

Guðrún Kvaran

Modern Icelandic lexicography

1. Introduction

Iceland is a small country of around 320,000 inhabitants. Nonetheless there is a vibrant book publishing industry and the general public buys a substantial number of books, both domestic and translated works. On account of how few people speak the language, extensive education in foreign languages is needed. The first foreign language taught in schools is English, the second language is Danish, owing to the fact that Iceland was part of the Danish state until 1944. Iceland is a member of the Nordic Council, which is a geo-political, inter-parliamentary forum for co-operation between the Nordic countries. The council places great emphasis on a mutual and internal understanding of Nordic, especially Scandinavian, languages, and Danish continues to be taught in Icelandic elementary and high schools. Other foreign languages taught in the upper levels of elementary schools and in high schools are primarily French, German and Spanish.

Language teaching and the use of foreign languages entails a need for dictionaries, with the country's language both as a target and source language. For a long time efforts were made to meet this need, but in the last decade little work has been done in this field and most of the available books are in serious need of extensive revision. Lexicography is time consuming and expensive work, and publishing companies show little interest. To the best of my knowledge no new or revised dictionary is in process in Iceland. However, several things are now happening towards improving the situation.

The University of Iceland had for sixty years an institute for lexicography, Orðabók Háskólans (The University Dictionary). This has now been merged with four other institutions in the field of Icelandic studies under the name the Árni Magnússon Institute for Icelandic Studies (www.arnastofnun.is). Originally the Institute for lexicography worked solely on a historical dictionary, work that was started in 1943. I will very briefly say something about this dictionary, but the main emphasis of this lecture will be on some novel developments, and in particular on a web-based dictionary called ISLEX. In addition, three projects of the department of lexicography of the Árni Magnússon Institute will be introduced, that are intended (in part) to strengthen and support dictionary making. Finally I will introduce an online website called "Snara". Dictionaries and some other reference works have been made publicly accessible there.

2. Historical dictionary of the Icelandic language

In the Nordic countries interest in historical dictionaries arose early. The reason why work was started on a historical dictionary in Iceland was simply that among the projects which the Arni Magnusson Institute in Copenhagen had embarked on and provided funds for in 1939, was the compiling of a historical dictionary of the ancient Icelandic language from the oldest known texts up to the middle of the 16th century. This date was based on the first printing of a book in Icelandic in 1540, a translation of the New Testament.

In 1943 the governing council of the University of Iceland agreed to begin a work on an Icelandic historical dictionary, ranging from 1540 and up to this present day, i.e., to continue where the Copenhagen dictionary project ended. By the millennium, most of the collection of slips, about two and a half million, with examples culled from printed books and manuscripts ranging over four centuries, had been made accessible on the Internet through the home page of the Institute and further under “Ritmálsafn Orðabókar Háskólans”. Work is underway to make a collection from the colloquial language, around 300,000 slips, accessible in the same way. Since the funding for the institute comes from the state treasury the access must be free and open to all.

With increased technology by the end of the last century, interest in making the collection of examples easier enjoyed support, and the institute has now accumulated a fairly good corpus of chosen texts for searching, with about 80 million running words. The corpus is open to the general public for searching and grows at a steady rate. It has been indexed by content for those who are looking for something in a special field, but the entire text in the corpus can also be searched. The access is through the home page of the Institute under the link “Gagnasöfn” (corpuses) and further under “Íslenskt textasafn” (Icelandic collection of texts). In addition the institute has, in collaboration with others, made a tagged Icelandic corpus, of which I will say more later on.

Due to lack of funding the historical project has hardly moved forward in the last decade, with more emphasis being placed on projects that were eligible for grants or more suitable for collaborations with others. Among other things, one concern was to strengthen the Institute in the field of dictionary making in general and making it better suited to take part in collaborative efforts. As a state funded institute it cannot, however, engage in direct competition with private publishers serving the general market.

3. The ISLEX-project

Although the work on the historical dictionary has been on hiatus, experience shows that the slip-collections, the corpus and several side projects are extremely useful for all kinds of work on Icelandic dictionaries. This is brought out, for example, by a collaboration that started in 2006 and reached an important milestone in 2011. This project, spearheaded by the department of lexicography, is an Icelandic-Scandinavian multilingual dictionary, usually called ISLEX. The headwords are Icelandic and the translations are into five languages: Danish, Swedish, Norwegian (both bokmål and nynorsk) and Faroese. The dictionary describes modern Icelandic. It is of a medium-size with about 50,000 lemmas. In addition it has a large number of collocations, phrases and idioms and many examples of use intended both for Icelandic and foreign users.

Icelandic is a language with a complicated inflectional system, and therefore it is very useful for foreign users to have a link to a database of inflections (nouns, adjectives, verbs) which I will come to later in this paper. ISLEX was opened to the general public on the “Day of the Icelandic language” on November 16th 2011, but the Faeroese part will only be ready in 2014, as the work on this part started later. Work on Finnish will commence at the beginning of the year 2013 with Nordic and domestic grants. How these two projects will fare is entirely dependent on grants, but it is increasingly difficult to procure funds for salaries, without which no translators can be hired.

