

The Icelandic Centre for Language Technology
(ICLT)
Annual Report
May 2006–August 2007

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

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, the *School of Computer Science* at Reykjavik University and the *Department of Lexicography* at the Árni Magnússon Institute for Icelandic Studies.

This first 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 May 2006–August 2007¹.

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

¹This first annual report covers an extended period – subsequent reports will cover 12-month periods.

- 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:

- Eiríkur Rögnvaldsson, appointed by the *Institute of Linguistics* at the University of Iceland (chairman)
- Hrafn Loftsson, appointed by the *School of Computer Science* at Reykjavik University
- Sigrún Helgadóttir, appointed by the *Department of Lexicography* at the Árni Magnússon Institute for Icelandic Studies

3 Current researchers and research areas

During the operating year, four researchers from the member institutions have taken part in the activities of ICLT: Eiríkur Rögnvaldsson (University of Iceland), Hrafn Loftsson (Reykjavik University), Kristín Bjarnadóttir (Árni Magnússon Institute for Icelandic Studies) and Sigrún Helgadóttir (Árni Magnússon Institute for Icelandic Studies).

During the last few years, these individuals have carried out research in the following areas of LT: BLARK (Basic Language Research Kit) development, context-sensitive spell checking, corpus annotation, morphology, part-of-speech tagging, shallow parsing, and speech recognition.

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:

- Project grant: *Tværsproglig søgning på tekster og ordbøger*. Project manager: Bente Maegaard (Copenhagen University) and Jürgen Wedekind (Copenhagen University). Collaborators: Gjert Kristoffersen (University of Bergen), Viggo Kann (KTH), Kimmo Koskenniemi (University of

Helsinki) and Eiríkur Rögnvaldsson. Grant: 0.3 million DKK. Sponsor: Nordplus Sprog.

- Project grant: *Context-sensitive spell checking*. Project manager: Eiríkur Rögnvaldsson. Collaborators: Hrafn Loftsson and Sigrún Helgadóttir. Duration: 2 years. Grant: 3.1 million ISK. Sponsor: The Icelandic Research Fund.
- Project grant: *Context-sensitive spell checking*. Project manager: Eiríkur Rögnvaldsson. Collaborators: Hrafn Loftsson and Sigrún Helgadóttir. Grant: 0.5 million ISK. Sponsor: University of Iceland (Research Liaison Office); first prize in a competition regarding exploitation of research/knowledge.
- Project grant: *The syntactic use of Old Icelandic POS tagged texts*. Project manager: Eiríkur Rögnvaldsson. Grant: 1.2 million ISK. Sponsor: University of Iceland Research Fund.
- Project grant: *Improved tagging accuracy of Icelandic Text*. Project manager: Hrafn Loftsson. Collaborators: Eiríkur Rögnvaldsson and Sigrún Helgadóttir. Duration: 2 years. Grant: 3.1 million ISK. Sponsor: The Icelandic Research Fund.
- Project grant: *Utilisation of a large lexicon for tagging Icelandic text*. Project manager: Hrafn Loftsson. Collaborators: Kristín Bjarnadóttir. Duration: 3 months. Grant: 0.33 million ISK. Sponsor: The Icelandic Student Innovation Fund.

In addition, the following grant awarded to ICLT researchers before May 2006 is still running:

- Project grant: *A tagged Icelandic corpus*. Project manager: Sigrún Helgadóttir. Duration: 4 years. Grant: 18.5 million ISK. Sponsor: The Ministry of Education, Science and Culture.

The following research project, which received a grant during the previous operating year (in January 2006), concluded in January 2007:

- Project grant: *Shallow parsing of Icelandic text*. Project manager: Eiríkur Rögnvaldsson. Collaborators: Hrafn Loftsson and Sigrún Helgadóttir. Duration: 1 year. Grant: 3.8 million ISK. Sponsor: The Icelandic Research Fund.

We remark that these grants, however, can only be used to support project specific activities, and not for activities related to the Centre as such.

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, NGSLT (Nordic Graduate School of Language Technology)

5.2 Membership in international organisations

ICLT is a member of CLARIN (Common Language Resource and Technology Infrastructure; <http://www.mpi.nl/clarin/>) 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/>).

During the operating year, Reykjavik University became a member of NGSLT. University of Iceland has been a member since the start of NGSLT in 2004.

5.3 International collaboration

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

- “Nordic Master Program in Language Technology” (NMPLT). An application to the Nordic Council of Ministers. Eiríkur Rögnvaldsson was the main applicant and coapplicants were from Denmark, Finland, Norway, Sweden and Russia.
- “Nordic Language Resource Infrastructure – preparatory phase”. An application to NordForsk. Main applicant was Kimmo Koskenniemi (University of Helsinki) and coapplicants were from Denmark, Norway, Sweden and Iceland.
- “Change and variation in Icelandic syntax”. A workshop with the aim of devising one or more collaborative research projects on the topic for which research funding could be sought in the US, Iceland and perhaps Europe. Main organisers were Anthony Kroch (University of Pennsylvania) and Sigríður Sigurjónsdóttir (University of Iceland). Participants were from USA, Iceland, UK, and the Faroe Islands.

