

ISLEX—An Icelandic-Scandinavian Multilingual Online Dictionary

Aldis Sigurðardóttir

Anna Hannesdóttir

Håkan Jansson

Halldóra Jónsdóttir

Lars Trap-Jensen

Pórdís Úlfarsdóttir

Society for Danish Language and Literature

This paper presents ISLEX, an inter-Nordic project based in Reykjavik, Iceland, with partners in Gothenburg, Bergen and Copenhagen. The aim of the project is to develop an online dictionary site with Icelandic as the source language and the three Scandinavian languages—Swedish, Norwegian (with two official standards) and Danish—as the target languages. The dictionary is planned to contain 50,000 lemmas, with a development period of six years. In 2011, or possibly sooner, the site will be publicly available on the Internet, free of charge. In this article, the main features of the project are presented with particular emphasis on database design, editorial principles and priorities.

1. Project background

ISLEX is organised as an overhead project consisting of four separate projects at four partner institutions. The largest editorial staff, responsible for the description of the source language (SL) and for the development and maintenance of the Icelandic database is located at *The Árni Magnússon Institute for Icelandic Studies* at the University of Iceland in Reykjavík. Partners and those responsible for the Scandinavian target language (TL) equivalents are *The Society for Danish Language and Literature* in Copenhagen, Denmark (for Danish), *The Department of Nordic Languages and Literature* at the University of Bergen, Norway (for the two Norwegian standard languages: Bokmål and Nynorsk), and *The Department of Swedish* at the University of Gothenburg, Sweden (for Swedish). The governments of all four countries have contributed to the funding of the project which runs for a period of six years before the final launch in 2011.

The motivation behind the initiative was an urgent need for a modern, up-to-date dictionary between Icelandic and the other Scandinavian languages. Put bluntly, existing dictionaries (mainly printed books) are either too limited in the stock of words covered, outdated, or falling short of the lexicographic standard required of dictionaries today.

1.1. Size and availability

ISLEX is designed as a medium-sized dictionary with 50,000 entry words. Particular attention is devoted to phraseology, and many entries contain a large number of phrases and fixed expressions. Other information types include example sentences, labels indicating register and subject field, grammatical information, pictures and recorded pronunciation.

The aim is to make ISLEX available to the public on the web, free of charge. Even though the final release is scheduled for 2011, a gradual introduction is being discussed and a preliminary version might be launched sooner. Updates would then be provided at regular intervals as the editorial work progresses and the functionalities of the site are implemented.

2. Target groups

Being the mother tongue of some 300,000 speakers and with only a limited number of non-native students and scholars outside Iceland, Icelandic is too “small” a language for ISLEX to focus only on one narrow user group. Two groups of users have been considered of primary importance from

the early planning phase: Scandinavians with elementary to advanced knowledge of Icelandic (teachers, students and translators) constitute the largest reception-oriented user group, and Icelanders with medium to advanced knowledge of a Scandinavian language (teachers, students and translators) are the most important production-oriented users. The second group is also the largest in terms of absolute numbers of future users. Apart from these groups, ISLEX will try to meet the needs of others who need to consult a dictionary between Icelandic and the Scandinavian languages—as long as the needs are not in conflict with those of the primary user groups.

The digital medium offers flexible ways of defining user profiles and rather than narrowing the target group beforehand, it seems worthwhile pursuing these possibilities in the interface design. This means that the staff will have to provide all the information relevant to the needs of the different user groups throughout the editing phase, even if it means supplying information which is seemingly redundant as seen from the point of view of one particular user group, for instance grammatical information on both SL and TL items, or multilingual metatexts. The point is that only the relevant information will be extracted and presented in the user interface once a specific profile has been selected by the user.

3. Database design and data structure

An important feature of the ISLEX dictionary is its four TLs: Danish, Norwegian Bokmål, Norwegian Nynorsk and Swedish. This had to be taken into account at the beginning of the project, and key elements in the choice of software were flexibility—to allow for variation in the TL solutions—and the possibility for remote access to the central software—in order to facilitate project management—. It was decided to develop a database especially designed to meet the requirements of ISLEX. This task was initiated in early 2004, and by March 2004 some 46,000 Icelandic words were imported into the system.

3.1. Data types

The dictionary is set in an object-relational database, the open-source system PostgreSQL for Debian GNU/Linux, using Perl as the programming language. The database handles several data types each of which has special properties.