ISLEX is now, at the end of 2012, the result of the collaboration of five Nordic countries. It will combine six Nordic languages and will therefore be an important contribution to the understanding between these nations – which are all closely related culturally and historically but speak different languages. The sixth language, Finnish, is a welcome addition, as Finnish is one of the Nordic languages even if it is of a different language family. The ISLEX-project has enhanced the relations between the participating universities and institutes and has received generous grants from Nordic funds. The dictionary is open to the public on the web, free of charge.

The ISLEX-partners are six Nordic institutes:

- Iceland: Stofnun Árna Magnússonar í íslenskum fræðum, Reykjavík.
- Denmark: Det Danske Sprog- og Litteraturselskab (DSL), Copenhagen.
- Norway: The University of Bergen.
- Sweden: Institutionen för svenska språket, Göteborg.
- Faroe Islands: Fróðskaparsetur Føroya, Tórshavn
- Finland: The University of Helsinki.

Each institute is a special unit team who takes care of the translations into their respective languages. They are also responsible for the financing of their part of the project. The project is mainly funded by the government of each participating country. Otherwise it would never have become reality, but substantial grants, mostly Nordic but also some domestic, have also been awarded for specific aspects of the project and collaborative effort, such as yearly consultation meetings.

ISLEX was designed from the beginning for the web and has never been intended as a printed book. It is operated and maintained in a central database, located in Reykjavík, but the editorial work was done simultaneously in the five countries as will be the fact with Faroese and Finish until the work has been completed. Many features of the web have been used, such as pictures, sounds, and hyperlinks, plus a range of search functions, etc. The database also functions as a communication tool.

From the start the recordings of the pronunciation of the headwords was on the list of desired features. Recently a grant was acquired for this work and recording of the lemmas and a great deal of the collocations will be finished at the end of 2012. The goal is also to record the phrases too.

The translations are in different stages of completion: the Swedish and Danish are nearly finished; and the Norwegian translations of the lemmas and a part of the collocations will be finished at the end of 2012. As the Faroese first joined the project in February 2011 they have still much work to do.

For those who are interested in looking something up, ISLEX can be accessed in four ways (a choice of different user profiles according to the metalanguage of the web page):

- Icelandic: www.islex.is,
- Danish: www.islex.dk,
- Norwegian: www.islex.no,
- Swedish: www.islex.se.

The dictionary is also available on mobile phones (iPhone, Nokia windows and Android).

There can be no doubt that the project would not have progressed as well as it did if the basic lexicographical data of the Institute in Reykjavík had not been available. The slip-collection on the Internet and the text corpus were extensively used for clarification and searching for examples and collocations, phrases and idioms.

The outcome of the ISLEX-project has been extensively used ever since it was launched, both in Iceland and in the Nordic countries. Originally the main emphasis was towards translators, but now the dictionary is in equal terms directed at children, teenagers and students in the Nordic countries with a view towards increasing their understanding of Nordic languages. Users can submit questions and comments, and the administrators of the project try to answer them and react to suggestions for improvements as far as possible. Currently means for continuing the project are being sought as a project like ISLEX will soon be outdated if nothing can be done to add new words, phrases and collocations and correct translations and misspellings.

After the formal presenting of ISLEX in 2011 interest has sprung up regarding collaborating on more dictionaries. In the summer 2012 the ambassadors of France and Germany to Iceland visited the department of lexicography of the Árni Magnússon institute to discuss the first ideas about collaborations on an Icelandic-German and Icelandic-French dictionary and talks are also underway about an Icelandic-Polish dictionary. A significant number of Poles have immigrated to Iceland in the last decades.

4. The Database of Modern Icelandic Inflection

As I mentioned before, Icelandic is a morphologically rich language with a complex inflectional system. It has up to 16 inflectional forms of a noun, 120 of an adjective, and 107 of a verb, not including variants. The endings that mark grammatical categories can, in some instances, have a number of variants, and stem changes are common, both in vowels and consonants. As an example I will show the word *köttur* ‘cat’ in all cases, plural and singular, with and without a suffixed article.

Case	Singular	With suffixed article	Plural	With suffixed article
Nom.	köttur	kötturinn	kettir	kettirnir
Acc.	kött	köttinn	ketti	kettina
Dat.	ketti	kettinum	köttum	köttunum
Gen.	kattar	kattarins	katta	kattanna

The database contains about 270,000 paradigms from Modern Icelandic, with over 5.8 million inflectional forms. It is accessible through the institute's home page and further under “Beyging orða” (the inflection of words). The database was created as a multipurpose resource, for use in language technology, lexicography, and as an online resource for the general public. It is e.g. much used by teachers and pupils as well as university students. This information is taken from an article by Kristín Bjarnadóttir, the editor of the database (2012). The article is available online (see the references section).

