Málið.is: A Web Portal for Information on the Icelandic Language

Steinþór Steingrímsson, Ari Páll Kristinsson & Halldóra Jónsdóttir

Málið.is is a new web portal on the Icelandic language and language use. Initially it includes six different resources, providing reliable information on Icelandic. Work is under way to add more resources to the portal, providing more diverse information. *Málið.is* is expected to serve the Icelandic speaking population in general, although students and writing professionals could be expected to be the principal target groups. The aim of the project is to strengthen the Icelandic language in the digital era by making it easy to access information, helping people becoming more proficient and confident language users. Usage statistics shows growing user base, suggesting the project is a success.

1. Introduction

The web portal *málið.is* was opened in November 2016. It is created and run by The Árni Magnússon Institute for Icelandic Studies (AMI) and gives access to a variety of databases on the Icelandic language created and/or maintained by the Institute. The portal provides abundant reliable and authoritative information, guidance and advice on Icelandic, all in one place. This concerns e.g. orthography, grammatical issues such as inflections, the lexicon, the history and etymologies of particular lexical entities, semantics, synonymity, terminologies and translations of technical vocabulary, and many other matters of language and its usage. Previously, these resources were accessible via a variety of different formats, user interfaces, web addresses, search methods and functions, which caused problems for many users as they were typically not aware of all options available.

The name (and web address) of the web portal, *málið.is*, translates to 'the language.is', which is descriptive and easy to remember. Access to *málið.is* is free of charge, and the portal is completely clutter free, showing no ads, notifications or other disturbances, only the results from the resources connected to the portal.

The portal is in many ways similar to the Danish web portal *sproget.dk*. And indeed, the Danish portal, initiated and operated by the Danish Language Council and Society for Danish Language and Literature, served as a model and an inspiration as we were planning this Icelandic web portal.

Students and professional language users, including writers, journalists, translators and others who work professionally with language could be considered the principal target groups for *málið.is*, as the portal gives access to abundant terminological data and directions for how to use the language "correctly". But we also

expect it to serve the Icelandic speaking public in general and thus the web portal strives for plain and non-technical exposition and conciseness, whenever possible.
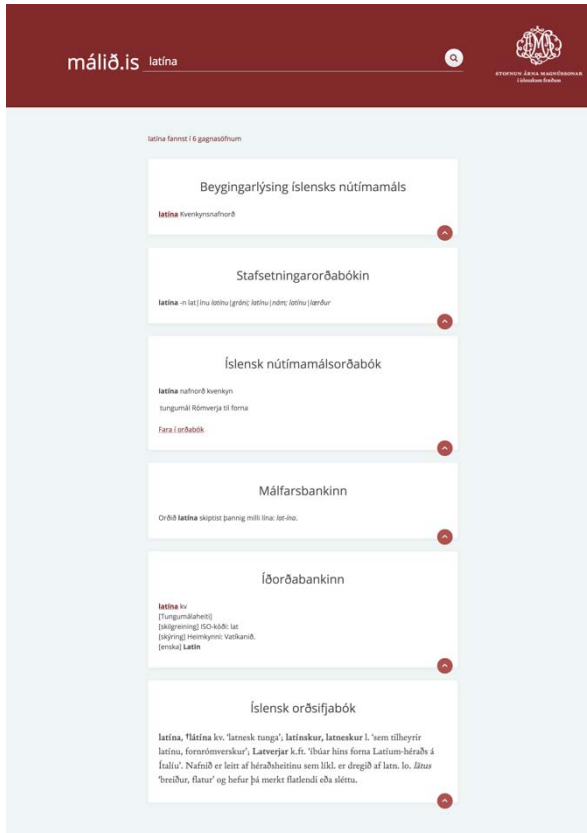

Figure 1: Example of search results on *málið.is*. The Icelandic word for ʻlatinʼ, *latína*, is found in all of the six available resources.

Some of the data available on *málið.is* are explicitly of a prescriptive nature, while other data have primarily descriptive functions. We do not expect users to be immediately familiar with this fundamental distinction, but by exposing users to multiple types of information we facilitate the comparison between them and enable users to better acknowledge the distinction. Indeed, one theoretical contribution of *málið.is* is that it highlights the difference between descriptive and prescriptive language resources, for the benefit of students and researchers.

2. Rationale for setting up the portal

The Árni Magnússon Institute for Icelandic Studies has the role to publish prescriptive dictionaries as well as descriptive ones, both contemporary and historical.

It publishes terminologies and word-lists on language for technical and specialized purposes, it offers language consultation and advice to the general public and moreover the institute compiles text corpora and other language resources. Thus, the institute has over the years been compiling various linguistic resources and diverse language materials. For the most part, these resources have been digitized by now. Six of these resources form the contents of *málið.is* in its initial phase.

Furthermore, the institute has the duty to provide information on the Icelandic language and assistance on language use. This takes a substantial time of some of the institute's employees. With the advent of *málið.is* we hope to get fewer trivial questions, as our clients can easily get their answers on our web portal. But although this could be reason enough to set up the portal, rather than to try easing our workload, the principal aim of *málið.is* is to make it easier for the writing and speaking public to access the information they need to be more proficient and confident language users.

