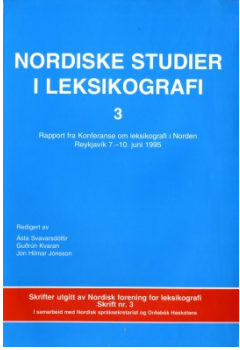


# NORDISKE STUDIER I LEKSIKOGRAFI

Titel:	A Dictionary of the Icelandic Family Sagas and Semantic Classification	
Forfatter:	Guðrún Ingólfssdóttir	
Kilde:	Nordiske Studier i Leksikografi 3, 1995, s. 167-174 Rapport fra Konferanse om leksikografi i Norden, Reykjavík 7.-10. juni 1995	
URL:	<a href="http://ojs.statsbiblioteket.dk/index.php/nsil/issue/archive">http://ojs.statsbiblioteket.dk/index.php/nsil/issue/archive</a>	

© 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 Dictionary of the Icelandic Family Sagas and Semantic Classification

Artikkelen handler om ett prosjekt som består i en leksikografisk beskrivelse av vokabularet i islendingesagaene på grunnlag av et korpus som inneholder en standardisert tekstutgave av samtlige sagaer. Beskrivelsen av substantivene utgjør den første fasen i prosjektet. For å komme fram til mest mulig presise definisjoner, blir substantivene først klassifisert ut fra forskjellige semantiske kriterier, noe som gir en oppdeling i omtrent 50 ulike semantiske kategorier. Denne framgangsmåten, som ikke har vært praktisert tidligere i en islandsk ordbok, bidrar til en konsistent behandling av de ord som knytter seg til de enkelte kategorier. Samtidig blir det lettere å avdekke relevante semantiske skilnader. Den semantiske klassifiseringen har vist seg å være til stor nytte for forskere på forskjellige områder.

### 1 Introduction

In this paper I will describe a part of a work in progress, bearing the working title *A Dictionary of the Icelandic Family Sagas*.<sup>1</sup> The project, at least in its first phase, only covers the Icelandic Family Sagas. The corpus is therefore finite, and furthermore, the Sagas only describe a rather limited subset of human affairs and actions. Because of this, a detailed semantic description of the vocabulary gives a clear and detailed picture of the world of the Icelandic Family Sagas. The dictionary is developed as a database.

The concordance itself is described by Eiríkur Rögnvaldsson (1995; this volume), and this is the first time that a concordance has been used as a basis for a dictionary in Iceland. The concordance facilitates the work on the dictionary considerably, in terms of explaining meanings, classifying into grammatical categories and, last but not least, of grouping formulaic (or quasi-formulaic) expressions or unusual syntactic structures. We have also noticed that semantic explanations are more precise when working on every example of a word at one time.

The following examples (cf. fig. 1–6) shows how the database is structured. As each word is entered into the database it is divided into its components. Affixes and inflectional endings are indicated with special symbols, '+' is placed between a suffix and a root, and '-' between a stem and an inflectional ending. Prefixes and suffixes are also entered into separate fields. Each word is marked for gender, and the frequency of each case-form is recorded. This work can proceed fairly quickly, as all words are already classified in the concordance, which also gives the frequency of every form of the word. Grammatical categories are given

<sup>1</sup>The main participants in this project are Bergljót Soffía Kristjánsdóttir and myself, but Margrét Guðmundsdóttir, Svanhildur Óskarsdóttir, Eiríkur Rögnvaldsson and Örnólfur Thorsson have also taken part. None of the participants is either a lexicographer or an engineer, and our specialization is mainly in medieval literature, apart from Eiríkur Rögnvaldsson and Margrét Guðmundsdóttir who are linguists.

categories are given as in other dictionaries, the difference being that in this dictionary the number of categories is greater, and the analysis is based on syntactic context. This will be useful for linguists or scholars studying the literary characteristics of the sagas. The definitions are arranged according to frequency, and phrases and constructions are explained under separate headings following the basic definition. There are also fields for linguistic, semantic, literary and other remarks.

