NORDISKE STUDIER I LEKSIKOGRAFI

Titel: A Concordance to Old Icelandic Texts and its Lexicographic Value

Forfatter: Eiríkur Røgnvaldsson

Kilde: Nordiske Studier i Leksikografi 3, 1995, s. 123-135
Rapport fra Konferanse om leksikografi i Norden, Reykjavik 7.-10. juni 1995

URL: http://ojs.statsbiblioteket.dk/index.php/nsil/issue/archive

© Nordisk forening for leksikografi

Betingelser for brug af denne artikel
Denne artikel er omfattet af ophavsretsloven, og der må citeres fra den. Følgende betingelser skal dog være opfyldt:

- Citatet skal være i overensstemmelse med „god skik“
- Der må kun citeres „i det omfang, som betinges af formålet“
- Ophavsmanden til teksten skal krediteres, og kilden skal angives, jf. ovenstående bibliografiske oplysninger.

Søgbarhed
Artiklerne i de ældre Nordiske studier i leksikografi (1-5) er skannet og OCR-behandlet. OCR står for 'optical character recognition’ og kan ved tegngenkendelse konvertere et billede til tekst. Dermed kan man søge i teksten. Imidlertid kan der opstå fejl i tegngenkendelsen, og når man søger på fx navne, skal man være forberedt på at søgningen ikke er 100 % pålidelig.
A Concordance to Old Icelandic Texts and its Lexicographic Value


The main subject of my paper will be a new concordance to the Íslendinga sögur (Icelandic Family Sagas), which will be published on CD-ROM later this year. In the first part of the paper, I will describe the concordance, but in the second part, I will consider its potential use in dictionary making.

1 The concordance

The concordance to the Family Sagas (Eiríkur Rúgvaldsson et al. 1995) is one of the first concordances to be published in Iceland. The very first computerized concordance to an Icelandic text is the one which professor Baldur Jónsson and his collaborators made to the novel *Hreîðrîð* by Ólafur Jóhann Sigurðsson. This concordance, which was published in a limited number of copies in 1978 (Baldur Jónsson 1978), differs however from the present one in various respects, the most important difference being that it is not lemmatized.

The first lemmatized concordance to a text in Icelandic appeared just before last Christmas. This is the concordance to the latest edition of the Bible, which was made by a group of specialists from different institutions (*Bibliulykill* 1994). This work is in many ways comparable to ours, but there are, however, several important differences. First, many of the most frequent words are omitted; for instance, all prepositions, conjunctions, and several adverbs, and also a few frequent verbs and nouns. All such words are included in our concordance. Second, the ordering of the occurrences of each word is different. In the concordance to the Bible, the examples are ordered according to the order in which they appear in the Bible. In our concordance, on the other hand, the examples are alphabetically ordered according to the following word.

A group of scholars started working on the concordance to the Icelandic Sagas in 1989. This group consists of Bergljót Kristjánsdóttir, Guðrún Ingólfsdóttir, Órnólfur Thorsson, and myself, but several others have also worked more or less on the project, which has been generously supported by the Icelandic Science Fund. It is based on a new edition of the Sagas, which appeared in 1985 and 1986 (*Íslendinga sögur* 1985–86). Some of the editors of that edition are also among the leaders of the present project, which can thus be seen as a continuation of the edition.
The main work on the concordance was done in 1989, and in November that year, the lemmatization was almost finished, so that preliminary results of some frequency studies on the vocabulary of the Sagas could be presented at a conference in Reykjavík; these results have been published in the journal Skáldskaparmál (Eiríkur Rögnvaldsson 1990). At that time, the grants that had been given to the project had been used up, but the project itself, however, was far from finished. The lemmatization had to be carefully checked and proof-read, the computer files had to be corrected, etc. But due to lack of money, the editors of the concordance have only been able to work on it in their spare time the last five years.