5.4 ICLT seminar series

During the operating year, ICLT initiated a seminar series in LT. Each talk has been attended by 20-65 people. Table 1 shows the talks given in the series.

Date	Venue	Title	Lecturer
05/12/2006	University of Iceland	Machine Translation	Stefán Briem
03/01/2007	Reykjavik University	An Icelandic speech recognition system	Arnar Þór Jensson Tokyo Institute of Technology
09/01/2007	University of Iceland	Speaker recognition	Jón Guðnason Imperial College
06/02/2007	Reykjavik University	An experiment with combination and integration of taggers	Hrafn Loftsson Reykjavik University
06/03/2007	University of Iceland	Automatic classification of semantic relations	Anna Björk Nikulásdóttir University of Heidelberg
17/04/2007	Reykjavik University	Achieving Dynamic Turntaking in Multimodal Humanoids	Kristinn R. Þórisson Reykjavik University

Table 1: Talks given during the operating year

5.5 LT conference

In May 2006, ICLT organised its first LT conference. The conference was attended by academicians, people from industry, and others interested in LT. Talks were given on various topics, i.e. morphology, tagging, parsing, speech recognition, speech synthesis, internet search using Icelandic, and LT in lexicon construction.

5.6 Masters program in LT

During the operating year, ICLT structured a new Masters program in LT, commencing in Fall 2007. The program, which is run jointly by Reykjavik University and the University of Iceland, 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).

5.7 Supervision of students

During the operating year, the members of ICLT supervised the following BSc/MSc students:

- Frank Arthur Blöndahl Cassata (Reykjavik University), a BSc student working on his final project: *Automatic Thesaurus Extraction for Icelandic*. Supervisor: Hrafn Loftsson.

- Anna M. Sigurðardóttir (Reykjavik University), a BSc student working on the project *Utilisation of a large lexicon for tagging Icelandic text*, funded by the *The Icelandic Student Innovation Fund*. 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:

- Helgadóttir, S. (2007). Mörkun íslensks texta [Tagging Icelandic text]. *Orð og tunga* **9**:75–107.
- Loftsson, H. (2006). Tagging a morphologically complex language using heuristics. In *T. Salakoski, F. Ginter, S. Pyysalo and T. Pahikkala (eds.), Advances in Natural Language Processing, 5th International Conference on NLP, FinTAL 2006, Proceedings*. Turku.
- Loftsson, H. (2006). Tagging Icelandic text: An experiment with integrations and combinations of taggers. *Language Resources and Evaluation*, **40(2)**:175–181.
- Loftsson, H. (2007). Tagging Icelandic Text using a Linguistic and a Statistical Tagger. In *Proceedings of Human Language Technologies 2007: The Conference of the North American Chapter of the ACL*. Rochester, NY.
- Loftsson, H. and Rögnvaldsson, E. (2007). IceParser: An Incremental Finite-State Parser for Icelandic. In *J. Nivre, H.-J. Kaalep, K. Muischnek and M. Koit (eds.), Proceedings of the 16th Nordic Conference of Computational Linguistics NODALIDA-2007*. Tartu
- Loftsson, H. and Rögnvaldsson, E. (2007). IceNLP: A Natural Language Processing Toolkit for Icelandic. In *Proceedings of InterSpeech 2007, Special session: Speech and language technology for less-resourced languages*. Antwerp.
- Rögnvaldsson, E. (2006) The Corpus of Spoken Icelandic and Its Morphosyntactic Annotation. In *P. J. Henrichsen and P. R. Skadhauge (eds.): Treebanking for Discourse and Speech. Proceedings of the NODALIDA 2005 Special Session on Treebanks for Spoken Language and Discourse*. Copenhagen Studies in Language 32, 133–145. Samfundslitteratur, Copenhagen.
- Rögnvaldsson, E. (2007) Textasöfn og setningagerð: Greining og leit. [Corpora and Syntactic Structure: Analysis and Search.] *Orð og tunga* **9**:51–73.

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. As stated in section 5.6, a new Masters program will start in Fall 2007. The new program will, hopefully, result in more LT research project being carried out in the field of LT in Iceland.

ICLT will continue with its seminar series, and in May 2008 it will organise an LT conference with the participation of international speakers. Moreover, ICLT will host an NGSLT course in Spring 2008.

8 Summary

During the past year ICLT has expanded its activities and achieved a level of recognition in the local LT community. Furthermore, the list of publications shows that the Centre is on track to make itself recognised in the global LT research community as well.

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