

The ISLEX Web Dictionary and the Analysis of its Use

Thórdís Úlfarsdóttir, Halldóra Jónsdóttir

The Árni Magnússon Institute for Icelandic Studies, Reykjavík, Iceland

E-mail: disa@hi.is, halldo@hi.is

1. Introduction

ISLEX is an online multilingual dictionary between modern Icelandic (as a source language) and the Nordic languages Danish, Norwegian, Swedish, Faroese and Finnish, as target languages. These languages all belong to the North Germanic language group and are closely related, except Finnish, which is of the Finnic language family.

Icelandic is spoken by 320,000 people. The language is of some historical importance as it is still very close to Old Norse, the common ancestor of the Nordic languages (of the Germanic group). Both Modern Icelandic and Old Norse are widely studied and researched in universities in many countries today, which obviously requires good dictionaries. These have not always been widely available which was the reason why ISLEX came into being. In this paper, section 2 gives background information, and 3 and 4 briefly describe the methods used to analyse the dictionary use and users, firstly by log files, and secondly by a user survey.

2. General Project Background

The ISLEX-project began in 2005 and was from the start designed for the web. At that time, only one online dictionary between Icelandic and the other Scandinavian languages existed, a retro-digitized Danish-Icelandic dictionary. ISLEX was opened on the web in 2011.

ISLEX is organized as a joint project of academic institutes in six Nordic countries. The largest editorial staff, responsible for the description of the Icelandic SL and for the development and maintenance of the database, is located at *The Árni Magnússon Institute for Icelandic Studies* in Reykjavík, Iceland. Its partners and responsible for the Scandinavian TLs are the following:

- *The Society for Danish Language and Literature* in Copenhagen (for Danish)
- *The Dept. of Nordic Languages and Literature* at the University of Bergen, Norway (for the two Norwegian standards: Bokmål and Nynorsk)
- *The Dept. of Swedish Language* at the University of Gothenburg, Sweden (for Swedish)
- *The University of the Faroe Islands* (for Faroese)
- *Helsinki University* (for Finnish).

The project has mainly been funded by the governments of the participating countries, as well as by Nordic funds like NordPlus Sprog.

The dictionary contains 50,000 headwords. In the early stages the vocabulary was divided into about 500 semantic fields, and the processing of the articles for both the SL and the TLs relied on those fields rather than working through the alphabet. The dictionary places a special emphasis on phrases and collocations, and contains a

large number of them which have all been translated into the TLs. It also provides inflections (for the SL and two of the TLs), many pictures and examples of use, plus the usual features of labels, cross references, etc. The pronunciation of each Icelandic headword can be heard by clicking on a loudspeaker icon below the word.

Illustration 1 displays a simple article (*eplakaka* 'apple cake, apple pie') in the user interface of ISLEX. The user has already chosen the metalanguage of the page, which is Danish in this case. The metalanguage can be seen by how the word class is presented (*sb.fem.*, according to the Danish convention) and by the Danish metawords *udtale* ('pronunciation') and *bøjning* ('inflection').

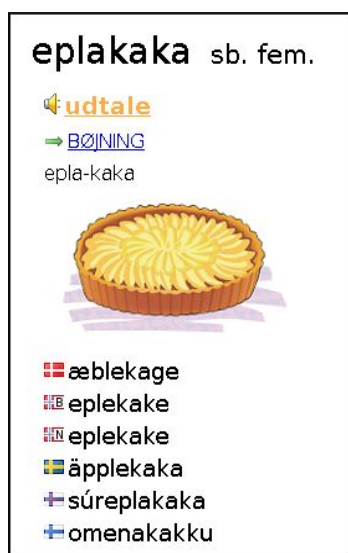


Illustration 1: The Danish user interface of ISLEX. The picture shows the headword *eplakaka* 'apple pie' with its word class ('feminine noun'). Below there is a loudspeaker icon which plays the pronunciation sound when clicked on. In the next line, BØJNING links to the inflectional paradigm of the headword. Next we see the two elements of the compound ('epla-kaka') and a picture of the cake below. Here, the user has chosen to see 'all languages', so they line up in a fixed order: Danish, two types of Norwegian (first Bokmål, then Nynorsk), Swedish and Faroese, followed by Finnish at the bottom.

The making of ISLEX has taken years, lots of funding and a great deal of effort. The editorial team was interested in finding out whether it was worth all the trouble, and in order to do that, the use of the dictionary was measured and analysed, as related in the next two sections.

3. Log files

From the start the ISLEX dictionary was well received by its users. Its use has been constantly recorded in the log files that are a part of the database. The log files can

give a variety of information, e.g. on the specific dictionary chosen, what words are looked up, etc. Here we will mention three important factors that the logs can tell us about: 1) the number of users, 2) the dictionary chosen by the user, and 3) the metalanguage selected by the user.

The number of users in one year (2015) is 100,000. The number has been slowly increasing since the launch of the dictionary in 2011, as more people have learnt about the dictionary's existence. With so few speakers of Icelandic (320,000), this number of users can probably be considered good achievement.