This does not mean, however, that the concordance has been inaccessible up to now. Since 1992, when the work was practically finished, it has been preserved at the Institute of Linguistics at the University of Iceland, both on a computer and in a laser print-out. Everybody has had unlimited access to both versions. In addition, the editors have answered numerous questions from all over the world, concerning words and phrases in the Sagas. However, with respect to the usefulness of the concordance, it has of course been a major drawback that it is not publicly available.

We have recently made a contract with the publishing house Mál og menning, which holds the copyright to the editions on which the concordance is based. Later this year, they are planning to publish a CD, which will include the concordance and also a text version of the Sagas. Both will be easily searchable by means of special Windows-based programs. There will be links between the text and the concordance, so that it will be possible to click on a certain word in the text and get all the examples of that word on the screen; or to click on a word in the concordance and get the surrounding text on the screen. The CD will also include several lists, such as a frequency list, a list of compounds, etc.

1.1 The making of the concordance

The concordance comprises all the texts in the edition on which it is based, except the þættir; the poetry is also omitted. The Sagas are around 40, but some of them exist in two widely different versions, so that 50 different texts are printed in the edition. This is around 5 megabytes of text, or nearly 900 000 running words; 2 079 pages.

We started by inserting special markers for each Saga, chapter numbers and page breaks. Then we could use WordCruncher, from Johnston & Co. in the United States, to generate a list of all the occurrences of each individual word-form. In this list, we have the word-form in question in the middle, with approximately 40 characters context in each direction, and references to Saga, chapter and page in the beginning of each line. At this stage, the file looks like the picture in (1a).

The next step is to prepare this file for lemmatization. We used WordPerfect macros to boldface the word in the middle, and to sort the examples of each word-form alphabetically, according to the following context. After that, the file looks as in (1b).

Up to this point, the process has been reasonably mechanic, but now comes the difficult part: the lemmatization itself, where we group together all the forms belonging to each individual lexeme, and make a distinction between all homonyms belonging to different lexemes. We considered using computer programs to make this easier, and we actually tried one such program, but we soon found out that its benefits did not compensate for the errors it made. So, the lemmatization had to be done manually, which was quite a task, considering
the size of the corpus. When it was finished, the concordance files were printed on a laser printer, giving the result shown in (1c).

(1a)

<table>
<thead>
<tr>
<th>Lexd</th>
<th>64:1634</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stðu út af asendamir og var einart pak</td>
<td>á húsinnu og ekki gróð. Þá mæli</td>
</tr>
</tbody>
</table>

(1b)

<table>
<thead>
<tr>
<th>Lexd</th>
<th>64:1634</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pakk</td>
<td>64:1634</td>
</tr>
<tr>
<td>Stðu út af asendamir og var einart pak</td>
<td>á húsinnu og ekki gróð. Þá mæli</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Pakk</th>
<th>64:1634</th>
</tr>
</thead>
<tbody>
<tr>
<td>Pakk</td>
<td>64:1634</td>
</tr>
<tr>
<td>og hvar að öðrum þar til er alt að pakk</td>
<td>af húsinnu. Vantafli týgði hér og svo</td>
</tr>
</tbody>
</table>

After that, the concordance was proof-read and the lemmatization rechecked. During that process, all available dictionaries were consulted, especially Fritzner’s Ordbog over det gamle norske sprog (Fritzner 1954), of course, but also Ásgeir Blöndal Magnussson’s etymological dictionary Íslensk orðsforbók (Ásgeir Blöndal Magnusson 1989), and several other works. This was a very time-consuming process, as one can imagine given the fact that the paper version of the concordance is more than 7000 pages with 100 lines on each page and ca. 100 characters per line. Since Old Icelandic is a highly inflected language, homonyms of different lexemes are very frequent, and therefore, it was necessary to read most of these lines carefully, because it is very often possible that a rare inflectional form
of a verb, for instance, is homonymous with a form that one would a priori think that could only be a noun.

(1)c

The final step was to make the necessary corrections to the computer files. As I said before, this was practically finished in 1992, even though individual corrections are still being made. Users of the concordance have sometimes noticed errors and inconsistencies which they have told us about. I can particularly mention Pórdís Úlfarsdóttir, who went carefully through the concordance in connection with Jón Hilmar Jónsson’s work on his book Orðastadur, which was published last year (Jón Hilmar Jónsson 1994). Pórdís gave us a list of errors that she had found, and we are very grateful to her and others who have assisted us in eliminating errors as far as possible.

