

In Fall 2007, ICLT started its new Masters program in LT. The program, which is run jointly by RU and the UI, is a two year interdisciplinary program (120 ECTS credits).

Students must either have a BA degree in the humanities (languages and linguistics) or a BS degree in computer science or related subjects (such as electrical and computer engineering). Furthermore, students have the possibility of taking courses at foreign universities, especially those participating in the Nordic Graduate School of Language Technology (NGSLT).

Two students were admitted to the program in Fall 2008.

5.7 Courses

The following courses, which are part of the joint LT program, were taught by members of ICLT during the operating year:

- Natural Language Processing. Fall 2008. Teachers: Hrafn Loftsson and Hannes Högni Vilhjálmsson (RU).
- Tölvur og tungumál [Computers and languages]. Fall 2008. Teacher: Eiríkur Rögnvaldsson (UI)
- Íslensk hljóðfræði og hljóðkerfisfræði [Icelandic phonetics and phonology]. Spring 2009. Teacher: Eiríkur Rögnvaldsson.
- Merkingafræði orða og orðanet [Lexical semantics and wordnets]. Spring 2009. Teacher: Anna B. Nikulásdóttir.

5.8 Supervision of students

During the operating year, members of ICLT supervised the following BSc/MSc students working on their LT theses/projects:

- Ida Kramarczyk (RU), an MSc student in LT. Thesis: *Improved Tagging Accuracy of Icelandic text*. Supervisor: Hrafn Loftsson.
- Martha Dís Brandt (RU), an MSc student in LT. Thesis: *Developing a shallow-transfer translation system using existing open-source tools*. Supervisor: Hrafn Loftsson.
- Kristjana Guðjónsdóttir (RU), a BSc student in Computer Science. Final project: *Prototype of an Icelandic Grammar Checker*. Supervisor: Hrafn Loftsson.
- Hlynur Sigþórsson and Pétur Sigurðsson (RU), BSc students in Computer Science. Final project: *Using IceNLP in Computer-Assisted Language Learning*. Supervisor: Hrafn Loftsson.
- Aðalsteinn Tryggvason (RU), a BSc student in Computer Science. Final project: *Named Entity Recognition for Icelandic*. Supervisor: Hrafn Loftsson.

6 Publications

In the period covered by this report, researchers in ICLT have published the following peer reviewed papers in the field of LT:

- Verena Henrich, Timo Reuter and Hrafn Loftsson. 2009. CombiTagger: A System for Developing Combined Taggers. In *Proceedings of the 22nd International FLAIRS Conference, Special Track: “Applied Natural Language Processing”*. Sanibel Island. Florida, USA.
- Hrafn Loftsson, Ida Kramarczyk, Sigrún Helgadóttir and Eiríkur Rögnvaldsson. 2009. Improving the PoS tagging accuracy of Icelandic text. In *Proceedings of the 17th Nordic Conference of Computational Linguistics (NODALIDA-2009)*. Odense, Denmark.
- Anton K. Ingason, Skúli B. Jóhannsson, Sigrún Helgadóttir, Hrafn Loftsson and Eiríkur Rögnvaldsson. 2009. Context-Sensitive Spelling Correction and Rich Morphology. In *Proceedings of the 17th Nordic Conference of Computational Linguistics (NODALIDA-2009)*. Odense, Denmark.
- Eiríkur Rögnvaldsson, Hrafn Loftsson, Kristín Bjarnadóttir, Sigrún Helgadóttir, Anna B. Nikulásdóttir, Matthew Whelpton and Anton K. Ingason. 2009. Icelandic Language Resources and Technology: Status and Prospects. In *Proceedings of the NoDaLiDa 2009 Workshop “Nordic Perspectives on the CLARIN Infrastructure of Common Language Resources”*. Odense, Denmark.

- Hrafn Loftsson. 2009. Correcting a POS-Tagged Corpus Using Three Complementary Methods. In *Proceedings of the 12th Conference of the European Chapter of the ACL (EACL 2009)*. Athens, Greece.
- Anna B. Nikulásdóttir and Matthew Whelpton. 2009. Automatic Extraction of Semantic Relations – For Less-Resourced Languages. In *Proceedings of the NoDaLiDa 2009 Workshop “WordNets and other Lexical Semantic Resources - between Lexical Semantics, Lexicography, Terminology and Formal Ontologies”*. Odense, Denmark.

7 Forthcoming activities

During the next twelve months, we plan to continue our work with the aim of achieving the objectives stated in section 1. Our LT Masters program will start for the third time in Fall 2009. The program will hopefully result in more LT research projects being carried out in the field of LT in Iceland. Furthermore, ICLT will try to continue attracting foreign speakers to its seminar series.

Finally, in August 2010 ICLT will host IceTAL 2010 (<http://icetal.ru.is>), the 7th International Conference on Natural Language Processing. The preparation for this conference will take up significant time of the activities of ICLT in the next 12 months.

8 Summary

During the past year ICLT has expanded its activities and achieved a level of recognition in the local LT community. By inviting foreign speakers to its seminar series, collaborating internationally in organisations and in research projects, and last but not least, publishing articles in international conferences, ICLT is gradually becoming known in the global LT community as well.

Further information is available from the ICLT web page at <http://www.iclt.is>.