Jóhannes Bjarni Sigtryggsson & Christopher Sanders

# The Copenhagen Old Norse Word-list[1]

## 1 Introduction

A list of headwords that will be treated in the dictionary of Old Icelandic and Old Norwegian prose — currently being published by the Arnamagnæan Commission in Copenhagen — has now been produced.[2] The article describes the background, production, and expected benefits to be derived from this list (subsequently referred to as "the word-list").

## 2 Background

The Arnamagnæan Commission's *Ordbog over det norrøne prosasprog // A Dictionary of Old Norse Prose* (subsequently ONP) has, as many readers of this journal will be aware, an extensive past history.[3] It was conceived before the use of computers in linguistics and lexicography was a commonplace, but has, since the middle of the 1980s, made extensive use of computer technology. One of the features that its early planners had not taken into consideration was the need for a comprehensive list of the headwords that the dictionary should contain in its published form. A list of this type, given that it should also specify word class and the number of citation slips (the number of tokens) that represent each headword, had gradually become essential to the planning of the production of a twelve-volume work. Words of different classes have varying

---

[1]The authors are, respectively, the editor of the word-list and, as one of the six current editors of the dictionary, the project supervisor.

[2]The underlying corpus of the dictionary is the prose works of Icelandic up to ca. 1540 and Norwegian prose up to ca. 1370.

[3]Since the appearance of the *Prøvehæfte // Prospectus* in 1983, the following two volumes have been published: *Registre // Indices* (1989) and *ONP 1 : a-bam* (1995) (for the explanatory booklet that accompanies *ONP 1: a-bam*, see footnote 5 below); for recent general descriptions of ONP see Rode (1991:67–71) and Sanders (1997:81–88).

inherent tendencies with regard to how long they take to edit and how much space they occupy on the printed page. Verbs and prepositions/adverbs notoriously take longer and are harder to reduce to comprehensible and useful dictionary articles than a great many nouns. Both nouns and verbs that are represented by very few tokens, can, however, be exceptionally time-consuming in terms of arriving at a semantic description, simply because there are "too few" citations — insufficient empirical material — on which to base an immediately convincing definition or explanation. Factors of this type need to be taken into consideration when planning the future production of the dictionary, year by year and volume by volume. The dictionary is regrettably produced in alphabetical order, so it was also evident that a list of this type would be a useful tool in enabling editors to relate words that are semantically or in other ways connected to one another, even though they are widely separated in terms of their position in the alphabet; most specifically cross-references could be incorporated in the dictionary in a reasonably consistent manner.

A further benefit is the advantages that the list could provide for anybody engaged in research on early Nordic languages and culture (see §8 below).

Because of a strict production schedule it was evident that the regular dictionary staff did not have the time available to produce a word-list of the type described above. The Arnamagnæan Commission applied to Lýðveldissjóður in Reykjavík for additional funds and in 1998 received a generous grant which provided the stimulus for work to be started. Further grants were made in 1999 and at the beginning of the year 2000, such that all the costs involved were generously met by Lýðveldissjóður. The project description entailed a full year's work for a person well-qualified in Old Icelandic and 0.85 man-years of work for a secretarial assistant. Jóhannes Bjarni Sigtryggsson was appointed from amongst a number of applicants for the post as the scholar responsible for the project and he commenced work in Copenhagen on September 15th 1998; Alex Speed Kjeldsen was employed as the principal student assistant and other assistants were subsequently brought into the project on the secretarial level.

## 3    The word-list and ONP's computer system

When the project started, all of the letter *a* and almost all of the letter *b* had been edited in final form, so for them no new word-list was needed. To understand further how the word-list was produced it is necessary to have some insight into the way in which the dictionary's database system operates. It is also necessary to appreciate that only a small percentage of the dictionary's total collection of citations has been keyed in, this process having only got as far as the word *en*. All of the dictionary's materials are in reality independent, but structurally correlated, database files.[4] The database system incorporates a program that can, at any stage of the editorial process, produce printouts for assessment and proofreading. When an operator instructs the database to produce the edited copy for a given sequence of words, the database first consults its

---

[4]Cf. Degnbol et. al. (1992:385–389).

word-file to extract the headwords that have been selected for editing and publication. This consultation of the word-file is the very start of the process, and, in order to avoid duplication of data, this word-file has been transmuted into the word-list that is being described here. The word-file also contains all of the information that generates the section at the end of a standard dictionary entry introduced by the abbreviation *Gloss.* for "glossaries",[5] but the preparation of that material is not part of the present project.