The basic and most common data type is *text* which is used for most of the information elements. In most cases, the text can be entered freely from the editor's keyboard whereas some elements have a fixed inventory which must be chosen from a drop-down menu, e.g. the labels indicating subject and register. There are two kinds of hyperlinks: *internal links* for cross-references within the dictionary allowing navigation between articles with a click, and *external links* to web pages outside the database. In ISLEX, they may be used for a number of purposes, but one purpose needs to be singled out.

As one of the most conservative of the modern Germanic languages, Icelandic has retained its inflectional richness, and it is therefore crucial to provide full information about the inflectional paradigms for the lemmas in the SL. This information is provided through linking to a database of Icelandic inflections, *Beygingarlýsing íslensks nútímamáls*, which was developed and is maintained by *The Árni Magnússon Institute*. The morphological database contains full paradigms for some 220,000 words which are directly available for the ISLEX project. The remainder of the words in the base serve as a valuable source from which examples can be taken to show the derivational and compounding word formation possibilities of a headword.

The data type *files* is for text files. These are sample pages from a concordance of various Icelandic texts which constitute helpful reference material for the editors of both the SL and the TL. They are added to the database by means of an uploading function.

Images constitute a special data type which are also inserted via an uploading function. Images can be either pictures or animated sequences and they are used in ISLEX with a dual purpose. They add to the understanding of a headword, and they have an aesthetic and entertaining function which in itself might attract the interest of a potential user or chance passer-by whose attention would otherwise be hard to maintain. See figure 1 for an example.








| kría n f | |
|---|---|
| 1 | HLUTAR kría |
| 2 | SAMSETN  |
| 3 | SKÝRING fuglategund (<i>Sterna paradisaea</i>) |
| 4 |  1 WWW  www.fauna.is |
| 5 | DA-jafn havterne  |
| 6 | NO-B-jafn rødnebbterne  |
| 7 | NO-N-jafn rødnebbterne  |
| 8 | SÆ-jafn silvertärna  |
| 9 | MSVIÐ dýr |
| 10 | STIG (0-3:) í-2 D-1 N-1 S-1 |
| brotið strúktúr XML | |

Figure 1. Example of an article with illustration: *kría* (“arctic tern”)

Finally, there is the data type *sounds*. This is mainly intended for the recorded pronunciation of a lemma. The plan is to have pronunciation for the lemmas as recorded speech. In addition, natural sounds may be used to facilitate understanding of a word meaning, for instance words designating objects or creatures with a characteristic sound, such as *bell*, *gun* or *sea lion*, or onomatopoeic words such as *beep*, *splash* and *crash*. Potentially, sounds may be used for video clips as well, even though none has as yet found its way into ISLEX.

Technically, the format of all text categories in the database is flexible. They can have user definable properties, i.e. colours, font size and font type (normal, bold face, italic or underscore), categories may be preceded by an icon or fixed text, and drop-down lists or show/hide function can be made for any textual element. Whether the possibilities are in fact utilized, is a matter of editorial principles.

3.2. Hierarchy

The ISLEX database is hierarchically structured, meaning that the microstructural information of an entry is arranged in recursive layers: each element may have one or more elements appearing at either coordinate or subordinate level, and the same holds for sub-elements.





































