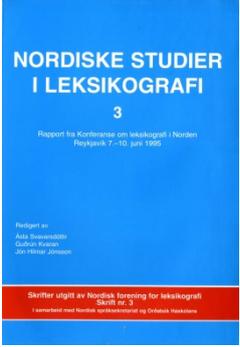


NORDISKE STUDIER I LEKSIKOGRAFI

Titel:	A Concordance to Old Icelandic Texts and its Lexicographic Value	
Forfatter:	Eiríkur Rögnvaldsson	
Kilde:	Nordiske Studier i Leksikografi 3, 1995, s. 123-135 Rapport fra Konferanse om leksikografi i Norden, Reykjavík 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

Den første del af dette foredrag er en beskrivelse af projektet *Konkordanse til de Islandske sagaer*, som udkommer på CD-ROM i slutningen af 1995. I den anden del af foredraget spekulerer forfatteren på den nytte som ordbogsredaktører kan have af en konkordanse som denne, og giver nogle konkrete eksempler som skal vise at den foreliggende konkordanse vil muliggøre en bedre og nøjagtigere ordbogsbeskrivelse af gammelislandsk, både i syntaktisk og semantisk henseende.

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 Rögnvaldsson 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 *Hreiðrið* 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, Örnólfur 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).

(1)a

			I*þak (1)	
Laxd	64:1634	stóðu út af ásendamir og var einart	I*þak	á húsinu og ekki gróíó. Þá mætti
			I*þakió (6)	
Njála	77:213	eigi fyrr en þeir höfðu undiö allt	I*þakió	af skálanum. Gunnar skýtur þá af
Egla	22:392	þurt og bræddur víðurinn en nærum	I*þakió	um ræriö, Þóróttur það menn sína
Gísl	13:864	húsin að drjúpa sem líklegt var er	I*þakió	tók að rofna. Gísl spratt upp
Gísl	18:917	þvo mikill að þegar tekur af skálanum	I*þakió	öörum mægin og hver að öörum þar til
Gísl	18:917	og hver að öörum þar til er allt er	I*þakió	af húsinu. Vatnfall fylgdi hér og svo
Vígl	2:1957	gulli í skuröina. Þetta herbergi var	I*þakió	blíyi og steint allt innan. Skiógaröur
			I*þakka (14)	
Njála	44:177	þá til nokkurs." "Allvel skal eg	I*þakka	þeim af þeir segja mér heim víg
Egla	16:384	öllu best varíó er eg hefí gert til	I*þakka	yövarra." Konungur segir að ekki
Finnb	31:654	jarl hefa gefiö sér, mág sinn. Þeir	I*þakka	honum haröla vel og riöa heim. Haföi
Fóstb	5:786	faríó haföi með þeim Vemundi. Þeir	I*þakka	henni sitt tillag er hún haföi þeim
Fóstb	5:786	Ög er þeir voru búinir til ferðar þá	I*þakka	þeir henni þarvist sína og allan
Fóstb	20:818	búinir þá fara þeir á konungs fund og	I*þakka	alla vingan þá er hann haföi þeim
Gísl	9:860	það var vant að þorkell var vanur að	I*þakka	bröður sínum verkiö en nú er hann
Grett	91:1093	ekki víst um aturkomu sína "Vil eg nú	I*þakka	yöur öllum," segir Þorsteinn, "Þversu
GunkV	11:1160	honum því að eg á honum lífgjöf	I*þakka."	Karl segir: "Eg vil fá að sjá
Halm	6:1205	sem konungur er og vil eg	I*þakka	hérivist mína." Þá kom það fram
Hallö	7:1235	þar nær sem konungurinn er. Vil eg	I*þakka	honum hirövist." Þá kom það fram
Laxd	40:1686	hann vildi. Þá mætti Kjartan:	I*þakka	viljum vér yöur konungur er þér gefiö
Svarf	8:1789	og riöu í braut. Nú	I*þakka	menn Þorsteini fyrir er þeir gefiö
Vígl	4:1911	þetta vil eg eigi veita." Þeir	I*þakka	honum vel og þótti þeim þó mikliö í
			I*þakkaö (1)	
Egla	65:469	voru. Var þetta verk honum allvel	I*þakkaö.	Þá kvaö Egill:

(1)b

þak (1)				
Laxd	64:1634	stóðu út af ásendamir og var einart	þak	á húsinu og ekki gróíó. Þá mætti
			þakiö (6)	
Gísl	18:917	og hver að öörum þar til er allt er	þakiö	af húsinu. Vatnfall fylgdi hér og svo
Njála	77:213	eigi fyrr en þeir höfðu undiö allt	þakiö	af skálanum. Gunnar skýtur þá af
Vígl	2:1957	gulli í skuröina. Þetta herbergi var	þakiö	blíyi og steint allt innan. Skiógaröur
Gísl	13:864	húsin að drjúpa sem líklegt var er	þakiö	tók að rofna. Gísl spratt upp
Egla	22:392	þurt og bræddur víðurinn en nærum	þakiö	um ræriö, Þóróttur það menn sína
Gísl	18:917	þvo mikill að þegar tekur af skálanum	þakiö	öörum mægin og hver að öörum þar til
			þakka (14)	
Fóstb	20:818	búinir þá fara þeir á konungs fund og	þakka	alla vingan þá er hann haföi þeim
Gísl	9:860	það var vant að þorkell var vanur að	þakka	bröður sínum verkiö en nú er hann
Fóstb	5:786	faríó haföi með þeim Vemundi. Þeir	þakka	henni sitt tillag er hún haföi þeim
Halm	6:1205	sem konungur er og vil eg	þakka	hérivist mína." Þá kom það fram
Finnb	31:654	jarl hefa gefiö sér, mág sinn. Þeir	þakka	honum haröla vel og riöa heim. Haföi
Hallö	7:1235	þar nær sem konungurinn er. Vil eg	þakka	honum hirövist." Þá kom það fram
Vígl	4:1911	þetta vil eg eigi veita." Þeir	þakka	honum vel og þótti þeim þó mikliö í
GunkV	11:1160	honum því að eg á honum lífgjöf	þakka."	Karl segir: "Eg vil fá að sjá
Svarf	8:1789	og riöu í braut. Nú	þakka	menn Þorsteini fyrir er þeir gefiö
Laxd	40:1686	hann vildi. Þá mætti Kjartan:	þakka	viljum vér yöur konungur er þér gefiö
Grett	91:1093	ekki víst um aturkomu sína "Vil eg nú	þakka	yöur öllum," segir Þorsteinn, "Þversu
Egla	16:384	öllu best varíó er eg hefí gert til	þakka	yövarra." Konungur segir að ekki
Njála	44:177	þá til nokkurs." "Allvel skal eg	þakka	þeim af þeir segja mér heim víg
Fóstb	5:786	Ög er þeir voru búinir til ferðar þá	þakka	þeir henni þarvist sína og allan
			þakkaö (1)	
Egla	65:469	voru. Var þetta verk honum allvel	þakkaö.	Þá kvaö Egill:

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 Magnússon's etymological dictionary *Íslensk orðsifjabók* (Ásgeir Blöndal Magnússon 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 7 000 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