Traditional Icelandic dictionaries provide limited information on inflection. The tradition is to give only certain inflectional forms as indicators of inflectional class, such as the genitive singular and nominative plural for nouns, either by showing the endings or, in the case of vowel change, by showing the whole inflectional form. Information on the inflection in other word classes is also fragmentary. The description of verbs is usually confined to the principal parts, i.e. three or four inflectional forms, depending on inflectional class, four for the so called strong verbs and three for weak verbs. The inflection of adjectives is very often omitted altogether, although it is not easily predictable from the base forms. This information is enough for most Icelanders, but certainly not all, and they are insufficient for foreign users. It is for this reason that at the start of the ISLEX-project it was decided not to list traditional inflectional forms, but instead put in a link to the *Database of Modern Icelandic Inflection*. This has proven to be a very good and popular decision.

5. The Tagged Icelandic Corpus

Although the institute has collected a reasonably extensive corpus to facilitate searching for examples for dictionary making and every type of language studies, as mentioned above, it has the drawback that it is not tagged and each word form must be searched for separately. Therefore it was decided to establish a different kind of collection in collaboration with the University of Iceland and the Reykjavík University. This is called *The Tagged Icelandic Corpus* and consists of about 25 million tokens of contemporary Icelandic texts collected from varied sources during the years 2006-2010. The corpus is intended for use in Language Technology projects, for linguistic research and for lexicography. The corpus is available for search through a web interface and has already proved to be useful. The permission for the use of texts in the corpus has been secured from all copyright holders. To get all the permits required much work, but the effort was worth while. The researchers can now obtain the texts and use them for their projects (Helgadóttir et al. 2012).

6. The Icelandic Word Net

The third project I would like to mention is the *Icelandic Wordnet* (Íslenskt orðanet). It is best described with the words of Jón Hilmar Jónsson, research professor at the Department of Lexicography, who has developed it with the assistance of Þórdís Úlfarsdóttir.

It is an ongoing lexicographic project which aims to analyse and describe the Icelandic vocabulary and its internal semantic relations on the basis of syntactic and morphological relations in word combinations and compounds. Disambiguation and lemmatisation of multi-word lexical units plays a central role in the construction of the wordnet. The multi-word lemmas are grammatically tagged, which enables grammatical sorting and an interaction between semantic and grammatical features. The wordnet combines the functions of a synonym and a concept dictionary. (Jónsson 2012, 64)

The product of the analysis is open and accessible as an online dictionary.

7. Icelandic dictionaries on the Web

As I mentioned earlier the publishing companies have at present no dictionary projects ongoing, neither mono- nor bilingual dictionaries, and this has been the case for several years. One of the largest publishing companies in the country has taken the route of setting up a collection of previously published dictionaries in one place, in cooperation with computer specialists and other publishers. This project is called *Snara*, and subscriptions are at a low fee of roughly 3 EUR per month. It now contains seventeen dictionaries, for example an English-Icelandic and Icelandic-English, Danish-Icelandic and Icelandic-Danish, along with seven other handbooks. According to information from the administrators of the collection, all Icelandic universities – seven in all – and two foreign ones, are subscribers and students and staff can make use of the collection free of charge. The same goes for all high schools in Iceland and a great number of elementary schools and communities. The collection is thus used extensively in language learning and foreign communication. It helps to meet a serious need but many of the books are growing old and urgently in need of revision.

8. Conclusion

I consider it unlikely that major printed dictionaries will be published in Iceland in the future. I also do not consider it likely that a publisher would embark on a new dictionary project without cooperating with the Árni Magnússon Institute. The models I envisage are electronic dictionaries made in collaboration between the Institute and local and foreign partners. A great mass of data is already at hand at the institute, and additions are constantly being made that greatly facilitate the work on new editions of bilingual dictionaries.

9. References

- Jónsson, J.H. (2012): Að fanga orðaforðann [To catch the vocabulary]. In: *Orð og tunga* 14, 39-65.
- Bjarnadóttir, K. (2012): The Database of Modern Icelandic Inflection. In: *LREC 2012 Proceedings: Proceedings of "Language Technology for Normalization of Less-Resourced Languages", SaLTMil 8 – AfLaT 2012*. Internet: www.lrec-conf.org/proceedings/lrec2012/index.html: Workshops, SaLTMil-AfLaT.
- Helgadóttir, S./Svavarsdóttir, Á./Rögvaldsson, E./Bjarnadóttir, K./Loftsson, H. (2012): The Tagged Icelandic Corpus (MÍM). In: *LREC 2012 Proceedings: Proceedings of "Language Technology for Normalization of Less-Resourced Languages", SaLTMil 8 – AfLaT 2012*. Internet: www.lrec-conf.org/proceedings/lrec2012/index.html: Workshops, SaLTMil-AfLaT.

10. Online addresses

Icelandic Text Corpus: www.arnastofnun.is/page/gagnasofn_textasafn.

Icelandic Word Net: www.ordanet.is.

Kristín Bjarnadóttir: www.lexis.hi.is/kristinb/kristinb.html#rit.

Online dictionaries: www.snara.is.

Ritmálssafn Orðabókar Háskólans: <http://lexis.hi.is/cgi-bin/ritmal/leitord.cgi>.

The Árni Magnússon Institute of Icelandic Studies: www.arnastofnun.is.