Dictionaries chosen. ISLEX currently has five TLs open to the public. Their use in October 2016 was as follows (the distribution is given in percentage):

| | |
|-----------------------------------|-----|
| Danish | 44% |
| Norwegian bokmål | 14% |
| Norwegian nynorsk | 11% |
| Swedish | 19% |
| Faroese | 10% |
| All dictionaries at the same time | 2% |

Table 1.

This is important information as a fundamental measure of a multilingual dictionary. As seen in table 1, it is no surprise that the Danish TL is the one most used, as Danish is an obligatory subject as a second language in Icelandic schools. Secondly, Faroese became available as a TL only in March 2015, and clearly it has already become established among the dictionary users. The use of the Norwegian and Swedish TLs are as expected, proportionally. As seen in the bottom row of table 1 there is an option of looking at all TLs at the same time (cf. illustration 1), and 2% of the users select to do so.

Metalanguage chosen. The user of ISLEX can choose between six different metalanguages (the language displayed on the web page; it also affects the form of grammatical notations of the headwords, etc.). The distribution of the metalanguages in October 2016 is as follows:

| | |
|-------------------|-----|
| Icelandic | 35% |
| Danish | 21% |
| Norwegian bokmål | 12% |
| Norwegian nynorsk | 8% |
| Swedish | 17% |
| Faroese | 7% |

Table 2.

The knowledge of the chosen metalanguage is extremely important as it may give us a clue to the users' native tongues. A little over a third choose Icelandic as a metalanguage, and one can conclude that in most cases they are native speakers of Icelandic. Danish, Norwegian and Swedish are fairly evenly spread, whereas the Faroese metalanguage may be somewhat higher than expected (7%), considering the few speakers of that language.

4. User survey

A user survey was conducted in March 2014 via the ISLEX web page. The general aim was to gain insight into the users' opinion of the dictionary and so to spot its strong and weak points. The survey was open for a period of two weeks. The answering of the questions was optional, so out of 6000 users that were active in that period, 416 users gave a response.

The questions in the survey were:

1. How useful is the dictionary?
2. How user friendly is the dictionary?
3. Where did you learn about ISLEX?
4. I find the Icelandic word I am looking for.
5. I find the Danish/Norwegian/Swedish word I am looking for.
6. Other comments, if any.

For each question, the user could choose between five options (by clicking on radio buttons) of the range (e.g. for question 1): *extremely useful, very useful, somewhat useful, not so useful, not at all useful*.

Question 6 did not have such options of course, as its objective was to invite the user to write a comment into a free text form. An example of the outcome of the survey is given in illustration 2.

Question 2. How user friendly is the dictionary?

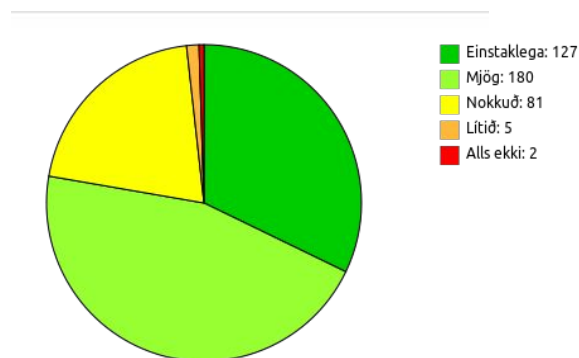


Illustration 2: "How user friendly is the dictionary?"

Answers:

Extremely user friendly 32%

Very user friendly 46%

Somewhat user friendly 21%
Not so user friendly 1%
Not at all user friendly 0%

The answers to the five questions and the written comments were important for the editors to be better able to evaluate the work. The room does not allow for more detailed analysis of the responses, but a later report will provide more details.

5. Conclusion

The user feedback resulted in some changes in the dictionary. Some wishes were realistic and easy to fulfill (like adding a picture), whereas in other cases the suggestions from the users would have been very complicated to bring about (such as redesigning the search mechanism). There are plans for a second survey in 2016. In order to make it as focused as possible, the authors are going to call for feedback on how to formulate the survey at a Nordic lexicological symposium held in Copenhagen in January 2016. What comes out of the upcoming survey remains to be seen, but listening to the voices of the users will continue to be a high priority of the ISLEX editorial team.

6. References

Paper in conference proceedings:

Rauset, Margunn, Hannesdóttir, Anna H. and Sigurðardóttir, Aldís. (2012). Ein-, to- eller fleirspråkleg ordbok? *Nordiska studier i lexikografi* 11, pp. 512-523.

Ulfarsdóttir, Thordis. (2014). ISLEX - A Multilingual Web Dictionary. *LREC 2014 Conference Proceedings, Reykjavík 26-31 May 2014*, pp. 2820-2825.

Journal article:

Hannesdóttir, Anna H. (2012). Ekvivalensrelationer i tvåspråkig lexikografi. *LexicoNordica* 19, pp. 39-58.

Websites:

ISLEX = *The ISLEX Dictionary*. Ulfarsdóttir, Thordis (Editor in Chief). Accessed at <http://www.islex.dk>, <http://www.islex.fo>, <http://www.islex.is>, <http://www.islex.no> and <http://www.islex.se>. (10 January 2016).