| kort n n | | | | | | | | | | | | | | | | | | | | | | |
|----------|---|--|--|---|--|--|--|--|---|--|---|---|---|---|---|--|---|--|---|--|---|---|
| 1 | LIDUR | <table border="1"> <tr> <td colspan="2">1</td> </tr> <tr> <td>1</td> <td>DA-jafn kort </td> </tr> <tr> <td>2</td> <td>NO-B-jafn kart </td> </tr> <tr> <td>3</td> <td>NO-N-jafn kart </td> </tr> <tr> <td>4</td> <td>SÆ-jafn karta </td> </tr> <tr> <td>5</td> <td>OSTÆDA kort af <landinu> <table border="1"> <tr> <td>1</td> <td>DA-jafn kort over <landet> </td> </tr> <tr> <td>2</td> <td>NO-B-jafn kart over <landet> </td> </tr> <tr> <td>3</td> <td>NO-N-jafn kart over <landet> </td> </tr> <tr> <td>4</td> <td>SÆ-jafn karta över <landet> </td> </tr> </table> </td> </tr> </table> | 1 | | 1 | DA-jafn kort  | 2 | NO-B-jafn kart  | 3 | NO-N-jafn kart  | 4 | SÆ-jafn karta  | 5 | OSTÆDA kort af <landinu> <table border="1"> <tr> <td>1</td> <td>DA-jafn kort over <landet> </td> </tr> <tr> <td>2</td> <td>NO-B-jafn kart over <landet> </td> </tr> <tr> <td>3</td> <td>NO-N-jafn kart over <landet> </td> </tr> <tr> <td>4</td> <td>SÆ-jafn karta över <landet> </td> </tr> </table> | 1 | DA-jafn kort over <landet>  | 2 | NO-B-jafn kart over <landet>  | 3 | NO-N-jafn kart over <landet>  | 4 | SÆ-jafn karta över <landet>  |
| 1 | | | | | | | | | | | | | | | | | | | | | | |
| 1 | DA-jafn kort  | | | | | | | | | | | | | | | | | | | | | |
| 2 | NO-B-jafn kart  | | | | | | | | | | | | | | | | | | | | | |
| 3 | NO-N-jafn kart  | | | | | | | | | | | | | | | | | | | | | |
| 4 | SÆ-jafn karta  | | | | | | | | | | | | | | | | | | | | | |
| 5 | OSTÆDA kort af <landinu> <table border="1"> <tr> <td>1</td> <td>DA-jafn kort over <landet> </td> </tr> <tr> <td>2</td> <td>NO-B-jafn kart over <landet> </td> </tr> <tr> <td>3</td> <td>NO-N-jafn kart over <landet> </td> </tr> <tr> <td>4</td> <td>SÆ-jafn karta över <landet> </td> </tr> </table> | 1 | DA-jafn kort over <landet>  | 2 | NO-B-jafn kart over <landet>  | 3 | NO-N-jafn kart over <landet>  | 4 | SÆ-jafn karta över <landet>  | | | | | | | | | | | | | |
| 1 | DA-jafn kort over <landet>  | | | | | | | | | | | | | | | | | | | | | |
| 2 | NO-B-jafn kart over <landet>  | | | | | | | | | | | | | | | | | | | | | |
| 3 | NO-N-jafn kart over <landet>  | | | | | | | | | | | | | | | | | | | | | |
| 4 | SÆ-jafn karta över <landet>  | | | | | | | | | | | | | | | | | | | | | |
| 2 | LIDUR | 2 | | | | | | | | | | | | | | | | | | | | |

Figure 2. Database view of the entry *kort* (“map”)

In figure 2, three hierarchical levels are given for the lemma *kort* (“map”). The number “1” is the topmost level indicating the first of three senses. At the second level are the four TL equivalents and a phrase labelled *OSTÆÐA* (collocation). The third level contains the four equivalents given for the collocation.

4. Work flow and editorial practice

The editorial process is organised around semantic fields where words are translated in packages with related meanings rather than applying a traditional alphabetical procedure. Each lemma is supplied with a marker indicating a semantic field. For example, the word *gulrót* (“carrot”) has the marker *gras/matur* (“vegetable/plant, food”) and the word *brúðkaup* (“wedding”) is marked with *fundur/samkoma/skemmtun brúðkaup/hjónaband* (“meeting/gathering/amusement, wedding/marriage”). In the editorial process each semantic field is processed as a whole to ensure consistency for words with similar meanings. Once the SL processing has been completed, the package is released to the TL editors who proceed in the same way with the translation task. The procedure is convenient for the TL editors as they can concentrate on one particular semantic field at a time, thus saving time and effort with respect to information seeking. It also sharpens their attention on the best choice of equivalent as they are forced to make contrastive comparisons. The flexible structure of the database allows for individual policies to be pursued by the different TL projects, resulting in slightly different end products (cf. section 6.1.).

Apart from these rather narrow means of communication through the database, the staff have somewhat special working conditions, having to collaborate in four different countries with only limited possibility of face-to-face meetings. It should be mentioned, though, that in addition to regular contact via wiki (a common web-based forum using the technology known from Wikipedia and other resources), e-mail and telephone, the staff meet biannually for discussion and coordination of common issues, and a third meeting for TL editors is currently under consideration. See also section 5.2. for a way of communicating via the database interface.