þak hk; þak (1); þakið (4)			
Laxd 64;1634	stóðu út af ásendamir og var einart	þak	á húsinu og ekki gróið. Þá mælti
GíslL 18;917	og hver að öðrum þar til er allt er	þakið	af húsinu. Vatnfell fylgdi hér og svo
Njála 77;213	eigi fyrir en þeir höfðu undið allt	þakið	af skálanum. Gunnar skýtur þá af
GíslS 13;894	húsin að djúpa sem líklegt var er	þakið	tók að rofna. Gísli spratt upp
GíslL 18;917	svo mikill að þegar tekur af skálanum	þakið	öðrum megin og hver að öðrum þar til
þakka so; þakka (13); þakkað (1)			
Fóstaþ 20;818	búinir þá fara þeir á konungs fund og	þakka	alla vingan þá er hann hafði þeim
GíslS 9;880	það var vant að borkell var vanur að	þakka	bróður sínum verkið en nú er hann
Fóstaþ 5;786	ferið hafði með þeim Vemundi. Þeir	þakka	henni allt tillag er hún hafði þeim
HallM 6;1205	sem konungur er og vil eg	þakka	hérivist mína." Þá kom það fram
Finnb 31;654	jarl hafa gefið sér, mér sínu. Þeir	þakka	honum harða vel og ríða heim. Hafði
HallÓ 7;1235	þar nær sem konungurinn er. Vil eg	þakka	honum hriðvíst." Þá kom það fram
YggL 4;1911	þetta vil eg eigi veita." Þeir	þakka	honum vel og þótti þeim þó mikilið í
GunnKV 11;1160	honum því að eg á honum lífgjafi að	þakka."	Kari segir: "Eg vil fá að sjá
Svarf 8;1789	og ríðu í braut. Nú	þakka	mann Þorsteini fyrir er hann tókst
Laxd 40;1696	hann vildi. Þá mælti Kjartan:	þakka	villjum vér yður konungur er þér gefið
Grett 91;1093	ekki víst um ásturkomu sína "vil eg nú	þakka	yður öllum," segir Þorsteinn, "hveru
Njála 44;177	þá til nokkurs." "Allveg skal eg	þakka	þeim ef þeir segja mér heim víg
Fóstaþ 6;786	Og er þeir voru búinir til ferðar þá	þakka	þeir henni þarvist sína og allan
Egla 65;469	voru. Var þetta verk honum allivel	þakkað.	þá kvað Egill:
þakja so; þakið (2)			
Vígl 2;1957	gulli í skurðina. Þetta herbergi var	þakið	blíði og steint allt innan. Skíðgarður
Egla 22;392	þunt og bræddur víðurinn en nærum	þakið	um næfri. Þórdífur bað menn sína
þökk kv; þakka (1)			
Egla 16;384	öllu best varið er eg hefði gert til	þakka	yðvarra." Konungur segir að ekki

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 Þórdís Úlfarsdóttir, who went carefully through the concordance in connection with Jón Hilmar Jónsson's work on his book *Orðastaður*, which was published last year (Jón Hilmar Jónsson 1994). Þórdís 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 Fritznér'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. *Njáls 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 orðtíðnibók* (Jörgen 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)

	<i>Íslendingasögur</i>			<i>Íslensk orðtíðnibók</i>
	<i>Lexemes</i>	<i>Running words</i>	<i>%</i>	<i>%</i>
<i>Nouns</i>	7 292	117 252	15.63	20.58
<i>Verbs</i>	1 447	203 148	27.09	20.65
<i>Adjectives</i>	2 851	31 947	4.26	7.14
<i>Adverbs</i>	706	173 484	23.13	23.25
<i>Pronouns</i>	52	94 800	12.64	14.88
<i>Conjunctions</i>	20	113 823	15.18	12.01
<i>Numerals</i>	33	6 292	0.84	1.18

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 orðtíðnibók*, 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 orðtíðnibók*.

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 orðtíðnibók* 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 orðtíðnibók*.

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 orðtíðnibók*. 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 orðtíðnibók*; 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 – hofsgoði
 atgervimaður – atgervismaður
 höfuðbani – höfuðsbani
 hugboð – hugarboð
- b. aðdráttamaður – aðdráttarmaður
 affaradagur – affarardagur
- c. drukknan – drukkun, geipan – geipun
 auðigur – auðugur, ástúðigur – ástúðugur
 hraustleikur – hraustleiki, hvatleikur – hvatleiki
- d. atgervi – atgjörvi
 gagnvert – gagnvart
- e. heyrinkunnigur – heyrunkunnigur
 hlægur – hlæglegur – hlægilegur

In Icelandic, such formal differences are often due to different ways of compounding. In the Icelandic Sagas, both *hofgoði* and *hofsgoði* 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 *aðdráttarmaður* 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 *-anl-un*, *-igurl-ugur*, *leikurl-leiki* and others, in words like *geipanlgeipun*, *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 retslig Øiemed, for at fremme en Retssag o. desl.
3. voldeligt Overfald paa en i hans hjem, hjemsoegelse hvorunder man bruger Magten mod dem som ere i Huset

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

(6)