# 4   How the word-list was produced

ONP's raw material is the handwritten slips that have amassed over the years. These contain the citations on which the grammatical (syntactic) and semantic analysis is based, and each has in its top right-hand corner the headword that was deemed appropriate for each token at the time that it was excerpted. Until now, no detailed count had been made of these slips, only approximate estimates of their quantity. There was an additional complication in that in the course of time two collections had developed: one that represented the corpus as it was at the beginning of the sixties ('gammel samling' // old collection), and another containing the slips that had subsequently accrued ('ny samling' // new collection). These two collections had to be combined before compilation of the word-list could start, and this was the first work-phase (a). Secondly (b), a primitive list of words was prepared on the basis of ONP's existing database file of compound words and the dictionary *Norrøn ordbok* (1975) (subsequently NO); thirdly (c), this initial list was compared with ONP's collection of dictionary slips. This final phase was the cornerstone of the project, since it was here that significant additions were made and earlier inaccuracies put to rights.

a)  In the process of combining the old and new dictionary collections, improvements were made to the physical organisation of the slips. With the help of differently coloured dividers, inserted between the filecards, it is now possible to quickly identify different types of slips and different types of vocabulary. This is invaluable for those who subsequently key in the slips and for dictionary editors and visiting researchers who need to consult not only the computer files but also the physical archives. It also greatly facilitated the third phase (c) of producing the word-list, where quick identification of the nature of the material that was to be compared with the inventory of words in NO was an important factor.

b)  Before the word-list in its present form was conceived, ONP had already prepared a database file of the compound words that had been registered in the course of excerpting. All the compounds in this file were copied to form the germ of the new word-list. The vocabulary of the dictionary NO was selected as the primary supplement to the embryonic word-list, so its inventory of headwords was additionally keyed in by the secretarial assistant.[6] NO was chosen because it contains

---

[5]See pp. 46–47 in Nøgle // *Key* (1995), the explanatory booklet distributed with *ONP 1 : a-bam.*
[6]The NO vocabulary for the letters *d-h* had been keyed in before the start of the project.

a reasonable representation of the contents of nineteenth-century scholarly dictionaries (Cleasby/Vigfusson and Fritzner) while incorporating the new material introduced in the supplementary volume to one of them (Hødnebø 1972).

**c)** The list thus arrived at, consisting of the compound words in ONP's archives and the vocabulary of NO, could now be augmented by the total vocabulary of ONP itself through comparison with the dictionary slips. This was the task of Jóhannes. Under considerable pressure of time hasty assessments had to be made. Apart from the appreciable number of additions to the word-list in the form of lexical items that had not been previously registered, it was necessary to check: i) that the tokens were correctly lemmatized (organised under the correct headword); ii) that the headwords were spelt in accordance with ONP's standard normalisation (see §6 below); iii) that the grammatical details were an accurate representation of the usages recorded on the slips (wordclass and gender). Furthermore a considerable number of selections had to be made. This is not the place to give a full presentation of the limits governing ONP's corpus, but a short explanation is appropriate (see §5 below).[7] During the process a number of what had been registered as single words were divided up into two or more headwords, and some items that had previously been separate were combined into one (see §5 below). Consultations sometimes had to be carried out with the regular editorial staff, and when everything was in place the number of handwritten slips designated as representatives of each headword was counted by the secretarial assistants, and this information recorded in the database file (for the statistical results, see §7 below).