In order to manage work flow, a numbering system is used to label the different editorial stages which a lemma must go through. A separate element in the database specifies how far a lemma has come in the process for each of the four national staffs. Once a lemma is imported into the database it is given a number (Í-1). This level indicates that the lemma exists, but that no editing work has yet taken place. When the lemma is finished on the SL side, it is promoted to level Í-2, and once all the lemmas of a thematic group are complete they are released to the TL editors for processing. From the outset, the TLs are marked with a zero (D-0, N-O, S-0) and each TL editor must make a change corresponding to the editorial status, using a scale from 1 to 3 or 1 to 6, dependent on the part of speech. If the contents of a lemma are changed by the SL team (for instance, as a response to a request put forward by a TL editor in a note, cf. 5.2.), its status is reset and has to be checked again by the TL editor. This procedure has proved convenient in the daily work and saves a lot of e-mail correspondence. An additional advantage is that the information is stored in the base and is readily retrieved by the editors through a combined search for thematic group and status level.

5. Information on the source language

5.1. Lemma selection

The language of description is modern Icelandic, and this is reflected by the stock of entry words. Words and word meanings not found in contemporary use, but belonging exclusively to Old Icelandic (Old West Norse) do not appear in the dictionary, with but a few exceptions. Students and translators of Icelandic occasionally come across words from medieval literature. For that reason, a limited number of significant words from the older language are retained in the lemma list, but they will contain a register label clearly indicating that the word is obsolete.

Overall, the lemma selection procedure for ISLEX has been based on a combination of several criteria. The most important considerations are:

- Frequency of the word in modern written texts.
- Frequency of the word in the spoken language, even if it has not (yet) found its way into the written language.
- Some less-frequent words are included if they are part of the established vocabulary of a specialised subject field: geology, law, medicine, music etc.

The criterion for inclusion of compound words is a combination of frequency and non-transparency of meaning. Icelandic contains a large number of compounds and is very productive in making new ones. The meaning of some compounds cannot be deduced directly from their components; they are lexicalised and behave like single units. Such compounds have been treated on a par with simplex words in the selection procedure. Other compounds, on the other hand, will be semantically clear from its parts, and for that reason the frequency threshold for inclusion is placed higher as they are less obvious lemma candidates.

5.2. Sense differentiating and contrastivity

The basic idea behind the ISLEX database is that the same SL description can be used as a starting point for a number of bilingual dictionaries of which the Scandinavian languages are the first. As a consequence, the semantic description of the Icelandic vocabulary is based on a monolingual analysis of the language, in principle independent of any TL. This may not seem an optimal solution as seen from a strictly lexicographical point of view, but nevertheless one that was judged justifiable given the economic and political reality which rules out repeated contrastive descriptions for every new bilingual project.

To a certain extent, however, it is possible to take some contrastive aspects into consideration in the ISLEX project, should the need arise. The database has a special element for each TL, reserved for staff internal communication. Here the TL editors can raise a problem, make a proposal or express a wish for a revision of the SL article if they find the SL description insufficient to deal with a contrastive problem they have encountered during their work with a particular word. For example, the TL editors may ask for splitting up a sense if the Icelandic word is rendered by two distinct equivalents in the TL, or an example or collocation may be unfortunate for contrastive reasons. At other times, the request may just be for an (extra) example or a picture, or any other matter in relation to the article at hand.

5.3. Phraseological Information

A priority of ISLEX is the many types of phraseological material: collocations, example sentences, idioms and other set phrases. Apart from the large phraseological dictionary Jónsson (2005), a comprehensive treatment of Icelandic phrases is nowhere to be found in reference works, and certainly not with full translations into other languages. So far, approximately half of the SL entries have been completed, of which about one third of the articles contain one or more phraseological items, giving a total of 16,000 examples, collocations or idioms. This number is not definite as new material may be added during the TL editorial work. For collocations and idioms, 90% of the material has been taken from the above-mentioned Jónsson (2005), and the rest from the Icelandic editors' own collections.

The phraseological material will be supplied with TL translations for all types, i.e. both collocations, idioms and example sentences, in order to comply with the needs of Scandinavian users with only elementary knowledge of Icelandic.

