

Digitizing And Opening Access To Six Icelandic Dictionaries

Steinþór Steingrímsson

The paper describes a project in progress. Six Icelandic dictionaries are being digitized and will be made accessible on the web. A web platform is being built to allow searching and browsing the dictionaries. Some of the dictionaries will also be made available to download in a machine readable format.

The six dictionaries to be made accessible online are:

- 1) The *Icelandic etymological dictionary* compiled by Ásgeir Blöndal Magnússon in the late 20th century and originally published in 1989.
- 2) The *Icelandic-Danish dictionary* by Sigfús Blöndal, published in 1920, with supplement published in 1963.
- 3) *Dictionary by Gunnlaugur Oddsson*. A Danish-Icelandic dictionary first published in 1819. Second edition edited by Jón Hilmar Jónsson and published in 1991.
- 4) *Icelandic-Latin-Danish dictionary* by Björn Halldórsson. Compiled late in the 18th century and first published in 1814. Second edition edited by Jón Aðalsteinn Jónsson and published in 1992.
- 5) *Nucleus Latinitatis*. A Latin-Icelandic dictionary translated by Jón Árnason from the Icelandic-Danish dictionary by Hans Gram, published early in the 18th century.
- 6) *Lexicon Islandicum*. An Icelandic dictionary with definitions in Latin. Compiled by Guðmundur Andrússon between 1650 and 1654 and first published in 1683. Second edition was edited by Jakob Benediktsson and Gunnlaugur Ingólfsson and published in 1999.

The motivation for making these dictionaries accessible online is to make it easier for researchers, students, lexicographers and others around the world, working with the Icelandic language or interested in the history of the language to use these out-of-print resources.

At the beginning of the project the dictionaries were at different stages of digitization. All but the Icelandic Danish dictionary have been printed in the last 25 years and were all laid out using LaTeX. The LaTeX documents still exist and the markup is quite regular, so relatively little preprocessing is necessary before splitting the dictionary into articles, which are inserted into a database for search and lookup from a web page. The Icelandic-Danish dictionary on the other hand is only available in print and has never been digitized in any way. Thus we had all the 1100+212 (original+supplement) pages photographed. Then we use ABYYY FineReader 12 to OCR-read the pages. The resulting text is then proofread, the articles separated and markup added before it is published online. We estimate the text processing to be approximately three months work.

On the web platform, the user can search or browse one dictionary or all the dictionaries at the same time. Pages for each of the dictionaries will provide information and explanations for the corresponding work, a short introduction to the work, explanations for abbreviations and other information that helps the user use the dictionary in question.

Furthermore five of the dictionaries (2-6 in the list above) will be made available to download in a machine readable format.

We plan to make the dictionaries available one by one throughout 2016. The first one, The Etymological dictionary (no. 1 in the list above), will be available by the end of January. Dictionaries 3-6 in the list above are planned to be made available by the end of April 2016 and the Icelandic-Danish dictionary by the end of October 2016. The platform will be available on

ordabaekur.arnastofnun.is.