## 2 The Corpus

Before coming to my main subject, i.e. the semantic classification system, a few words on the nature of the corpus are due. Many people think that the true meaning of every word in a certain text can be found by looking the word up in a dictionary. While it is of course true that the basic meanings of words, and even some constructions, are defined in dictionaries, these definitions usually only take ordinary, every-day language into account. The problem is that words are not isolated objects, and in fact it can be maintained that words only have a meaning in some context. A word in isolation is therefore only a skeleton, without flesh and blood. In traditional dictionaries, however, the symbolic meaning or various connotations of words are rarely given, because these are more closely connected with fiction and often refer to mythical reality. The lexicographer, on the other hand, is of course always interpreting ordinary human language, and trying to find the basic meaning of words, and in a certain sense also the true meaning.

In our project, the fields for semantic and literary remarks can be used for explanations that fall outside the scope of an ordinary dictionary. There we can play with the fiction which is a part of every word. To find out the basic meaning of words, the lexicographer must often tear the words from the earth that in fact gives them their life and colour. The greater flexibility of our form gives full scope to the context of the words, and therefore we may be seeing some new possibilities in dictionary making. We must admit that this form is in many ways both complicated and time-consuming, but as pointed out by Eiríkur Rögnvaldsson (1995; this volume), the soil, i.e. the syntactic structure, is more significant here.

Orð	brjóst% Kyn	hk	Fsk	Vsk
Nf.et.	brjóst	2		Pf.et. brjóst 32 18
Bgf.et.	brjósti	17	3	Ef.et.
Nf.ft.				Pf.ft.
Bgf.ft.				Ef.ft.
e Með fs.			á/fyrir/í brjóst; á/í brjósti, !e-ð gengur úr brjósti	
l Orðasb.			e-m býr e-ð í brjósti	
m Klasi			venja af brjósti, bera e-ð í brjósti	
Merking 1			bringa, barmur	
Merking 2			kirtill sem framleiðir mjólk	
Merking 3			hugur, innræti	
Merking 4			fremsti hluti e-s %hlíf, vörn	
Merkingarl. aths.			í Flóamanna sögu (24;754,772) er Þorfinnur vaninn af brjósti föður síns svo það eru ekki bara konubrjóst sem framleiða mjólk: 'Latur hann [Þorgils] nú saxa á geirvörtuna á sér og kemur þar blóð út. Síðan latur hann teygja það og kom þar út blanda og eigi lét hann af fyrr en það var mjólk og þar fæddist sveinninn við.' (Flóam 23;750)	

Figure 1.

Some disadvantages of working with a closed corpus should be mentioned, as shown in the example of the word *brjóst* 'breast' (cf. fig. 1). By looking the word up in the concordance, two examples of the construction *að venja af brjósti* 'to wean (a child)' can be found. In the first one, from *Fjótsdæla saga*, a woman weans her daughter. In the second example, from *Flóamanna saga*, a young boy is being weaned, but the breast in question is that of his father. These examples can indeed give a defect picture of medieval society, but on the other hand they also show that there is nothing new beneath the sun.

The advantages of working with a closed corpus are in fact far greater than the disadvantages. To name an example, the word *rass* 'ass' only appears four times in the Icelandic Family Sagas (cf. fig. 2). In all cases it refers to the ass of Guðmundur ríki Eyjólfsson, a great chieftain in the northern part of the country in the beginning of the 11th century. The word is used by the enemies of Guðmundur when trying to make fun of him, humiliate him, and thus question his power and authority. Their point of attack is Guðmundur's ass, the true symbol of his supremacy and position, doubly apt as he was a very fat man.

Therefore the word *rass* does not really appear in its basic meaning in the texts, but only as a symbol of something else, i.e. as a personification. In our dictionary or database we show the basic meaning, but in addition the contextual meaning of these examples is shown in the field **Bókmenntl. aths.** 'literary remarks'. Such connotations are not usually found in ordinary dictionaries, but our database gives us the scope for such contextual meanings alongside the basic meaning.<sup>2</sup>

Orð	rass	Kyn	kk	Fsk	Vsk
Nf.et.	rass		2	Df.et.	rass 2
Dgf.et.				Ef.et.	
Nf.ft.				Df.ft.	
Dgf.ft.				Ef.ft.	
a Frl.so.	rass sveitist				
q Acccinf	kveða rass hafa freistað/leitað flestra lækjanna				
Merking 1	afturendi, both				
Bókmenntal. aths.	2LjósA, 2LjósC. Orðið er einungis notað um afturendann á Guðmundi ríka á Möðruvöllum, en hann er áþreifanlegt tákni um veldi hans (sbr. Guðrún Ingólfsdóttir. 1993. Að eiga mikið undir sér. Orðaforði heyrjaður Guðrúnu kvaran. Reykjavík:37-38.)				