<i>heimsókn</i>	kv;	<i>heimsókn</i>	(5);	<i>heimsóknar</i>	(1);	<i>heimsóknina</i>	(1);	<i>heimsóknir</i>	(1)
Laxd	63;1632			og eigi mundi eg veita honum slíka				<i>heimsókn.</i>	
Gullþ	17;1138			að víg þeirra Helga skyldu á standast				<i>heimsókn</i>	
Vopnf	17;2003			eg brátt fara til Hofa og veita Bjarna				<i>heimsókn</i>	
HallM	10;1218			komi fyrir víg Galta og þar með				<i>heimsókn</i>	
HallÓ	12;1249			komi víg Einarssonar og þar með				<i>heimsókn</i>	
Svarf	15;1800			og snýr hann þá málinu öllu til				<i>heimsóknar.</i>	
Reykd	14;1755			víggunum en víg Narfa skal koma fyrir				<i>heimsóknina</i>	
Vatn	29;1878			Már kvað þá sýna af sér óvingjarlegar				<i>heimsóknir</i>	

Eða hvað er enn þá? Hann svarar og fjórráð við Þóri. Síðan var saman og sækja hann með eldi ef vér getum við Kolfinnu fyrir mannamun en fyrir við Kolfinnu fyrir mannamun. En fyrir þetta voru ráð Hrólfs nefgiltu og og brotttöku konunnar með en hinir og lést eigi mundu láta sinn hlut

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 *Sýnihefti sagnorðabókar* (Ásta Svavarsdóttir et al. 1993), which Orðabók Háskólans 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 *heimila*, 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.

(7)

heimill lo; heimil (3); heimila (5); heimilan (1); heimilast (1); heimill (4); heimilt (10); heimul (2); heimull (1); heimult (6)	
Ljósc 23:1697	Þorkell hitti þá brátt og mælti:
Ljósc 13:1673	hugsað hafa hver niður skal koma en
Vatn 16:1862	mælti: "Það er vel gert. Er þér og
Valjó 3:1831	Ljótur svarar: "Þú skalt eiga
Laxd 29:1575	beiða yður herra að þér létuð oss
Ljósc 8:1667	fylkismenn sögðu að þeir bræður sætu
Fósth 23:832	yfir þér að fjandinn á ekki þig svo
Vatn 17:1863	hitti stýrimann að máli og lét honum
GíslS 12:863	þar til húas og lét Þorvaldur honum
Hænsþ 5:1421	til væri." "Það muntu eiga allra
Valjó 8:1839	ná nokkrum þeirra." Narfi kvaðst
Fjót 19:712	"Óngvan sá eg nema sjálfan mig en
Vatn 31:1881	þar var öllum mönnum matur
VígGl 2:1907	konunga eða annarra höfðingja? Er þar
Eirík 7:529	vorum malt og mjöl og korn og er yður
Fósth 11:800	og skemmta þér hér? Nú er það
Njála 13:142	fyrir Glúmi ef hann er lengur, en
Hænsþ 10:1427	reka hingað fé sitt. Skal þeim það
GíslS 32:890	þú hefir sagt. Og mun þér nú þykja eg
HalMV 1:1221	Stýrimaður mælti: "Það skal ykkur nú
Eyrb 49:601að	maður skyldi jafnörgum mönnum eiga
Fígam 30:759	þar að vera "og allt mitt góss er þér
Valjó 6:1835	á skipinu fara. Ámundur kvað þeim
Vatn 12:1858	rammra hluta en fési. En það er
Fósth 3:779	Húskarlinn mælti: "Gakk inn þú.
Laxd 62:1630	segir hann, "að fylgja þér heim því að
Njála 136:296	mælti til Fíosa: "Því eru borð sett að
Laxd 70:1642	gangi fram." Þorkell svarar:
Egla 61:458	innngöngu í höllina. Honum var það og
Egla 71:478	til þessar ferðar þá mun yður það
Egla 73:482	sæja hestum sínum. Þorfinnur bóndi lét
Egla 33:408	og sagði svo að það var skylt og
Egla 41:417	umsjá." Þórir sagði að það var
Heimil	mun vist með mér sem fyrir þótt vant sé
heimil	munu þar til vor orð." Hann
heimil	vor mörk sem þú vilt höggva láta en eg
heimila	hegning ef oftar verður en legg nú
heimila	mörk ybra að höggva húsavið.
heimila	sveitarvist þeim sem þeir viðu.
heimila	til iltra hluta sem þú viðir gert hafa
heimila	vist með sér ef hann vildi. Hrafn kvað
heimila	hest sinn. Ríður hann nú við hrynjandi
heimilast	að veita öðrum þitt en eigi mitt."
heimill:	"Til þess er eg nú búinn. Er mér og
heimill:	er greiði við ykkur sá sem þér viljið
heimill	og hestaskipti og allur annar
heimill	vor tilbeini" - þá var Hákon
heimill	að hafa af sífki sem þér viljið og
heimill	að þú sért hér ef þú vilt það því að
heimill	á Glúmur að lofa það, en ekki er það
heimill:	Eg hefi hey ærin. Eru hér og nógar
heimill	eiga að gera af sífki er mér sýnist."
heimill	og kann vera að þið séuð hér eigi verr
heimill	róm í himinríki sem standa mættu í
heimill	til þess að þér megi þá betur líka við
heimill	þar að vera. Þóðvar kvaðst það mundu
heimill	þeim er fara vilja með mér. Hinum er
Heimul	mun þér gisting." Þorgeir segir:
heimul	mun þér gisting hér vera nálangt. En
heimull	er matur þeim er hafa þurfa."
Heimult	er þér mágur að eg gangi með máli þessu
heimult	gert. Gangu þeir Egill inn með helming
heimult	og allan farargreiða þann er þér viljið
heimult	skyldi það. Gangu þeir Egill þá inn í
heimult	um systur Þóris fóstbróður síns að hann
heimult	þó að Þórólfr vildi fleiri menn hafa

