

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

Halldóra Jónsdóttir, Ari Páll Kristinsson, Steinþór Steingrímsson

The Árni Magnússon Institute for Icelandic Studies

E-mail: halldo@hi.is, aripk@hi.is, steinst@hi.is

Abstract

Málið.is is a web portal on the Icelandic language and language use. It currently includes seven different resources, providing reliable information on the language. Six of the seven resources made available on the portal are living projects, constantly being updated. *Málið.is* provides easy and fast access to these, accessible on and designed for desktop and mobile devices. One of these resources contains a number of specialized terminologies. Thus it is a special characteristic of *málið.is* among language portals for other languages that it provides access to abundant domain specific vocabulary. Work is under way to add more resources to the portal, providing more diverse information. The aim of the project is to strengthen the Icelandic language in the digital era by making it easy to access information on the language, helping people becoming more proficient and confident language users.

Keywords: e-Lexicography, Icelandic, dictionary portal, orthography, phraseology, etymology, terminology

1 Introduction

Users' access to all kinds of information has been subject to great changes over the past two decades, as the demand for easy access to dictionaries and other lexical information has increased. Moreover, online linguistic information that is free of charge can play an important role in language policy and planning, not least for small- and medium-sized language communities, such as the Icelandic one. Iceland is a nation of 340,000 people, and approximately 90% of the population are native speakers of Icelandic.

Málið.is is a new web portal on the Icelandic language and language use. It is operated by The Árni Magnússon Institute for Icelandic Studies, a publicly funded academic institute, dedicated to a variety of tasks in the field of Icelandic language and literature. The web portal opened in November 2016 and at present provides access to seven different resources created and/or maintained by the Institute. Together, the resources provide reliable in-depth information on Icelandic. Work is still in progress to add more resources to the portal, which will give more diverse information.

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.

Users of the portal can find abundant data and directions on language use and usage, inflections, semantics, stylistic variation, etymology and more. But it also gives access to large amounts of terminological data, as it searches through a database of more than sixty terminological glossaries, each for a specific field of study. The principal target group for *málið.is* are students and professional language users, including writers, journalists and translators. But as the web portal also serves the general Icelandic speaking public, as we strive for plain and non-technical exposition and conciseness.

All but one of the dictionaries and datasets available on *málið.is* are living projects, constantly being updated and revised. The portal runs a separate copy of all the databases, adapted for fast search and access on the *málið.is* website. Every night the server automatically launches a program that checks for recent changes and updates the *málið.is* database accordingly. When searching on *málið.is* users get the local copy of the data, but as most of the resources included in *málið.is* have their own website, users are also provided with a link to the corresponding entries in each of the dictionaries or databases.

2 Rationale for Setting up the Portal

The main aim of the *málið.is* project is to strengthen the Icelandic language in the digital era by making it easy to access coded linguistic information, thus helping the public to become more proficient and confident language users. Digital resources for Icelandic are far more limited than for most other national languages in Europe. For that reason, it is even more important that the good quality data that already exists becomes as clearly noticeable as possible for users.

The Árni Magnússon Institute for Icelandic Studies plays a 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, offers language consultation and advice to the general public, and compiles text corpora and other language resources. Moreover, researchers at the Institute have edited and published various dictionaries and other diverse language materials over the years. This work constitutes the core of the material we present on *málið.is*.

A great inspiration in preparing *málið.is* was the Institute's participation in the COST Action IS1305: *A European Network of e-Lexicography* (European Network of e-Lexicography 2013), and its ongoing discussions about convenient access to scholarly and advanced dictionaries. In Iceland we have also seen growing demands in recent years for dictionaries and other language resources to be made available free of charge to the general public. Furthermore, there is widespread interest in Iceland among lay people 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 & Kristinsson 2013), and the A great number of the queries the Institute gets from the general public shows this. With *málið.is* we try to fulfil these needs as well as possible.

3 The Resources

Seven resources are currently accessible through the portal. Initially we wanted the information disseminated through the portal to be highly relevant to everyday use of Icelandic. The decision on what was to be included was therefore based on that, and on the state of the resources' database – whether we could connect it to the portal with relatively little effort. These are the seven resources:

- **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.
- **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).
- **Dictionary of Modern Icelandic** (*Íslensk nútímamálsorðabók* 2018). The Dictionary of Modern Icelandic is a new dictionary only available online. It is compiled at the Department of

Lexicography at The Árni Magnússon Institute for Icelandic studies. It contains approximately 50 thousand words. The work on the dictionary commenced in 2013 and is ongoing.

- **Language Usage Database** (*Málfarsbankinn* 2018). A compilation of short articles giving advice on language use. It contains more than 7,000 articles on subjects from grammar and syntax to well-crafted and elaborate language.
- **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 is constantly growing.
- **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.
- **Icelandic Wordnet** (Jón Hilmar Jónsson 2018). *Íslensk orðanet* (Icelandic Wordnet) is a semantic database with elements that allow it to be used much like an onomasiological dictionary or a thesaurus.