Figure 2.

Another example, showing the importance of the context, is the word *brottaka* 'removal of something from somebody' (cf. fig. 3). There are three examples of the word in the corpus, from three different sources, i.e. *Kjalnesinga saga*, *Kormáks saga*, and *Reykdale saga*. Although the basic meaning of the word is 'the removal of something from somebody', all the examples are of men stealing other men's wives. This specific negative usage is dominant in the Icelandic Family Sagas, and the usage shows that a wife, in medieval society, was often thought of as her husband's property. These examples also show that one result of working with a rather closed world is that a very clear picture emerges of at least some aspects of the words, making clear the interactive influence of language and society.

<sup>2</sup>Similar observations are made in the description of the Cobuild Project (cf. Moon 1987:99).

Orð	brott tak-að	Kyn	kv	Fsk	Vsk
Nf.et.				Þf.et.	brotttöku 2
Þgf.et.	brotttöku	1		Ef.et.	
Nf.ft.				Þf.ft.	
Þgf.ft.				Ef.ft.	
c Andlag spyrja brottöku e-s					
e Með fs. fyrir brottöku e-s; í brottöku e-s					
Merking 1		brotnám; rán			
Bókmenntal. aths.		1Kjaln, 1Korm, 1Reykd. Vert er að gefa gaum að því að orðið er einungis haft um rán á konum			
Aðrar athugas.		sjá brotttekja			

Figure 3.

It is not only the syntactic structure or the reference to symbolic reality that has effect on the meaning of words. Things like human behaviour also characterize the words and form their meaning. This shows better than anything else how complicated the meaning of every word can be, how delicate the interplay of the corpus and the words of that corpus can be, and how difficult the task of the lexicographer is. Despite all the traditions in dictionary making, it is indeed good to question them from time to time.

In our project, the text leads the way, and a broad overview of the whole text is necessary, as the true meaning of a word can be hidden at some distance to the actual occurrence of the word itself. To name an example, the word *blíðmæli* is explained in Fritzner's *Ordbog over Det gamle norske Sprog* (1883–96) as 'venlig Tale' (i.e. 'gentle talk'), quoting one example from *Flateyjarbók* and two from *The Lives of Saints*, but none from the Icelandic Family Sagas. At first sight, this meaning is also the right one for the occurrence in the Icelandic Family Sagas (cf. fig. 4). Closer scrutiny of the story itself indicates that the word may have the supplementary meaning 'flattery', i.e. 'gentle talk and flattery.'

Orð	blíð+mæli-ið	Kyn	hk	Fsk	Vsk
Nf.et.				Þf.et.	
Þgf.et.				Ef.et.	
Nf.ft.	blíðmæli	2		Þf.ft.	
Þgf.ft.				Ef.ft.	
a Erl.so.	blíðmæli tjóa				
o Samsetn	#ógnir né blíðmæli				
Merking 1		vingjarnleg orð			
Málfræðil. aths.		ef marka má Fritzner er orðið ótvírætt ft í fornu máli			
Merkingarl. aths.		vert er að kanna hvort orðið hefur merkingarpáttinn smjaður, fagurgali!			
Bókmenntal. aths.		1LjósA, 1LjósC			

Figure 4.

Another such example is the word *útibúr* 'out-house, store-house' (cf. fig. 5). Dictionaries, including Fritzner's, explain it as 'a store-house, detached from other houses on the farm'. In the Icelandic Family Sagas, however, there is one example where the store-house is adjacent to the kitchen.