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; alldjarflega (8)

Þórb 13:2043	Síðan ganga þeir saman og berjast
Eyrb 62:616	menn til varnar og barðist sjálfur
ÞorSH 2:2062	koma milli klæða sér og barðist þá
Gullþ 15:1136	manna. Þá kom Gunnar að og barðist
Grett 2:955	fram þá hélt ekki við. Þórir barðist
Þórb 11:2039	öðrum stór sár. Steingrímur barðist
Grett 4:957	af skipunum. Víkingar lögðu að
Gullþ 12:1131	við túngarð í Rauðsdal og börðust

alldjarflega
alldjarflega
alldjarflega
alldjarflega
alldjarflega
alldjarflega
alldjarflega

. Fann Þórbur það brátt að Stóri var
. Hann gekk mjög út á virkið er hann
. Lítu síðar heyrðu þeir mælt í
. Þórir það sína menn hlífa sér og gæta
og féll á skipi sínu með mikilli
og varð fjögurra manna bani. Skiptist
og þóttu hinir komnir í stilli.
því að Raubur var frækn maður. En svo

(9)

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

HallM 4:1199	að og sækja að þeim en þeir verjast
Kjaln 16:1457	Kolfinnur hjó hart og tíðum og sótti
Eyrb 58:613	sköfustu en þeir Óspakur vörðust
JökBú 2:1463	þeir báðir að Eitli en hann varðist
Gullþ 20:1142	hvertveggi mjög sár. Þorgils varðist
Gullþ 8:1127	Hyrningur Hallsson kom heim og segir
Gullþ 13:1132	hjalla einum. Varðist Þorbjörn þaðan
Grett 82:1079	sótu að fast en Illugi varði þá báða

alldrengilega
alldrengilega
alldrengilega
alldrengilega
alldrengilega
alldrengilega
alldrengilega

. En þó kom þar sem mælt er að ekki má
. Hjóst þá allmjög skjöldur Búa. En er
. Varð þeim þó handfátt og urðu þeir
. Þar kom um síðir að þeir gátu drepið
en féll þó fyrir þeim Gunnari og Grími.
frá rtr þeirra Þóris. Hallur
með stokkinum því að vopn hans höfðu
. En Grettir var með öllu óvígur bæði

In (8) we see that the adverb *alldjarflega* and the verb *berjast* almost always go together; and (9) shows that the same goes for the adverb *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) *Orðastaður*, 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 they are. It is not clear on which principles the excerption has been based, and which of the excerpted examples actually appear in the dictionary.