1.2 Vocabulary and word frequency

The concordance has already proved to be very useful in itself. Let me first mention its use as a frequency dictionary. For the first time, we now have an overview of the vocabulary of a whole literary genre; the Icelandic Family Sagas. Of course, there exist dictionaries of Old Icelandic, especially Fritzner’s (1954) Ordbog over det gamle norske sprog; and as is well known, the Arnamagnæan Commission in Copenhagen has been working on a dictionary of Old Norse for several decades. These works, however, comprise not only narrative texts like the Sagas; they also cover other genres such as the law, lives of saints, etc. The vocabulary of these genres is remarkably different from that of the Sagas.

Now we know that the vocabulary of the Sagas is somewhere between 12 000 and 12 500 words — the exact figure depends on our definition of lexeme, and besides, differences between manuscripts can of course affect the figure. We can also find out the vocabulary of each individual Saga. Njöls saga, for instance, uses around 3 200 different words. It appears that the Sagas use unusually few words. Unfortunately, however, we cannot show this statistically, since there exist no comparable studies of Modern Icelandic texts — except for the Bible, which is hardly representative of Modern Icelandic.
The Institute of Lexicography has recently published a frequency dictionary of Modern Icelandic, *Íslensk orðtögníbók* (Jørgen Pind et al. 1991). It is possible to compare several figures from this work to the results of our study of Old Icelandic. This is done in (2) below.

(2)  

<table>
<thead>
<tr>
<th>Lexemes</th>
<th>Running words</th>
<th>%</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nouns</td>
<td>7,292</td>
<td>117,252</td>
<td>15.63</td>
</tr>
<tr>
<td>Verbs</td>
<td>1,447</td>
<td>203,148</td>
<td>27.09</td>
</tr>
<tr>
<td>Adjectives</td>
<td>2,851</td>
<td>31,947</td>
<td>4.26</td>
</tr>
<tr>
<td>Adverbs</td>
<td>706</td>
<td>173,484</td>
<td>23.13</td>
</tr>
<tr>
<td>Pronouns</td>
<td>52</td>
<td>94,800</td>
<td>12.64</td>
</tr>
<tr>
<td>Conjunctions</td>
<td>20</td>
<td>113,823</td>
<td>15.18</td>
</tr>
<tr>
<td>Numerals</td>
<td>33</td>
<td>6,292</td>
<td>0.84</td>
</tr>
</tbody>
</table>

The first column shows how many lexemes in the Sagas belong to each part of speech. As you see, the nouns make up almost 60% of the vocabulary. I must point out that adverbs and prepositions are grouped together. This is done to facilitate the comparison with the results from *Íslensk orðtögníbók*, and besides, it is often very difficult or even impossible to draw a line between these two parts of speech.

In all the other columns, the figures refer to running words but not to lexemes. In the second column, we see that the relative frequency of running words belonging to each part of speech is widely different from the relative frequency of lemmas. The last two columns show percentages; the first of them shows the percentage of running words in each part of speech in the Sagas, whereas the second shows comparable figures from *Íslensk orðtögníbók*.

As you see, the figures are rather similar. There is, admittedly, a considerable difference in the relative frequency of nouns. The reason is that we have omitted all proper names from our figures for nouns in the Sagas. It must be noted that proper names are no doubt much more common in the Sagas than they are in the texts on which *Íslensk orðtögníbók* is based. If we had chosen to include proper names in our figures, the relative frequency of nouns would have been higher in the Sagas than in *Íslensk orðtögníbók*.