Orð	úti+búr%	Kyn	hk	Fsk	Vsk
Nf.et.	útibúr	4		Pf.et.	útibúr 11 4
Þgf.et.	útibúri	12	4	Ef.et.	útibúrs 3 2
Nf.ft.				Þf.ft.	útibúr 2
Þgf.ft.				Ef.ft.	
a Frl.so.	útibúr stendur í túni				
b Frl.sf.	útibúr er opið/óbrunnið				
c Andlag	brjóta upp útibúr, eiga/sjá útibúr				
e Með fs.	+í útibúr; að/af/á/í/úr útibúri, vaka yfir útibúri; til útibúrs				
f Með ek.	mikill/(ákaflega) stór				
g Sem ek.	dyr útibúrs				
j Ópersl.	útibúr skortir eigi				
Merking 1	afhýsi eða lítið hús (oftast) laust frá íbúðarhúsum þar sem matur var geymdur & matarbúr, matargeymsla				
Merking 2	skemma, geymsla				
Málfræðil. aths.	vera áfast útibúri er þgf með lo				
Aðrar athugas.	í einu tilvikum er útibúrið áfast eldhúsi. Sjá búr				

Figure 5.

Working so closely with the text, examples of meaning not found in other dictionaries are often found. Such an example is the word *tiltekja* (literally, 'what a man takes to doing'), where extant dictionaries have the explanation 'an expedient, contrivance', but not the additional meaning 'intention, plan' (cf. fig. 6).

Orð	til+tek+j-a%	Kyn	kv	Fsk	til	Vsk	j
Nf.et.	tiltekja	5	1	Pf.et.			
Þgf.et.	tiltekju	1		Ef.et.			
Nf.ft.				Þf.ft.	tiltekjur	8	
Þgf.ft.	tiltekjum	1		Ef.ft.			
a Frl.so.	e-m mislíkar tiltekja						
b Frl.sf.	tiltekja er meyleg, tiltekja þykir ill/óvarleg/undarleg						
c Andlag	hafa/spyrja tiltekju, vita tiltekju e-s; ráða tiltekju sinni						
e eð fs.	fyrir tiltekju sína, (ræða) um tiltekju; yfir tiltekju e-s						
f Með ek.	annar/hann/hún/sá/sjá/sinn/þessi/pinn						
o Samsetn	tiltekja og drottinssvik/hneisa						
p Andh.							
Merking 1	tiltæki, uppátæki; verk, athöfn						
Merking 2	áform						
Merkingarl. aths.	merk. 2 er ekki að finna í öðrum orðabókum						

Figure 6.

### 3 The List of Semantic Classification

I will now turn to the list of semantic classification. At the outset of the project we decided to define each part of speech separately, starting with the nouns. We thought this procedure would make the task of defining words easier and ensure greater consistency. Soon we also saw that we could improve this method considerably by classifying all the nouns semantically. Such a semantic classification of nouns has never before been used in an Icelandic dictionary.

It has often been maintained that the subject matter of the Icelandic Family Sagas is very narrow, and this is in some respects true. The list below, containing the table of contents of the semantic classification, shows, however, that the semantic classes touch upon nearly every aspect of life.

As shown below in (1), there are 21 primary classes in the list as it now stands, and the total number of classes comes to 54, when the secondary classes are included. The classes are ordered by size, and the relative size of the semantic classes is interesting in itself. The largest class (about 8 pages) is the class *Human nature*. For readers of the Icelandic Family Sagas, this is hardly surprising, as their main subject is man and nature, and *Nature* is the second largest of the classes (about 4 pages). Next in line are the classes *Culture and Learning* and *Conflict* (about 3 pages each). Then comes *Economy* (2½ pages), and *Language of Legal Affairs and Administration* (more than 2 pages). *Human Condition* and *Buildings* are 2 pages each, but other semantic classes are 1 page or less.

The size of the classes emphasizes the well-known fact that the Icelandic Family Sagas describe a feudal-rural community, where man, nature and feelings are in place of honour. It is, however, rather difficult to guess what the result of applying the same classification to all the words in a dictionary such as Fritzner's *Ordbog* would be, as his corpus is very much larger. The outcome would certainly be bound to be interesting!