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

(10)

<i>uns</i> st; <i>uns</i> (36)			
Fljót	18:709	Arneiðarstaði í kvöld og þóð mún þar	<i>uns</i>
Njála	142:306	Vil eg eiga rétting allra orða minna	<i>uns</i>
Grett	52:1033	enginn varðhöld á sér. Hann fór nú	<i>uns</i>
Þórð	9:2032	og ríður undan sem mest getur hann	<i>uns</i>
Eirík	3:522	ríða skyldu. Einar fór suður aftur	<i>uns</i>
Eirík	4:524	sagt. Þýr Þorbjörn skip sitt og fer	<i>uns</i>
Vopnf	2:1988	bolði mikla á hávu skafi. Hann fer	<i>uns</i>
BandK	4:32	þá fer Oddur með hinn tuttuganda mann	<i>uns</i>
GunKe	6:1153	þau skildu. Gekk Gunnar þá leið sína	<i>uns</i>
Þorhv	6:2056	yfir Lagarfljót og upp með fljótinu	<i>uns</i>
Fljót	21:720	Mývatns og fara svo vestur til sveita	<i>uns</i>
Vatn	40:1891	taka við honum og fór hann um sveitir	<i>uns</i>
Fljót	18:711	vermans. Nú dregur þó saman með þeim	<i>uns</i>
Svarf	24:1817	Karl Karisson vex upp með móður sinni	<i>uns</i>
GíslS	24:879	fer hann í Geirþjófsfjörð og er þar	<i>uns</i>
Grett	30:1001	annan mann. Þeir voru fimm saman, ríðu	<i>uns</i>
GíslS	3:854	er til komu. En þau Gíslí fara	<i>uns</i>
Njála	6:132	Hún bjó sig skjótt og síðan ríða þau	<i>uns</i>
Grett	90:1092	með góðu foruneyti og fóru alla leið	<i>uns</i>
Fljót	23:723	þeir ríða nú út eftir héraði	<i>uns</i>
Hrafn	14:1414	undan. Ríða þeir Sámur þá allt að einu	<i>uns</i>
Þórð	9:2033	ríða eftir þeim. Ríða þeir leið sína	<i>uns</i>
Fljót	13:697	hvorutveggju með hinu vestra landinu	<i>uns</i>
GíslS	7:857	þeir fara dag þann og um nóttina	<i>uns</i>
GíslS	29:887	þessu máli. Sveinarnir fara nú	<i>uns</i>
BjHft	30:114	þær er upp liggja í dalinn frá Völlum	<i>uns</i>
Korm	7:1476	Þetta ríð taka þeir, fara	<i>uns</i>
Vopnf	18:2004	Fara þeir nú sem þeir mega mest	<i>uns</i>
Fljót	18:708	til Öss og upp í heiði til Gínguskarðs	<i>uns</i>
Harð	36:1290	sté á skip reiður mjög og fara nú	<i>uns</i>
GíslS	11:862	Hann ljær þeim hestana og ríða þeir	<i>uns</i>
Fljót	26:727	nú ofan að ánni og upp eftir nesinu	<i>uns</i>
Njála	131:287	í braut og fyrir norðan jökul og svo	<i>uns</i>
Heið	22:1373	og styðjið hana á baki og ríðið svo	<i>uns</i>
Grett	19:985	hans alla hluti vel. Leið nú svo fram	<i>uns</i>
Harð	38:1292	ofan Indriðastíg hjá Þyrli og beið þar	<i>uns</i>
		eg kem á morgun, nema eg finni eigi	<i>uns</i>
		eg kem máli mínu til rétra laga.	<i>uns</i>
		er hann kom í Vatnsfjarðardal og fór	<i>uns</i>
		hann kemur heim til Þverár og unnr	<i>uns</i>
		hann kemur heim. Nokkuru síðar	<i>uns</i>
		hann kemur í Brattahlíð. Tekur Eiríkur	<i>uns</i>
		hann kemur í sauðabás og rekur þaðan	<i>uns</i>
		hann kemur skammt frá bæ þeim er	<i>uns</i>
		hann kom heim. Urðu hans menn honum	<i>uns</i>
		hann kom í Atlavík snemma morguns.	<i>uns</i>
		hann kom til Hjelgafells, því að hann	<i>uns</i>
		hann kom til Öttars og það hann	<i>uns</i>
		hann kom upp í ásinu. Þá hallar burt af	<i>uns</i>
		hann var tólf vetra gamall og töluðu	<i>uns</i>
		haustar. Þá fer hann enn til Þorkels	<i>uns</i>
		þar til er þeir komu á	<i>uns</i>
		þau koma í Friðarey til Styrkás og	<i>uns</i>
		þau koma til þings. Unnur gekk	<i>uns</i>
		þau komu í Noreg. Tóku	<i>uns</i>
		þeir koma á Arneiðarstaði. Er gengið	<i>uns</i>
		þeir koma á heiðarbrúna. Sá hann þá	<i>uns</i>
		þeir koma á Miklabæ í Öslandshlíð.	<i>uns</i>
		þeir koma fyrir nes það er gengur	<i>uns</i>
		þeir koma í dal þann er gengur upp af	<i>uns</i>
		þeir koma í Geirþjófsfjörð og liggja	<i>uns</i>
		þeir koma í Hólmsland og töluðu þá	<i>uns</i>
		þeir koma í Saurbæ. Tók Bersi vel við	<i>uns</i>
		þeir koma mjög svo í öndverðan dalinn.	<i>uns</i>
		þeir koma svo upp að ein brekka var	<i>uns</i>
		þeir koma þar að er Geir flaut dauður	<i>uns</i>
		þeir komu á Mosvöllu og þaðan inn	<i>uns</i>
		þeir komu á Völvöllu. Þá var af	<i>uns</i>
		þeir komu til Svínafells. Flosi sendi	<i>uns</i>
		þér komið fram að Faxalæk. Hann fellur	<i>uns</i>
		Þorfinns var heim von.	<i>uns</i>
		Þorsteinn fór til blóthúss síns sem	<i>uns</i>