It must also be noted that we have chosen to count all instances of participles, both past and present, as verb forms; the only exception being present participles used as nouns, such as *eigandi*. The obvious alternative would have been to classify the participles as either verbs or adjectives according to their syntactic status in each case, as is done in *Íslensk orðtögníbók*. We actually tried this in the beginning, but we soon came to the conclusion that it was impossible to make a principled decision in all cases, and the only consistent solution would be to count all participles as verbs. This decision, of course, results in relatively more occurrences of verbs and fewer occurrences of adjectives than it would have done if we had followed the same principles as the authors of *Íslensk orðtögníbók*; but if we take this difference into account, I think we can say that the figures in the last two columns are very similar.
1.3 Other uses

We are pleased to say that the concordance has already been used and quoted in numerous publications in different disciplines, such as medieval literature, historical syntax, history, folklore, ethnography, law, zoology, physics, etc. In our view, one of the most important features of the project is its interdisciplinary character. It brings together scholars from various fields of study, who are working on some aspects of Medieval Iceland. They can use the concordance to locate places of interest in the Sagas, and thus, they can get a unique overview of their subject. Thus, the concordance has already inspired several studies, and the insights these scholars get by using the concordance will in turn be of tremendous use in the semantic description of numerous words in the Sagas.

2 How the material affects the structure of the dictionary

My second main subject in this talk is the use of this kind of material, i.e. a concordance, in dictionary making. In what way does it affect the final form of the lemmas in a traditional dictionary if the material is a concordance, but not accumulated by a traditional excerption? The effects are numerous and of various kinds, but the most important are those listed under (3):

(3) a. Frequency information facilitates the selection of citation forms
   b. The semantic description of very common words will be more accurate; different senses of a word can be more easily ordered by importance, and various subtle semantic differences can be more easily detected
   c. Formal categorization will be more prominent, and syntactic features (such as case government) are listed more systematically
   d. The selection of text examples (citations) will be more accurate, and the examples will be more typical
   e. All kinds of collocations and word patterns are more obvious, and therefore more likely to be mentioned

In the following, I will discuss each of these effects in turn.

2.1 Frequency

As is well known, it is not always necessary nor feasible to list every word that occurs in a given corpus as a separate dictionary entry with its own description. On the contrary, there are numerous cases where two or more words which differ somewhat in form should rather be considered as belonging to the same lexeme, and listed under one citation form. Such examples can be of various types, and some of them are shown in (4) below.
(4) a. hofgoði – hofgoði  
   atgervimaður – atgervismaður  
   höfuðbani – höfuðsbani  
   hugboð – hugarboð

b. aðdráttamaður – aðdráttarmaður  
   affaradagur – affarardagur

c. drukknan – drukknun, geipan – geipun  
   auðigur – auðugur, ástúðigur – ástúðugur  
   hraustleikur – hraustleiki, hvatleikur – hvatleiki

d. atgervi – atgjörvi  
   gagnvert – gagnvart

e. heyrinkunnigur – heyrumkunnigur  
   hlælegur – hlæglegur – hlægilegur

In Icelandic, such formal differences are often due to different ways of compounding. In the Icelandic Sagas, both hofgoði and hofgoði are found, as shown in (4a). We can explain this difference by saying that in the former, the first constituent of the compound is the stem, whereas in the latter, the first constituent is the gen.sg. form. However, there is little doubt that these two should be considered as belonging to the same lexeme.

We also find a number of compounds where the first constituent sometimes has the gen.sg. form, but in other cases the gen.pl. form. This is most frequent in words where the first part has a gen.sg. ending in -ar; then the only difference between the gen.sg. and the gen.pl., which always ends in -a, is the -r. Since the number (sg. or pl.) of the first constituent in such compounds is (usually) not semantically distinctive, and since the -r is often not clearly pronounced, such vacillation in number is common in Modern Icelandic; and many similar examples can also be found in the Sagas, such as aðdráttamaður and aðdráttamaður, which are shown under (4b) above.

There are also various examples of suffixes having more than one form in the Sagas; for instance -an/-un, -igur/-ugur, leiður/-leiði and others, in words like geipan/geipun, auðigur-auðugur, hraustleikur-hraustleiki, as shown in (4c) above. We also find words with and without breaking, such as atgjörvi and atgervi, as shown under (4d); and various other types, cf. (4e).

