

# Icelandic Documentation Center for Language Technology

by

Eiríkur Rögnvaldsson

## 1. Introduction

The Icelandic Documentation Center went through a complete reorganization in 2004. As a result of an agreement with the government-funded Language Technology Project, the Center's website has merged with the website [www.tungutaekni.is](http://www.tungutaekni.is), which the LangTec project has been running since its start in 2001. The new combined website retains the URL and the graphic design of the *tungutaekni* site, whereas the content is structured in a similar manner as in the other documentation centers. The Center will continue to be run by the Institute of Lexicography, as the LangTec project came to an end on December 31<sup>st</sup>, 2004.

## 2. The website

The reorganization of the website does not only involve changes in the URL and the graphic design. The information on the site has also been thoroughly updated and is now much more comprehensive and detailed than before. The number of different items listed in the categories *Players and Teams* and *Resources and Tools* has more than tripled since 2002. This of course reflects the growth in Icelandic language technology during the last couple of years, but it is also a witness of our success in harvesting relevant information.

Furthermore, the structure of the site has been changed in accordance with decisions made at the NorDokNet meeting in Bergen in October 2003 (cf. Kimmo Koskenniemi's article in the 2003 Yearbook). Each piece of information and each news item is now kept on a separate page, and each page has been indexed with one or more keywords from the keyword vocabulary of Language Technology World (LT-World, <http://www.lt-world.org>). The pages have also been furnished with meta tags especially designed for the SiteSeeker search engine (<http://www.euroling.se>). All the pages are available in both Icelandic

and English. The website now consists of some 310 html pages and around 20 pdf files; most of the latter are articles on Icelandic language technology.

### **3. Establishing Icelandic LT terminology**

The Documentation Center sees it as one of its missions to establish standard Icelandic terminology for language technology. We have put considerable work in translating the LT-World keyword list into Icelandic. A preliminary version of the translated list has been put on the web ([http://www.tungutaekni.is/info/LT\\_Keywords.html](http://www.tungutaekni.is/info/LT_Keywords.html)) and we hope to have a final version ready early in 2005. It must be pointed out that it is much more difficult and time-consuming to compile a list like this for Icelandic than for the other Nordic languages. There are two main reasons for this. One is that language technology is a very young field in Iceland, and many of its subfields are still relatively unknown in the country. Therefore, only a small language technology vocabulary has been established in Icelandic.

The other reason is the well-known reluctance of Icelanders to adopt foreign words and stems into their language. If we look at multilingual term lists like the one found (in slightly different versions) on the websites of many of the Nordic documentation centers (for instance at [http://www.tungutaekni.is/info/terms\\_en.html](http://www.tungutaekni.is/info/terms_en.html)), the significance of this restriction is quite obvious. In many cases, the same word (with variations in spelling, of course) is used in all the languages except Icelandic (see for instance *acoustics*, *articulation*, *automata*, *corpora*, *dialogue*, etc.). However, in none of the 41 terms in the list, the Icelandic term bears any resemblance at all to the other 10 languages. The Icelandic terms thus have to be made from scratch, so to speak.

In order to get the translations of the terms as correct and precise as possible, and to avoid proposing new terms where there are already established Icelandic terms in use, we will consult experts from various related fields and ask them to check our translations and correct them. We also plan to cooperate with the Terminology Committee of the Icelandic Society for Information Processing. The committee is currently preparing a new edition of the *Dictionary of Computer Terms (Tölvuorðasafn)*, and we want as much as possible of language technology terms to be included in this dictionary.

### **4. Other activities**

The coordinator of the Center participated in the NorDokNet meeting in Copenhagen in April, but noone from the Center was able to attend the meeting in Helsinki in October. The coordinator also took part in the Nordic language technology delegation that visited the Baltic countries in the end of October.

### **5. Future plans**

As a sort of continuation of the above-mentioned LangTec project, there are currently plans to form a consortium or network on language technology in Iceland. If everything goes as planned, the members of this consortium/network will be the University of Iceland, Reykjavík University, and several commercial companies in the telecommunications and software industry. One of the main objectives of the consortium/network is to promote language technology in

Iceland, and to disseminate knowledge of language technology projects, products, materials etc. Thus, it is clear that the Documentation Center will play a central role in this collaboration. We will of course continue to add material to our website and keep it up to date, but we will also rely on other means of communicating with people and disseminating knowledge, such as meetings with interest groups and companies, writing articles, and organizing seminars and conferences.

Eiríkur Rögnvaldsson  
[eirikur@hi.is](mailto:eirikur@hi.is)