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, *þar 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 Ingólfssdóttir, 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

- Ásgeir Blöndal Magnússon 1989: *Íslensk orðsifjabók*. Reykjavík: Orðabók Háskólans.
- Ásta Svavarsdóttir/Guðrún Kvaran/Jón Hilmar Jónsson/Kristín Bjarnadóttir 1993: *Sýnihefti sagnorðabókar*. Rannsóknar- og fræðslurit 3. Reykjavík: Orðabók Háskólans.
- Baldur Jónsson 1978: *Orðstöðulykill að Hreiðrinu*. [Fjölrít.]
- Bíblíulykill*. 1994. Orðalyklar að Bíblíunni 1981. Reykjavík: Bíblíulykilsnefnd/Hið íslenska Bíblíufélag.
- Eiríkur Rögnvaldsson 1990: Orðstöðulykill Íslendinga sagna. *Skáldskaparmál* 1, 54–61.
- Eiríkur Rögnvaldsson/Bergljót Kristjánsdóttir/Guðrún Ingólfssdóttir/Örnólfur Thorsson 1995: *Orðstöðulykill Íslendinga sagna*. Forthcoming on CD-ROM.
- Fritzner, Johan 1954: *Ordbog over Det gamle norske Sprog*. Nytt uforandret opptrykk av. 2. utgave (1883–1896) med et bind tillegg og rettelser redigert av Didrik Arup Seip og Trygve Knudsen. Oslo: Trygve Juul Møller.
- Íslendinga sögur*. 1985–86. Editors: Bragi Halldórsson, Jón Torfason, Sverrir Tómasson, Örnólfur Thorsson. Reykjavík: Svart á hvítu.
- Jón Hilmar Jónsson 1994: *Orðastaður*. Orðabók um íslenska málnotkun. Reykjavík: Mál og menning.
- Jörgen Pind/Friðrik Magnússon/Stefán Briem 1991: *Íslensk orðtíðnibók*. Reykjavík: Orðabók Háskólans.