In cases like these, the lexicographer is often faced with several problems. It is often difficult to decide whether to group two or more different forms under one headword. Even if it can be shown that two different forms stem from the same lexeme historically, it is by no means evident that they should be given a single lexical entry in the dictionary. It is perfectly possible that each form has developed a special meaning which makes it natural to list both forms separately.

If we decide to group the different forms together in a single dictionary entry, it is often difficult to select the headword. The most straightforward solution would perhaps be to select the most frequent form as the citation form, but it is often not easy to find out which of the forms is most frequent. It would for instance not be wise to base the choice on the number of examples that have been excerpted from texts.
A concordance can be of a great help in solving these problems. Since the concordance contains all the occurrences of every single word in a given corpus, it is easy to find the frequency of any particular form. This gives the lexicographer a more solid ground, on which to build the selection of a citation form. However, it is clear that frequency is not the only factor to consider in this respect; the selection must also fit into the system, so to speak.

A concordance also makes it easier to decide whether two different, but related forms actually mean the same, and hence should be listed under the same dictionary entry. By the careful examination of all the examples that the concordance makes possible, one can sometimes detect subtle semantic differences that would otherwise not be noticed.

2.2 Meaning

It is a well-known tendency for traditional excerption to give a somewhat skewed picture of the meaning or use of individual words. Lexicographers tend to pick up unusual examples, and hence, such examples often get a more prominent status in the dictionary than they deserve. When a dictionary is based on a concordance, this problem can be avoided, because in principle, at least, the description is based on all the examples found in the corpus. Therefore, the most frequent meaning and use should get prominent status in the description.

It is very important in this respect that the lexicographer who writes the final description for the published dictionary has access to all stages of the material. When a lexicographer is writing a dictionary entry on the basis of examples that have been collected in a traditional excerption, he is completely dependent on his examples. Of course, he can, in principle, look up the citations in the excerpted texts, but in practice, it is impossible to do so, except in a limited number of cases. Therefore, the lexicographer does not know how typical his data are.

I'll just show you one example. Fritzner (1954) gives the following semantic definition of the word *heimsókn*:

(5)  
*heimsókn:*  
1. Besøg  
2. Besøg som man aflægger i retsag Øiemed, for at fremme en Retssag o. desl.  
3. voldeligt Overfald paa en i hans hjem, hjemmøgelse hvorunder man bruger Magten mod dem som ere i Huset

In our corpus, we find the following examples of this word:

(6)  
<table>
<thead>
<tr>
<th>heimsókn lev.</th>
<th>heimsókn (5); heimsóknar (1); heimsóknina (1); heimsóknir (1)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Laxd</td>
<td>63:1632 og eigi munda og veita honum síka helmsókn.</td>
</tr>
<tr>
<td>GullI</td>
<td>17:1138 so víg þeirra Hálga skyltú út standari helmsókn.</td>
</tr>
<tr>
<td>Vopaf</td>
<td>17:2003 og bragt fara til Hofs og veita Bjaran</td>
</tr>
<tr>
<td>HallM</td>
<td>10:1218 komi fyrir viga Galls og þar með helmsókn.</td>
</tr>
<tr>
<td>HallO</td>
<td>12:1249 komi víg Eimur fórlansnar og þar með helmsókn.</td>
</tr>
<tr>
<td>Svefr</td>
<td>15:1800 og anfr hann þa málinu tilu til helmsókn.</td>
</tr>
<tr>
<td>ReykI</td>
<td>14:1755 vígumum en víg Nafís stúl koma fyrir helmsóknina.</td>
</tr>
<tr>
<td>Veita</td>
<td>29:1878 Már kvað þa sýna af sér døvingarnlegar helmsóknir.</td>
</tr>
</tbody>
</table>

From these examples, it looks as if the most frequent meaning is the one under 3. in (5) above. Admittedly, Fritzner bases his description on many more texts, but however, there are
reasons to believe that a close consideration of all the examples would change the structure of the lexical entry.