5.4. Examples

The dictionary includes many example sentences. These are not necessarily genuine citations taken from authentic texts, but are often inspired by authentic examples and adapted to the dictionary by the editors. The examples show the words in a typical linguistic context illustrating characteristic syntactic behaviour, idiomatic usage or providing additional semantic or encyclopaedic information to the lemma.

5.5. Labels

Register labels are used in a way similar to most dictionaries today. The labels used in ISLEX are of two kinds: subject labels and register labels. Word senses belonging to particular subject fields are labelled accordingly, e.g. *physics*, *anatomy* or *law*. Register labels indicate a restriction in usage carrying particular connotations based on style or social parameters such as age, social class or geographic distribution, e.g. *archaic*, *formal* or *derogatory*. In both cases, the inventory is fixed and must be chosen from a drop-down menu. The labels are presented both in the SL and the TL in order to offer the information in the mother tongue of both primary user groups. An example is shown in figure 3. Similar labels are used for the TL equivalents (cf. section 6.6.).








| böggur n m | | |
|------------|-----------|--|
| 1 | HLUTAR | böggur |
| 2 | SVIÐ | tölvur |
| 3 | NOTKUN | slangur |
| 4 | SKÝRING | villa í forriti (enska bug) |
| 5 | DA-jafn | bug  |
| 6 | DA-jafn | programmeringsfejl  |
| 7 | NO-B-jafn | bug  |
| 8 | NO-B-jafn | programfeil  |
| 9 | NO-N-jafn | bug  |
| 10 | NO-N-jafn | programfeil  |
| 11 | SÆ-jafn | bugg  |

Figure 3. The article *böggur* (“bug”) has been supplied with the labels *tölvur* (“information technology”, element 2) and *slangur* (“slang”, element 3)

6. Information about the target languages

6.1. The target languages

A basic function of bilingual dictionaries is to facilitate contact between speakers of different languages, i.e. they are intended as an aid in translation in a broad sense. However closely related—in what almost can be described as a symbiosis—there is a theoretical gap between the two disciplines sharing translation as the object of study: translation science and bilingual lexicography. The key concept in both disciplines is *equivalence*. While the definition of the concept is constantly at issue in translation theory it is rather poorly defined in the discipline of lexicography.

In the Icelandic-Scandinavian projects, the structure of the database allows the different TL projects to adapt an individual policy concerning the final product of that specific bilingual project. This applies e.g. to the depth of the contrastive lexical analysis of the relationship between the Icelandic entity and the means provided by the TL expressing the different qualities of the lemma, as well as to the presentation of the result of the contrastive analysis. The overall project allows various levels of theoretical ambitions regarding the development of a genuine contrastive lexical dimension in the bilingual lexicographical description.

6.2. Editing the target languages

As already mentioned, there are four TLs involved in the ISLEX project: Danish, the two official Norwegian standards, Bokmål and Nynorsk, and Swedish. The TL lexicographers, spread across Scandinavia, are working within one network which has the Icelandic database at its centre. The editing process goes on simultaneously in the different languages and everything that is inserted in an article in the database is immediately displayed in the editorial interface for the lemma in question. For practical reasons, however, the editor of one TL avoids working with a semantic field at the same time as a colleague of another TL.

There is an ongoing debate among the editors whether it is an advantage or a disadvantage to be the last one to enter the scene of a semantic field. On the one hand, the lexicographer may, in many cases, take advantage of the work already done by the two previous colleagues. Interestingly, even though the four TLs are closely related and, indeed, in many aspects similar to each other, they are autonomous enough to require individual lexicographic solutions when it comes to the presentation of relevant properties of a SL unit in the TLs—solutions that are often asymmetrical with respect to the equivalents chosen—. However, the equivalents provided by the first in the field in question can, in some cases, bring appropriate alternatives to the fore and thus save some time and effort for the following colleague, while in other cases there is a risk of misleading the successive editors into an intricate, quadrilingual mishmash of false friendship. On the other hand, the first editor is encumbered with a substantial amount of initial research in order to come to grips with a semantic field—often a time-consuming analysis of (semi-)technical terms.

6.3. *Equivalents and paraphrases/explanations*

The ISLEX database offers each of the TLs two alternatives for the presentation of the translation equivalent of a lemma. One alternative is meant for equivalent or semi-equivalent lexical units consisting of a single word or a lexical collocation of words. The other alternative is meant for explanations, either in the form of a paraphrase or an encyclopaedic explanation.