The web portal will grow in the coming years and work is under way to add more resources. Potential datasets to be added include a written language archive with usage examples dating from 1540 to the late 20th century, historical dictionaries and bilingual dictionaries. These resources could make the portal more useful to students of the language as well as for everyday users of Icelandic.

4 Expanding the General with the Specific

The interests and needs of regular users of language for special purposes (LSP) often tend to be overlooked when information on language and language use is compiled and disseminated through web portals and the like. Among the most innovative and important special features of *málið.is* is that a large number of specialized terminologies, containing domain specific vocabulary, can be accessed through the portal along with information on words and usage in the language in general. Thus, the language portal *málið.is* is particularly useful to authors of textbooks, translators, and scientists, to name but a few groups that write and read LSP texts on a regular basis. *Málið.is* provides information on grammar, orthography, phrases, semantics, etc., of the general vocabulary of Icelandic, and, in addition, it retrieves content from the about 60 different terminologies – containing about 180,000 terms in total – that constitute the Icelandic Term Bank (ITB).

Among the roles of the Árni Magnússon Institute for Icelandic Studies is to facilitate the collection of domain specific terms, and the coordination of their usage and definitions. The Institute operates the ITB (opened on the Internet in 1997) for this purpose. Since Icelandic is the national language of Iceland, it is an important national language policy goal that Icelandic terminologies exist and are readily accessible to language users in a variety of domains of society, science and technology. The Institute is responsible for both the ITB and for the databases that contain information on general Icelandic vocabulary. This favorable situation made it a natural choice for the language portal *málið.is* to incorporate the voluminous terminological data at hand.

Through *málið.is*, one can of course search for specialized terms by choice, as well as for other words, as the need may arise. But the portal also “recruits” new users of terminological resources as they become aware of existing terminologies when they search for information among the other types of

language resources and dictionaries. A number of users may become aware of a domain specific usage of a number of Icelandic words, for the first time, by learning of it when using *málið.is*.

For example, someone might type in the search string *ferill* ‘trail, course’ in order to find out, for example, something about the orthography or the (complex) inflection of this particular Icelandic noun. In addition to such information, *málið.is* instantly exposes the user to abundant data on *ferill* as a term in a number of specialized terminologies; more precisely, 17 hits, retrieved from 11 different terminologies, in the following domains: statistics, physics, aviation, computer science, medicine, economics, engineering, seamanship, political science, geography, and epidemiology, as shown in Figure 1. By clicking on a red underlined string in the results screen, the user is transferred to that specific database.

5 Concluding Remarks

The web portal *málið.is* has been open to the public, free of charge, since November 2016 and more and more language users find their way to it. Bringing seven resources together in one portal is a new way to provide access to reliable in-depth information on Icelandic in the digital era, and by doing this users are directed to resources they do not know exist and are otherwise unlikely to find. Four of the resources were accessible on various websites run by the Institute, prior to the opening of *málið.is*. The others had either not been made accessible or were available only in print.

By giving access to specialized terminologies through the portal we aim to serve users of terminological collections as well as helping others become aware of the domain specific usage of a number of words.

Work is under way to add more resources. Amongst others four historical dictionaries, which have never been published online before, will be made available through *málið.is* as an experiment to explore if there is a demand for such resources on a portal like this one, and therefore whether they belong there.

References

- Bjarnadóttir, Kristín (2012): The Database of Modern Icelandic Inflection. In: *Language Technology for Normalization of Less-Resourced Languages SALT MIL 8 - AfLaT 2012*, 13–18. Accessed at: <http://aflat.org/files/saltmil8-aflat2012.pdf> [28/03/2018].
- European Network of e-Lexicography (2013): Cost Action IS1305. Accessed at: <http://www.elexicography.eu/> [01/12/2017].
- Hilmarrsson-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. Accessed at: <http://islenskan.is/images/ritreglur-IM-2016.pdf> [28/03/2018].
- Íslensk nútímamálsorðabók (2018): Halldóra Jónsdóttir & Þórdís Úlfarsdóttir (Eds.). Accessed at: <http://islenskordabok.arnastofnun.is> [28/03/2018].
- Jónsson, Jón Hilmar (2018): “Íslenskt orðanet: Tekstbasert kartlegging og presentasjon av leksikalske relasjonjer”. Rapport fra Konference om leksikografi i Norden. Island 30. maj-2. juni 2017 [forthcoming].
- Magnússon, Ásgeir Blöndal (1989): *Íslensk orðsifjabók*. Reykjavík: Orðabók Háskólans.
- Málvarsbankinn* (2018): Jóhannes B. Sigtryggsson (Ed.). Accessed at: <http://malvar.arnastofnun.is> [28/03/2018].
- Sigtryggsson, Jóhannes B. (Ed.) (2016): *Stafsetningarorðabókin*. 2. Ed. Accessed at: www.malid.is [28/03/2018].