When we started preparing the concordance, we were planning to exclude most function words; conjunctions and prepositions, and also many or most adverbs and pronouns. We did not think that examples of these words would be of any interest, since there are many examples of some of them on each page of the text. But when the real work started, we soon found out that a concordance could tell us many things about these words.

It is evident that only a limited number of examples of these words make the basis of their description in a traditional dictionary. Here we can, in principle, base our description on all the examples, and hence, we should be able to present a much more coherent description, both formally and semantically.

2.3 Syntactic characteristics and formal classification

In linguistic definitions of the lexicon, we usually read that this is the place where information on all unpredictable features of individual words is stored. This includes phonetic and phonological features (pronunciation), inflection, syntactic features, and meaning. This definition of course applies to the mental lexicon, but not to dictionaries, but by and large, I think we can say that these features are also the ones we can expect to find in a good dictionary.

Traditional dictionaries usually do justice to three of the above-mentioned fields. The phonological features can often be deduced from the spelling, and of course, many dictionaries show phonetic transcription. Inflection is usually shown by mentioning inflectional class, showing the principal parts of verbs, etc. The main part of the entry is, then, the semantic description.

Syntactic features, however, are usually not systematically represented. It is of course shown to which part of speech each lexical entry belongs; but features such as the case government and argument structure of verbs, for instance, are usually not mentioned. True, we can often see from the citations whether some verb takes one or two objects, or whether it governs accusative, dative, or genitive; but the point is that information on this is not systematically present, and it is sometimes lacking. One of the reasons for this is probably that the excerption of texts is not done with syntactic characteristics in mind, and therefore, there is simply no basis for including such things in the dictionary.

Here we have, once again, one of the problems with traditional excerption. Lexicographers have the tendency to pick up unusual or exceptional examples. This is fine, of course; but the danger is that such examples will be overrepresented in the material, at the expense of the normal use of words. If we find, for instance, one example where a certain verb governs a different case than it usually does, we are likely to pick up this example; and later, it might end up in a published dictionary, perhaps as the only text example which shows the case government of this verb.

By using a concordance, such dangers can be avoided. Since we have direct access to all the examples of each individual word in the corpus, we can simply count how often each verb takes each case, and make that information a part of the lexical entry, either directly or indirectly.
When the structure of a dictionary is based on a concordance of the kind we have made, it is bound to affect the final form of the lemmas in various ways. The main effect is probably that syntactic characteristics will be more prominent than they would otherwise, but semantic characteristics will tend to be less prominent. However, it must be emphasized that syntactic and semantic characteristics often go together, of course.

It is likely that in a traditional excerption, the meaning will be the dominant factor. The lexicographer tends to pick up those examples that exemplify the meaning of the word in question; but he will be less likely to select his examples according to their syntactic status. In many ways, it is more straightforward to let formal characteristics govern the structure of the lemma than to let the semantics do the job. One reason is that the formal classification is usually rather clear-cut; the syntactic status of the word in question is normally reasonably clear, so that the formal classification is not problematic. Semantic classification often presents much more difficult problems, and the lexicographer will have to rely on his intuitions to a much greater extent.

I can mention here that in the Sýnilehti sagnordabókar (Ásta Svaavardsóttir et al. 1993), which Orðabók Háskólaun published two years ago, formal classification is dominant, but semantic classification subordinate. I think this booklet shows well the merits of that structure. However, it must be kept in mind that this work is not based on a concordance, but rather on material from a traditional excerption of texts; and as I said above, this might mean that certain syntactic constructions are not justly represented.

2.4 Selection of text examples

It is very important that the text examples in a dictionary are carefully chosen. The appropriate examples can shed a new light on the meaning of a word, and be more illuminating than a long and tedious definition or explanation. In a dictionary which is based on material from a traditional excerption, we can always expect the selection of examples to be more or less arbitrary. The examples in the material can have been collected for various reasons; they may be of interest semantically, syntactically, or morphologically, for instance, but that does not mean that they are typical of the use of the word in question.