As can be expected, quite often the TL equivalent does not fully match the meaning of the Icelandic word. In this context, it is not possible to exemplify all the different forms of full, partial or zero lexical equivalence or different kinds of pragmatic equivalence that occur between the SL units and one or all of the TLs. Let us, however, consider some aspects of the problems involved and their possible solutions.

In case of near-equivalence the editor has the option to add an explanatory note in a separate sub-element. The note may comment on either the SL lemma or on the TL equivalent. As an example, consider first the lemma *afréttur* (“pasture land”) which has been given the Danish equivalent *græsgang*. A problem here is that the equivalent only partly covers the meaning of the Icelandic word because of the difference in farming tradition and the natural landscape in the two countries. In this case, the Danish editor has added an explanatory note commenting on the Icelandic lemma: “i vildmarken; fælles græsning for et områdes fårehold” (“common grassland for sheep in the countryside”). Please note that, although the comment regards the Icelandic lemma, the problem is confined to the Icelandic-Danish contrastive relation and does not affect Norwegian or Swedish.

An example of a more symmetrical equivalence structure would be the Icelandic word *afabróðir* which is translated as *grandonkel* (“great-uncle, parent’s uncle”) in both Norwegian and Danish. However, as the Icelandic word is constructed as a compound of *afi* “grandfather” (genitive *afa*) and *bróðir* “brother”, it is necessary to supply the lemma with an additional note restricting the meaning to “grandfather’s brother”. In both Scandinavian languages, the equivalent *grandonkel* may be used for any uncle of one’s parents, including the brother of one’s grandmother, which is not covered by the Icelandic word.

In other instances, a comment is needed if the Scandinavian equivalent is polysemous. A short gloss will clarify which of the meanings is the relevant one.

Another typical situation is a semantic restriction which applies to one or more of the equivalents. Figure 4 shows an example, the lemma *áttatíu* (“eighty”), where the note applies to only one of the TLs. Icelandic, Swedish and Norwegian all use the decimal system of counting whereas the Danish numerals are derived from a vigesimal system. So the unmarked Danish equivalent is *firs* (etymologically “four times twenty”, cf. French *quatre-vingt*). However, a decimal numeral does exist in Danish, *otti*, but this word is restricted in usage to cheques, receipts and financial or legal documents. Therefore, it is appropriate to bring the word *firs* as an equivalent, but a note is needed to draw attention to the usage restriction.






| áttatíu num | |
|-------------|---|
| 1 | HLUTAR átta-tíu |
| 2 | SKÝRING 80 |
| 3 | DA-jafn firs  |
| 4 | DA-jafn otti  1 KOM-da især på checks, dokumenter m.m. som supplement til beløbsangivelse med cifre |
| 5 | NO-B-jafn átti  |
| 6 | NO-N-jafn átti  |
| 7 | SÆ-jafn áttio  |
| 8 | MSVIÐ tala |
| 9 | STIG (0-3:) Í-2 D-1 N-1 S-1 |

Figure 4. A note informs that the Danish numeral *ottí* (“eighty”) can be used as an equivalent for *áttatíu*, but only under special circumstances (bank transactions)

Other notes may comment on the orthography, on an abbreviation or the syntactic behaviour of the Scandinavian equivalent.

It is important to recognize the distinction between notes relating to the Icelandic SL word and notes relating to the Scandinavian TL equivalents. The presentation of the notes in the end product is dependent on the user group and the user situation, and therefore the notes are entered into different elements in the database.

6.4. *Word combinations and phrases*

A large portion of the entries at the sub-lemma level consists of collocations and idioms; they are indexed under their respective lexical headwords, but also as collocations or idioms. This means that they can be retrieved and processed independently by the TL editors. The flexibility of the database also allows each of the TL projects to have their own indexation for local purposes. One such project is an ongoing Swedish investigation of collocations and idioms. The SL idioms will be classified by their meaning and linked to Swedish collocations and idioms expressing similar meaning—independently of the lexical items they are composed of (cf. Hannesdóttir and Jónsson (2001) for a methodological discussion).