## 5   Corpus considerations

ONP does not treat place-names or personal names (except cognomina/nicknames). Poetic vocabulary and unassimilated foreign words (as distinct from assimilated, or partly-assimilated, loanwords) are registered in the printed volumes but without analysis or explanation. These delimitings of what areas of vocabulary ONP is to cover have been the subject of much consideration throughout the history of the project and were finally crystallized when the index volume, *Registre // Indices*, was produced in 1989. Factors of time and money were significant, but important too was an awareness that ONP's excerpting in some of these debated areas was insufficient. Quite a number of slips have been written in the course of time containing onomastic material and foreign words, yet the excerpting was too irregular and inconsistent for a scholarly analysis. From the point of view of planning it is, however, necessary to know how many slips of this type there are in the ONP archives. Consequently there is a detailed system in ONP's database procedure for tagging the various types of deselected or "marginal" material. This had to be executed in full detail during the third phase (c) of the preparation of the word-list.

---

[7]See also *Nøgle // Key*, p. 30.

# 6   Caution!

The word-list from the letter *c* (of which there are a small number of representatives — all of foreign origin) to the end of the alphabet was completed in the allotted time. In order to achieve this it was necessary to accept a considerable degree of discrepancy between the word-list on the one hand and the finally edited dictionary on the other. The level of differentiation is not the same. In the dictionary itself there are so-called alternative forms printed in bold type alongside the main form of the headword.[8] A decision on whether or not to incorporate an alternative form in an article is the result of detailed editorial assessment which could not be undertaken while producing the word-list. The number of cross-references to be contained in the word-list was also deliberately kept to a minimum, again awaiting the editorial process, so a number of forms or spellings that are potential alternative forms are not immediately searchable in the word-list, but are "concealed" under a rubricized main form. A similarly hard-handed treatment was necessary when dealing with compound words that contain productive final elements (for example, since the form *morginn* was selected as the headword for that particular entry, all compound words that have 'morning' as their final element are registered under the spelling *-morginn*, even though spellings in *-morgunn*, *-myrginn*, *-merginn*, and *-murginn* are recorded on the slips; similarly all compounds that end with the element 'sea' are registered as *-sjór*, despite various occurrences in *-sjár* and *-sǽr*). Some of these necessary crudities will be eliminated when the words are subsequently edited for publication. In order to avoid unnecessary disappointments during a search, future consultors of the word-list will do well to familiarize themselves with ONP's guidelines on normalisation, which will soon be made available on the dictionary's homepage.[9] It is also useful to acquaint oneself with ONP's practice with regard to cardinal forms; thus, for example, adverbs which manifest themselves as declined forms of an adjective are registered and cited as integrated elements in the article for the relevant adjective (*bráðan*, *bráðum* and *brátt* under *bráðr*, etc.).

# 7   Statistical results for the entire alphabet

**a)** number of handwritten dictionary slips: 726.092

**b)** to this must be added ca. 6% to accommodate entries containing marginal material for which there are no handwritten slips (containing for example onomastic, poetic, and foreign vocabulary):

ca.     43.500
ca.    770.000

**c)** to this must be added ca. 17% increase due to supplementary excerpting:[10]

Result: *approximate total number of tokens* ca. 900.000

---

[8]See *Nøgle // Key*, pp. 31–32.

[9]www.onp.hum.ku.dk

[10]During the production of *ONP 1 : a-bam* it was established that it was necessary to extract supplementary citations from previously published dictionaries and glossaries.

# 8   The future

Although the initial impulse to make the word-list was administrative, the benefits to future word studies both within the dictionary and in a wider scholarly circle should be considerable. It will, to take random examples, be possible to search for all the nouns that end in *-hvalr* (should one be interested in whales, see fig. 1) or all the nouns that end in *-leikr* or *-leiki* (should one wish to continue the research highlighted by the Norwegian scholar Alfred Jakobsen (1970), concerning, amongst other things, the distribution of these two suffixes; cf. also Walter 1972). The ability to conduct this type of search will, amongst innumerable other benefits, make up for the previous lack of a word-list with reverse alphabetisation for Old Norse.[11]

### ONP's word-list

About the word-list                    Instructions for searching                    Dansk version

Search string: hval          Exact spelling:  ☑
Search mode: component        Word-type:       headword
Word class:  sb.              Gramm. specif.:  unspecified

[ Find ]

*Search string:* hval        *Spelling:*        exact
*Search mode:* component     *Word-type:*       headword
*Word class:*  sb.           *Gramm. specif.:*  unspecified