2.5 Collocations

It is by no means obvious in what order the examples of each word form should appear in a concordance. We decided to order the inflectional form of each lexeme alphabetically, as shown in (7) below. There we have first all the examples of the form heimil, then all the examples of the form heimília, and so on. If you look at the examples of each form, you see at once that they are alphabetically ordered according to the following word or words.

It must be admitted that the decision to choose this particular order was not built on much considerations, but nevertheless, we think that this decision has proved to be correct. The reason is that this ordering reveals how common it is that the same string of words occurs many times in the corpus. The reasons for such recurrent patterns can of course vary. In some cases, it is fairly clear that one author is imitating another, and even though that can be of a great interests to philologists, such information should hardly enter the dictionary.
A Concordance to Old Icelandic Texts and its Lexicographic Value

Often, however, it is evident that some word pattern or collocation is at stake, and that kind of information should be a part of the dictionary. A few examples of such patterns are shown in (8) and (9).

(8) alldjarflega ao alldjaflega (8)

(9) alldrengilega ao alldrengilega (7); alldrengilega (1)

In (8) we see that the advverb alldjarflega and the verb berjast almost always go together; and (9) shows that the same goes for the advverb alldrengilega and the verb verjast. This is not mentioned in any existing dictionary of Old Icelandic or Old Norse, as far as I know, and
it would probably not be fair to claim that it should. On the contrary; I think that we must have access to a concordance to see this. However, there can be no doubt that this is not a coincidence, and information on this should be found in a dictionary. It may be noted in this connection that the concordance to the Sagas has already been used in a published dictionary; this is Jón Hilmar Jónsson’s (1994) Orðastáður, which is a dictionary of collocations.

Examples of this kind are numerous in the concordance. It is true, of course, that one can sometimes infer something of this kind in the citations in the published dictionaries. The trouble is, however, that it is difficult to know what these examples really show; how typical the connection that the concordance to the

---

In (10) we see another similar example. The word under consideration is the conjunction **uns**.

(10)