In Iceland, there is widespread interest among the general public in matters of the Icelandic language, in particular to etymological speculations, the coining of purist neologisms and matters of standard vs. non-standard grammar (Hilmarsson-Dunn and Kristinsson, 2013). A great portion of the queries the institute receives show this, so we tried to fulfil these needs as well as possible from the start.

There is an urgent need to make as much language advice and information available online as possible. In recent years there have been growing demands in Iceland for dictionaries and other language resources to be made available free of charge to the general public. This project aims to satisfy these demands.

3. The resources

Initially six resources were made accessible through the portal. They were chosen based on two main criteria: (1) How appropriate they were to the portal — we wanted to start with information most relevant to everyday use of Icelandic. (2) The state of the resources' database — whether we could connect it to the portal with relatively small effort. Based on this, the following six resources were adapted in time for the portal's launch:

1) The Database of Modern Icelandic Inflection (Kristín Bjarnadóttir 2012). The DMII contains 278,000 paradigms from Modern Icelandic, with over six million inflectional forms. It was created as a multipurpose resource, for use in language technology, lexicography, and as an online resource for the general public.

2) Spelling Dictionary (Jóhannes B. Sigtryggsson (Ed.) 2016).
The 2nd edition of the Icelandic spelling dictionary, revised in accordance to updates to official Icelandic orthography, published by the Ministry for Education and Culture in 2016 (Íslensk málnefnd 2016).

3) Dictionary of Modern Icelandic (*Íslensk nútímamálsorðabók* 2017).
    Dictionary of Modern Icelandic is a new dictionary only available on-line. It is compiled at the department of lexicography at The Árni Magnússon Institute for Icelandic studies. It contains approximately 50,000 words. The work on the dictionary commenced in 2013 and is ongoing.

4) Language Usage Database (*Málfarsbankinn* 2017).
    A compilation of short articles giving advice on language use. It contains more than 7000 articles on subjects from grammar and syntax to well-crafted and elaborate language.

5) The Icelandic Term Bank (http://www.ordabanki.hi.is).
    The Icelandic Term Bank (ITB) contains around 60 bilingual or multi-lingual glossaries, and one monolingual glossary, containing terminologies in various fields. The glossaries contain terms in Icelandic, usually with corresponding terms in other languages, most commonly English. Often the terms are accompanied by definitions or explanations. In total the ITB contains more than 180,000 terms and it is constantly growing.

6) Icelandic Etymological Dictionary (Ásgeir Blöndal Magnússon 1989).
    Icelandic Etymological Dictionary (*Íslensk orðsifjabók*) was compiled by Ásgeir Blöndal Magnússon and originally published in 1989. It is the first and only Icelandic etymological dictionary and contains approximately 25,000 entries.

In fall 2017 *Íslenskt orðanet* (Icelandic wordnet) will be added. This is a resource that was considered for the initial launch, but was deferred due to technical complications. *Íslenskt orðanet* is a semantic database with elements that allow it to be used much like an onomasiological dictionary or a thesaurus (Jón Hilmar Jónsson 2018).

Other resources to be added are the Written Language Archive (*Ritmálssafn*), a database of language use with usage examples from 1540, when the New Testament was first published in Icelandic, to recent decades; also: four historical dictionaries, compiled in the 17th, 18th and 19th centuries; and possibly ISLEX, an Icelandic-Scandinavian dictionary with six language pairs. Adding ISLEX to *málið.is* could make the portal more useful to Icelandic students of Scandinavian languages and to Scandinavian users, who might benefit from some of the other resources too.

4. Usage

Usage of the portal has grown somewhat from its initial launch. The AMI does not have any advertising budget for the project and therefore has to make do with word-of-mouth. The project received valuable media coverage upon launch and in

the first six months we received on average 700 different users on working days. Most weeks we had around 2500 different users. The usage went down to 1500 weekly users during the summer 2017, but in August it picked up and the last two weeks of August and first week of September 2017, the last weeks in our data, show record numbers, 3200-3500 weekly users. This suggests that the word is spreading and the user base is growing.
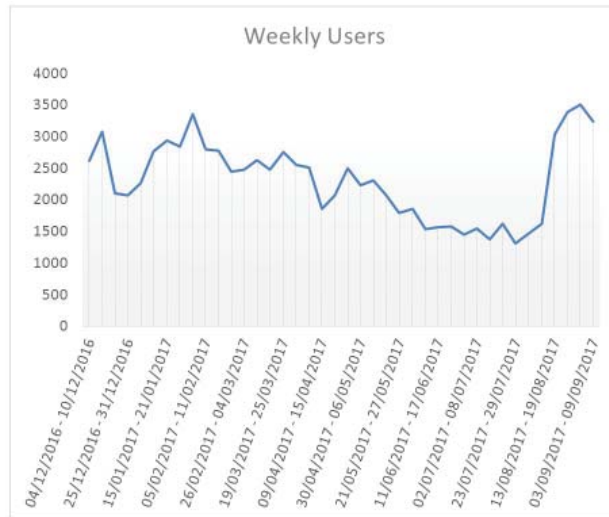

Figure 2: Weekly users.