Items found: 123

─────────────────────────────────────────────

**almenningshvalr** *sb. m.*

Citation slips: 1

─────────────────────────────────────────────

**andhvalr** *sb. m.*

Citation slips: 2

─────────────────────────────────────────────

*Fig. 1. An example of the net interface to ONP's word-list: the inquirer has asked for information about all the noun (sb.) headwords that include the element* hval*; this allows for all items ending in* -hvalr *(including of course* hvalr *itself) and all compounds containing* (-)hval- *or* (-)hvals- *or* (-)hvala-. "Exact spelling" (top right option) was requested to eliminate forms in* hvál(l).*

*The result is 123 items; at the bottom of the screen there is the beginning of the alphabetical list of items that each search generates.*

---

[11] An extremely useful list of this type, *Orðmyndir í stafrófsröð eftir niðurlagi*, for Modern Icelandic is to be found in *Íslensk orðtíðnibók* 1991:628–1066.

At the time of writing, the dictionary's entire database system is being converted from a DOS platform to a custom-designed Windows-based application of the editors' own making. The software that is necessary to provide internet access to ONP material has been installed, and there will be open access to the word-list on the world wide web (www.onp.hum.ku.dk) after the Easter recess, 2001.

How long it will take before a click on a headword can also provide a remote user with the underlying citations in the manner now practised by Orðabók Háskólans — that is a more difficult conundrum.

# 9   Sources

Degnbol, Helle, Guðrún Ása Grímsdóttir, Bent Chr. Jacobsen, Jette Knudsen, Eva Rode & Christopher Sanders. 1992. En ordbog er en database in Ruth Vatvedt Fjeld ed., *Nordiske studier i leksikografi: Rapport fra Konferanse om leksikografi i Norden 28–31 mai 1991*, Skrifter utgitt av nordisk forening for leksikografi, skrift nr. 1, pp. 385–389. Oslo.

Hødnebø, Finn (ed.). 1972. *Rettelser og Tillæg = Johan Fritzner: Ordbog over Det gamle norske Sprog* 4. Oslo, Bergen, Tromsø.

*Íslensk orðtíðnibók*. 1991. Orðabók Háskólans. Reykjavík.

Jakobsen, Alfred. 1970. Avledninger på *-leik* i nordisk, særlig norsk. *Maal og Minne*, pp. 104–124.

*Norrøn ordbok.* 1975. Leiv Heggstad, Finn Hødnebø, and Erik Simensen (ed.). Oslo.

Nøgle // Key = *Ordbog over det norrøne prosasprog. ONP 1: Nøgle // Key.* 1995. Den arnamagnæanske kommission. København.

*Ordbog over det norrøne prosasprog.* ONP 1: Nøgle // Key. 1995. Den arnamagnæanske kommission. København.

ONP 1: a-bam = *Ordbog over det norrøne prosasprog*. 1995. Helle Degnbol, Bent Chr. Jacobsen, Eva Rode, Christopher Sanders & Þorbjörg Helgadóttir (eds.). Den arnamagnæanske kommission. København.

Prøvehæfte // Prospectus = *Ordbog over det norrøne prosasprog. // A Dictionary of Old Norse Prose. Prøvehæfte // Prospectus.* 1983. Den arnamagnæanske kommission. København.

Registre // Indices = *Ordbog over det norrøne prosasprog. // A Dictionary of Old Norse Prose. Registre // Indices.* 1989. Den arnamagnæanske kommission. København.

Rode, Eva. 1991. Ordbog over det norrøne prosasprog. *Lexiconord: Leksikografi i Norden*, Nordisk Språksekrekretariat rapport 14, pp. 67–71. Oslo.

Sanders, Christopher. 1997. *Ordbog over det Norrøne Prosasprog* (A Dictionary of Old Norse Prose) in K.H. van Dalen-Oskam, K.A.C. Depuydt, W.J.J. Pijnenburg & T.H. Schoonheim, eds, *Dictionaries of Medieval Germanic Languages: A Survey of Current Lexicographical Projects*, International Medieval Research 2, pp. 81–88. Turnhout.

Walter, Ernst. 1972. Adjektivabstrakta auf *-leikr* (*leiki*) in lateinisch-altwestnordischen Übersetzungtexten. *Medieval Scandinavia* 5, pp. 34–41.