(1)

<i>Human nature</i>	People Characteristics Thinking, Perception, Feelings
<i>Nature</i>	Animals Heaven and Earth Weather Vegetation Source of Energy, Metals, Minerals
<i>Conflict</i>	Warfare Armament
<i>Culture and Learning</i>	Story-telling and Studies, Poetry and other Arts Language Written Language Quantity, Directions, Numbers Time Games and Sports
<i>Economy</i>	Agriculture Fishing and Hunting Trade and Merchandise Handwork (Needlework, Sewing, Knitting) Construction
<i>Language of Legal Affairs and Administration</i>	
<i>Buildings</i>	Houses and Dwellings Other Buildings

<i>Human condition</i>	Death and Life Healing Wounds and Injury Diseases Human body Cleaning and Leftovers
<i>Transport</i>	Travels (in general) Travels on Land Ships and Sailings
<i>Tools, Instruments and Vessels</i>	
<i>Respect and Power</i>	Wealth Gifts Festivals, Celebrations and Hospitality
<i>Clothes and Ornament</i>	
<i>Religion</i>	Places/Houses Objects Creatures Events
<i>Family</i>	Children and Upbringing Family Mariage
<i>Food and Drink</i>	
<i>Women</i>	
<i>Fortune and Misfortune</i>	
<i>Inhabited Area</i>	Countries Urban Area and Rural Area
<i>Dreams</i>	
<i>Form</i>	
<i>Deficiency/Lack, Difference</i>	

When we started to classify all the nouns semantically, we tried to give an overview of most of the human and cultural aspects of everyday life, but we also had to take notice of the kind of corpus we were working on. We then proceeded to classify every noun according to its basic meaning. In the pilot list we did not take any notice of secondary meaning, but each word could be placed in more than one semantic class. Three versions of this list have now been made, and in the second version the semantic classes are arranged like in a thesaurus.

The semantic classes are not all classified by this system (e.g. *Language of Legal Affairs and Administration, Clothes and Ornament, Food and Drink, Tools, Instruments and Vessels, Deficiency/Lack, Difference*, etc.). Some of them are too small, and others have not been fully defined yet. The question of how exact the classification should be also remains open.



In the beginning, the classification was not very thorough, mainly because the size of most of the semantic classes did not really call for too much detail, i.e. the lexicographer could easily gain a fairly good grasp of the whole of each class without it.

We have, however, made some experiments with a more detailed classification, such as shown in the class *Religion*. After finishing the work on this semantic class, the person concerned revised the classification, giving the end result shown in (1) above. This semantic class is rather simple, and I think that the result is convincing. I am not totally satisfied, however, because this kind of classification is often bound to make things very complicated. The problem is that the system allows words to be a part of more than one primary class, and furthermore, they can also be a part of more than one secondary class. This, I fear, will make the use of this list rather confusing. I am aware of the fact that not everyone agrees with me.

The list of semantic classification has been revised a number of times, in order to make it a better tool for scholars who are working on the Icelandic Family Sagas. Many scholars from different fields of study have obtained material from the list, using the information in connection with the concordance. This considerably speeds up the search for all examples pertaining to the things they are researching.

There are mainly two advantages of using a list of semantic classification, such as ours, in a dictionary project. Firstly, it gives the lexicographer a unique perspective of the semantic field being worked on each time, and prevents concepts with the same meaning from being defined in different ways. At the same time, it can reveal subtle semantic differences, which would otherwise be unnoticed. Secondly, it can sometimes reveal more clearly how related concepts differ, e.g. in usage, etc. — Those are the same features as found in the Cobuild Project.

When we decided to classify the nouns semantically, we mainly had two things in mind. In the first place, we wanted to improve the processing method, and secondly, we felt the need for some progressive steps to be seen, by making some parts of the project useful before the completion of the project, especially as it is such a huge one. The list of semantic classification is such a progressive step. It is both a useful tool in the dictionary project itself, and in various research on the Icelandic Family Sagas. — Later this year, when the concordance will be published, the list of semantic classification will also be published with it, and thus get a still greater distribution.

## References

- Eiríkur Rögnvaldsson 1995: A Concordance to Old Icelandic Texts and its Lexicographic Value. [This volume, 123–135.]
- Fritzner, Johan 1883–96: *Ordbog over Det gamle norske Sprog. Omarbejdet, forøget og forbedret Udgave*. I–III. Kristiania: Den norske Forlagsforening.
- Fritzner, Johan 1972: *Ordbog over Det gamle norske Sprog IV. Rettelser og tillegg ved Finn Hødnebo*. Oslo/Bergen/Tromsø: Universitetsforlaget.
- Moon, Rosamund 1987: The analysis of meaning. In: J. M. Sinclair (ed.): *Looking up*. An account of the COBUILD Project in lexical computing. London: Collins.