6.5. *Free example sentences*

Working with four TLs simultaneously can be a demanding task for the SL editors. Finding illustrative examples that reveal both the characteristics of the SL unit and also the contrastive relationship to each of the TLs is quite a challenge. Sometimes an unforeseen translation problem calls for a substitution of the example sentence; at other times it may lead to insertion of extra material in addition to the example(s) already presented. In extreme cases, it happens to be the last of the TL editors who faces the problem—and the work of the previous two (the Norwegian editor takes care of both Norwegian TLs) has been in vain—. In most cases, however, the work runs smoothly and the examples fulfil their explanatory function by highlighting features that cause problems for the different user groups, such as structural change (e.g. nominalization vs. verbalization), grammatical aspects, idiomaticity and pragmatics.

6.6. *Labels*

In order to ensure consistency the TL editors are provided with a fixed set of labels from a drop-down menu, but in addition it is possible to enter more specialized comments in an open text field, if needed. Labels regarding the TL equivalents may be added by the TL editor. For the TL side, there is one element reserved for labels concerning register (formal, informal, old fashioned, derogatory etc.) and another for grammatical comments (for example change of part of speech, if the equivalent is used in the plural or in definite form only). An example is given in figure 5.




| 1 HLUTAR | | refsi-réttur | |
|----------|-------|--------------|--|
| 2 | LÍÐUR | 1 | |
| | | 1 | SKÝRING refsilög, þ.e. þær réttarreglur almennt er varða refsiverða háttsemi og viðurlög við henni |
| | | 2 | DA-jafn strafferet  |
| | | 1 | TILFØJELSE retsregler |
| | | 3 | SÆ-jafn brottsbalk  |
| | | 1 | NOT-sæ vanligen i bestämd form |
| | | 4 | SÆ-jafn strafflag  |
| | | 1 | NOT-sæ vanligen i bestämd form |
| | | 2 | NOT-sæ-sn historiskt |
| | | 1 | NOT-sæ2 ersattes 1962 av brottsbalken |

Figure 5. In the article refsiréttur (“criminal law”), the Swedish equivalent strafflag in element 4 has been supplied with the information 1) that it is mostly used in the definite form, and 2) that it is a historical term which was replaced by brottsbalk (as in element 3) in 1962

Since a bi-directional use of the dictionary is foreseen, the meta-language ought to be changeable in the final interface so that each of the user groups could have the information in their mother tongue (cf. section 7).

6.7. Grammatical information

At the initial editing stage there will be no grammatical information entered in the database for the TLs. The different TL projects will seek separate solutions for obtaining inflectional and other grammatical information corresponding to the resources locally available for the different languages. The aspiration is to provide the information by automatically connecting to existing databases for the different languages, but at the moment of writing the matter has not yet been finally settled.

7. User interface and perspectives

ISLEX is a work in progress and its user interface is presently only in the planning phase. The examples shown in this article have all been extracted from the editors’ view of the database, and a database interface is of course different from the final web interface designed for the general public. As yet, detailed specification requirements for the final design remain to be developed, but the task will be addressed in the near future. It goes without saying that some of the more basic features of online reference works will be implemented, such as search with wildcards and advanced search for the contents of microstructural information types other than the lemma itself (cf. 7.1), and it was suggested earlier that a bi-directional use of the dictionary could be rendered possible by introducing specific profiles for users of different mother tongues. At the moment of writing, the question how and to what extent different user profiles will be implemented is not fully settled, but an important consequence of such a decision would be that all metatexts relating to the relevant elements should be available in both the SL and all the TLs. Another consequence is that some information types would be visible while others would be suppressed for a given user profile. For example, grammatical information about the SL is crucial for a Scandinavian student of Icelandic, but less relevant for a native speaker of Icelandic – and vice versa for the Scandinavian equivalents.

As the translation work continues in Denmark, Norway and Sweden, a number of TL equivalents are added to the database. It is tempting to offer a look-up function for words in the TL equivalents from the user interface, thereby turning the Scandinavian languages into SLs and Icelandic into the TL. Technically, it is no problem, and editorially the function has been foreseen: equivalents are entered in a specific element for this purpose—and only one equivalent is entered in each element—named differently from the element used for a

paraphrase or an explanation. However, one must bear in mind that the dictionary was never designed to operate in the opposite direction, and within the framework of the current project, resources do not suffice to allow human checking. So using the database in reverse can by no means guarantee correspondence between the languages. Despite this, it will undoubtedly be a much sought-after feature of ISLEX which will have to be considered.