In total, more than 50,000 users visited the site between November 2016 and September 2017. This is according to Google Analytics, which counts each browser in each device as a different user.[1] The actual number of users is therefore somewhat smaller. For reference, there are no more than 350,000 speakers of Icelandic in the world. Most of the portal's users (92%) are in Iceland.

The users look for all sorts of things but there seem to be some emerging patterns. Upon checking the most common words searched for on individual days, we can see that uncommon words used in news stories get many hits for a short period of time. People also search for words that are commonly misspelled, probably to make sure they spell them correctly. The distribution of search strings has a very long tail, as expected. Fewer than 300 words get more than 50 hits and 164,000 search strings get less than 10 hits.

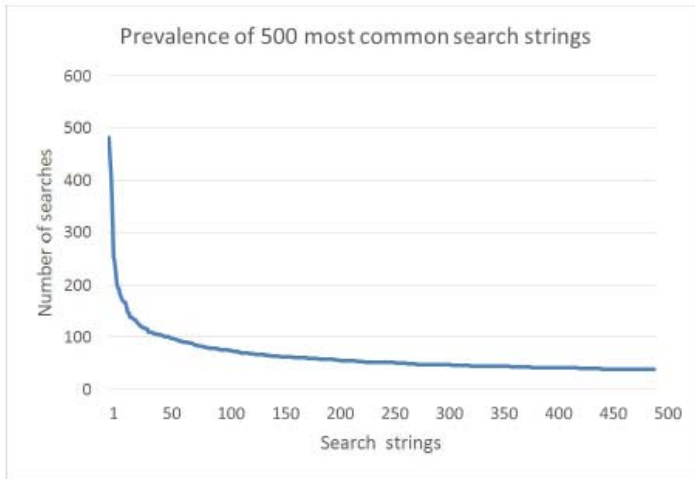[1]   <https://support.google.com/analytics/answer/2992042?hl=en>.

Figure 3: A few strings get looked up often, but as can be seen from the graph there is a long tail of search strings that have few lookups. The graph only shows the 500 most common, but the tail is a lot longer as over 170,000 search strings have been entered.

5. What next?

Most of the resources accessible through *málið.is* are constantly being updated. They will continue to be so while new resources will be added to the portal, as outlined in section 3. The data on users, shown in section 4, indicates there may be room for improvement in promoting the portal. Steps have been taken to inform teachers of the portal, but currently no plans for further promotion have been made.

References

Ásgeir Blöndal Magnússon (1989): *Íslensk orðsifjabók*. Reykjavík: Orðabók Háskólans.

Hilmarsson-Dunn, Amanda & Ari Páll Kristinsson (2013): The language situation in Iceland. In: Robert B. Kaplan, Richard B. Baldauf, Jr. & Nkonko M. Kamwangamalu (Eds.): *Language Planning in Europe: Cyprus, Iceland and Luxembourg*. London / New York: Routledge, 100–169.

Íslensk málnefnd. (2016): *Ritreglur. Auglýsing mennta- og menningarmálaráðuneytis nr. 695/2016 með leiðréttingum*. <http://islenskan.is/images/ritreglur-IM-2016.pdf> (Retrieved September 10, 2017).

Íslensk nútímamálsorðabók (2017): Halldóra Jónsdóttir & Þórdís Úlfarsdóttir (Eds.). <http://islenskordabok.arnastofnun.is> (Retrieved September 10, 2017).

Jóhannes B. Sigtryggsson (Ed.) (2016): *Stafsetningarorðabókin*. 2. Ed. <www.malid.is> (Retrieved September 10, 2017).

Jón Hilmar Jónsson (2018): Íslenskt orðanet: Tekstbasert kartlegging og presentasjon av leksikalske relasjoner. In: *Nordiske Studier i Leksikografi* 14 (this rapport). Reykjavík: NFL & Stofnun Árna Magnússonar í íslenskum fræðum, 1–17.

Kristín Bjarnadóttir (2012): The Database of Modern Icelandic Inflection. In: *Language Technology for Normalisation of Less-Resourced Languages SALT-MIL 8 – AfLaT 2012*, 13–18 <http://aflat.org/files/saltmil8-aflat2012.pdf> (Retrieved September 10, 2017).

Málfarsbankinn (2017): Jóhannes B. Sigtryggsson (Ed.). <http://malfar.arnastofnun.is> (Retrieved September 10, 2017).

Steinþór Steingrímsson   Ari Páll Kristinsson
IT-ansvarlig   Forskningsprofessor
steinthor.steingrimsson@arnastofnun.is   ari.pall.kristinsson@arnastofnun.is

Halldóra Jónsdóttir
Ordbogsredaktør
halldora.jonsdottir@arnastofnun.is

Árni Magnússon-instituttet for islandske studier
Laugavegi 13
101 Reykjavík, Island