| uns st; uns (36) | Fjöt 18:709 | Arnleif Hrafnson | fólk og búað rað þar uns | eg kem á morgun, nema eða finni eiga | eg kem milli tils til ráða lagi.
|               | Írð 9:2032 | og rúðar undan sem mest getur hann uns | hans kemur skammt frá þeim er | hans kemur í Atlaflið semm gormun.
|               | Eirík 4:524 | sagt. Írðar þeir fór slafið og fer uns | hans kom til Þjóðsaga semm gormun | hans kom til Þjóðsaga semm gormun.
|               | BækiK 4:32 | það fóra oddur med hinn tattunganda maður uns | hans kemur til Þjóðsaga semm gormun | hans kemur til Þjóðsaga semm gormun.
|               | Þóvr 6:2056 | yfir Lagarfljótt og upp með fjóðinu uns | hans kemur til Þjóðsaga semm gormun | hans kemur til Þjóðsaga semm gormun.
|               | Fjöt 21:720 | mývæna og fara svo vestur til svoða uns | hans kemur til Þjóðsaga semm gormun | hans kemur til Þjóðsaga semm gormun.
|               | Vatn 40:1891 | taka við honum og fór hann um svoði uns | hans kemur til Þjóðsaga semm gormun | hans kemur til Þjóðsaga semm gormun.
|               | Glís 24:879 | fóra þann í Geirfjóðsflókið og er þar uns | hans kemur til Þjóðsaga semm gormun | hans kemur til Þjóðsaga semm gormun.
|               | Grett 30:1001 | sannan mánna. þeir voru fimm saman, þyða uns | hans kemur til Þjóðsaga semm gormun | hans kemur til Þjóðsaga semm gormun.
|               | Njóla 6:132 | rúða bjað sig skjótt og söða rúða þar uns | hans kemur til Þjóðsaga semm gormun | hans kemur til Þjóðsaga semm gormun.
|               | Grett 90:1092 | með göðu þramsveyti og fóru alla leið uns | hans kemur til Þjóðsaga semm gormun | hans kemur til Þjóðsaga semm gormun.
|               | Fjöt 23:723 | þeir rúða nö til efir hleði uns | hans kemur til Þjóðsaga semm gormun | hans kemur til Þjóðsaga semm gormun.
|               | Hrafn 14:1414 | undan. Rúða þeir Sámar þar all að ekki uns | hans kemur til Þjóðsaga semm gormun | hans kemur til Þjóðsaga semm gormun.
|               | Fjöt 13:697 | hvertímgjöf með þeim vestur landsins uns | hans kemur til Þjóðsaga semm gormun | hans kemur til Þjóðsaga semm gormun.
|               | Glís 7:857 | þeir færa dag þann og um söttina uns | hans kemur til Þjóðsaga semm gormun | hans kemur til Þjóðsaga semm gormun.
|               | BækiK 30:114 | þar er upp liggja í dalinn frá Völlum uns | hans kemur til Þjóðsaga semm gormun | hans kemur til Þjóðsaga semm gormun.
|               | Korm 7:1476 | þetta rúð taka þeir, færa uns | hans kemur til Þjóðsaga semm gormun | hans kemur til Þjóðsaga semm gormun.
|               | Vagam 18:2004 | Fara þeir nö sem þeir mega mest uns | hans kemur til Þjóðsaga semm gormun | hans kemur til Þjóðsaga semm gormun.
|               | Fjöt 18:708 | til þis og upp f heiti til Ógosprekkars uns | hans kemur til Þjóðsaga semm gormun | hans kemur til Þjóðsaga semm gormun.
|               | Harð 36:1290 | þöf á að skipta þádiv og fára nö | hans kemur til Þjóðsaga semm gormun | hans kemur til Þjóðsaga semm gormun.
|               | Glís 11:852 | Hann lítur þeim þanstana og rúða þeir uns | hans kemur til Þjóðsaga semm gormun | hans kemur til Þjóðsaga semm gormun.
|               | Fjöt 25:727 | þu ofan á dani og upp efir nesið uns | hans kemur til Þjóðsaga semm gormun | hans kemur til Þjóðsaga semm gormun.
|               | Njóla 131:287 | í brætt og fyrir nöðan þökull og svo uns | hans kemur til Þjóðsaga semm gormun | hans kemur til Þjóðsaga semm gormun.
|               | Heið 22:1373 | og stjóðu hans á baki og rúðu svo uns | hans kemur til Þjóðsaga semm gormun | hans kemur til Þjóðsaga semm gormun.
|               | Grett 19:985 | ofan Indriðaþátt fjáð Pyrrh og þeir þar uns | hans kemur til Þjóðsaga semm gormun | hans kemur til Þjóðsaga semm gormun.
|               | Harð 38:1292 | ofan Indriðaþátt fjáð Pyrrh og þeir þar uns | hans kemur til Þjóðsaga semm gormun | hans kemur til Þjóðsaga semm gormun.

When we look at the examples we see that in a great majority of them, or 31 out of 36, the verb *koma* follows *uns*. Note that *uns* is a temporal conjunction, and it is impossible to deduce from its meaning that it has closer ties to *koma* than to any other verb. Another temporal conjunction in Old Icelandic, *par til*, for instance, does not have any comparable ties to any particular verb.

I could add hundreds of examples similar to those that I have mentioned. In some of the cases, information on word combinations or ties between words clearly should be found in a dictionary; in other cases, this may not be so clear. The point is, however, that the concordance gives us a unique overview of such patterns, and makes it possible to see things that simply could not be seen without such a tool.
3 Conclusion

In the first part of this paper, I described the making and the structure of a forthcoming concordance to the *Íslendinga sögur*, whereas in the second part, I talked about its potential use in dictionary making, as I see it. During the last few years, Guðrún Ingólfsdóttir, Bergljót Kristjánsdóttir and others have actually been using the concordance as a basis for a new dictionary of the Sagas. In another paper in this volume, Guðrún gives a short description of their work.

References


Baldur Jónsson 1978: Orðstöðulykill að Hreðrinu. [Fjölrit.]