7.1. Searching across information elements

In a structured database such as ISLEX, any information element may be queried, and should be offered as an option. Two examples may illustrate the point.

By searching across the elements “examples” and “phrases” the user will find a richer example material than is presented in any single entry. The search facility may thus prove useful as a guide to correct or idiomatic language usage. This is particularly helpful for solving language production problems.

A combined search for an adjective plus its noun marker (cf. section 4) is another useful option for language production purposes. For example, if the user looks up “part of speech = adjective” plus “noun marker = *planta*” (“plant”), it would yield the following result: *einær* (“annual”), *ffjölaer* (“perennial”), *jurtkennndur* (“herbacious”), *sígrænn* (“evergreen”), *sjálfsáinn* (“self-seeded”). The result is retrieved from different entries, but put together it reveals useful information about the vocabulary used for plant words and possibly also about collocational characteristics.

Again however, the same warning applies as was mentioned for the reversed use: it is a feature which was not deliberately planned and should, if implemented, be used with corresponding caution.

7.2. Access to phrases as independent elements

Phraseological material is marked as a special information type in the database and as such multi-word units are fully searchable.

There are plans to store the phrases independently of the rest of the dictionary and give access to them as a separate dictionary of Icelandic phrases. In order to enable effective search possibilities, a system of keywords will be implemented, similar to the feature provided by the online dictionary <http://idioms.thefreedictionary.com>. If the user wants to look up the idiom *it's raining cats and dogs* the keywords to type in the search box would be *rain*, *cat* or *dog*. This can be done fairly easily as the keyword function is already built into the database.

8. Concluding remarks

Although ISLEX is only just half-way through the project period, we are often reminded how much an Icelandic-Scandinavian dictionary is needed through repeated inquiries from impatient users. This is one reason why an early launch is contemplated even if it implies that only a part of the entries would be complete and perhaps some information types still lacking (such as pronunciation and pictures).

Although no final decision has been taken on the issue, we are encouraged by the feedback that we get. We are also grateful for the financial support which the project has received. Not only is the project supported by the governments of all four countries, which in itself is rather unusual, but it has also received donations from the Nordic Language Programmes (Nordplus Sprog) and a private company (Baugur Group). We realize that sponsoring, whether public or private, is becoming increasingly necessary for the development of lexicographical quality products. Only in that way will it be possible for lesser spoken languages to hold their own and meet the challenges of a globalized world.

References

- Hannesdóttir, A.; Jónsson, J. H. (2001). “Að hafa í sig og á—Ísländsk fraseologi i ett isländskt-svenskt perspektiv”. *LexicoNordica* 8. Oslo: Nordisk forening for leksikografi i samarbeid med Nordisk Språkråd. 67-91. [“To put in and put on (i.e. ”to have food and clothes”)—Icelandic phraseology in an Icelandic-Swedish perspective”. *LexicoNordica* 8. Oslo: Nordic Association of Lexicography in collaboration with the Nordic Language Council. 67-91].
- Jónsson, J. H. (2005). *Stóra orðabókin um íslenska málnotkun*. Reykjavík: JPV útgáfa. [Large Dictionary of Icelandic Language Usage. Reykjavík; JPV Edition].
- Kernerman, I. (1994). *Kernerman Dictionary News* 1. Also available online: <http://kictionaries.com/newsletter/kdn1-4.html>.
- Martin, W. (ed.) (2007). *International Journal of Lexicography* 20 (3). Oxford: Oxford University Press.
- Svensén, B. (2004). *Handbok i lexikografi. Ordböcker och ordboksarbete i teori och praktik*. Stockholm: Norstedts Akademiska Förlag. [“Practical Lexicography: Dictionaries and Dictionary-Making in Theory and Practice”. Stockholm: Nordstedt’s Academic Publishers].

Electronic resources

- Bjarnadóttir, K. *Beygingarlýsing íslensks nútímamáls* [“Inflectional database of contemporary Icelandic”] [on-line]. <http://www.lexis.hi.is/beygingarlýsing>. [Access date: 30 March 2008].
- LEXIN dictionaries online* [on-line]. Stockholm: The Language Council of Sweden. <http://lexin.nada.kth.se> [Access date: 30 March 2008].
- TheFreeDicctionary [on-line]. Huntingdon Valley: Farlec, Inc. <http://idioms.thefreedictionary.com> [Access date: 30 March 2008].