
Working on an Historical Dictionary: The Swedish Academy Dictionary Project

Per Stille (*per.stille@svenskaakademien.se*) and Bo-A. Wendt (*bo.wendt@svenskaakademien.se*), Dictionary Staff of the Swedish Academy, Lund, Sweden

Abstract: The Swedish Academy Dictionary is one of the world's largest dictionary projects. Work on it was started in 1884 and it will be completed by 2017. The dictionary describes the written standard language of Swedish from the beginning of the sixteenth century until the present. It is therefore both historical and contemporary. In this article the background and scope of this huge project are presented with a description of what an entry contains, and how it is constructed. The problems of describing a language simultaneously from both an historical and a contemporary aspect are also dealt with.

Keywords: CHRONOLOGY, CONTEMPORARY DICTIONARY, DIACHRONY, ENTRY, HISTORICAL DICTIONARY, LANGUAGE HISTORY, LEXICOGRAPHIC DESCRIPTION, STANDARD LANGUAGE, SYNCHRONY.

Opsomming: Om aan 'n historiese woordeboek te werk: Die woordeboekprojek van die Sweedse Akademie. Die woordeboek van die Sweedse Akademie is een van die wêreld se grootste woordeboekprojekte. In 1884 is begin om daaraan te werk en dit sal teen 2017 voltooi wees. Die woordeboek beskryf die geskrewe standaardtaal van Sweeds vanaf die begin van die sestiende eeu tot die hede. Dit is gevolglik sowel histories as kontemporêr. In hierdie artikel word die agtergrond en omvang van hierdie enorme projek aangebied met 'n beskrywing van wat 'n inskrywing bevat, en hoe dit saamgestel is. Die probleme om 'n taal gelyktydig van sowel 'n historiese as 'n hedendaagse gesigspunt te beskryf, word ook behandel.

Slutelwoorde: CHRONOLOGIE, HEDENDAAGSE WOORDEBOEK, DIACHRONIE, INSKRYWING, HISTORIESE WOORDEBOEK, TAALGESKIEDENIS, LEKSIKOGRAFIESE BESKRYWING, STANDAARDTAAL, SINCHRONIE.

Introduction

To write a combined historical and contemporary dictionary covering more than five centuries is an immense task. A large amount of material has to be handled. It is inevitable that a project of such huge scope will take a considerable time to complete. In order to keep the same format and standard throughout this whole period, thorough planning is essential. In this article the project of the Dictionary of the Swedish Language published by the Swedish Academy

(SAOB) is presented with an account of how it was established, what its guidelines are, and how the content of the entries is constructed.

The Swedish language

Swedish is spoken by about nine million people in Sweden and some 300 000 in Finland. It belongs to the Germanic family of the Indo-European languages and is closely related to the other Scandinavian languages, Danish, Norwegian, Icelandic and Faroese. A thousand years ago, these languages appear to have constituted one fairly uniform language. Swedes, Norwegians and Danes are still, with some effort, able to understand each other. In the Viking Age (A.D. 700–1000) Scandinavian had some impact on English. Nouns such as *birth, egg, fellow, husband, leg, sister*, adjectives such as *flat, ill, weak*, verbs such as *get, give, take, want*, and even a pronoun such as *they* (with *them* and *their*) are borrowings from Scandinavian. The importance of the Scandinavian languages today is more modest. A rare exception is the Swedish word *ombudsman*, which has found its way into legal terminology in many languages.

The oldest stages of Swedish are known only from runic inscriptions. Swedish literature developed only in the thirteenth century. At that time Latin was the main language of written texts and Swedish only slowly made its way into new domains. This situation also inspired the first attempt to create a Latin-Swedish dictionary in order to enable people to translate from Latin into Swedish or *vice versa*, the *Latinskt-svenskt glossarium efter Cod. Ups. C 20* from the fifteenth century.

It is customary to distinguish between Swedish before 1526, Old Swedish, and after 1526, Modern Swedish. The shift which of course was not so abrupt, had to do with several factors. In medieval times, there was no complete translation of the Bible in Swedish, but the Reformation resulted in a translation of the New Testament in 1526 and of the whole Bible in 1541, which had a great impact on the written language. Equally important was the centralized bureaucracy that was built up by King Gustavus Wasa. It used a modernized language, heavily influenced by Low German.

During the period of Early Modern Swedish (1526–1732), several bilingual or multilingual dictionaries were written to facilitate the learning of other languages, such as Latin, German, English and French. It also brought the first attempt at a descriptive monolingual dictionary of Swedish, *Glossarium-Sveo-Gothicum Eller Swensk-Ordabok* (1712), by the bishop Haqvin Spegel. It was a rather small dictionary with less than 10 000 entries. The eighteenth century thus saw much linguistic work in the fields of grammar and lexicography, but the need for a great standard dictionary was yet to be filled.

The Swedish Academy Dictionary

In 1786, the Swedish Academy was inaugurated by King Gustavus III. The main objective was to work for the "purity, vigour and majesty" of the Swedish

language, which in modern language would be its clarity, expressiveness and prestige. To attain this, the Academy was to produce a dictionary and a grammar (the grammar appeared in 1999). What the king had in mind was surely a normative dictionary such as *Le dictionnaire de l'Académie française* (1694). At first the Academy did not have the staff or economic resources for such a huge undertaking, so in spite of several attempts, nothing really substantial could be achieved. Not until 1883 was a feasible plan drawn up and the work begun. Three other large dictionary projects served as model: *A New English Dictionary* (reissued as *The Oxford English Dictionary*), *Deutsches Wörterbuch* and *Woordenboek der Nederlandsche Taal*. They all are a combination of an historical and a contemporary dictionary and illustrate the definitions with plenty of authentic quotations.

The dictionary, *Ordbok över svenska språket utgiven av Svenska Akademien* (SAOB), or the Swedish Academy Dictionary of the Swedish Language, covers the period of Modern Swedish. It does not try to cover all periods of the language as *The Oxford English Dictionary* does, because there is already a dictionary of Old Swedish, *Ordbok över svenska medeltids-språket*, by K. F. Söderwall (1884–1918, suppl. 1925–73). The scope is more or less limited to the written standard language. This means that spoken language, dialects and technolects are included only to a limited extent.

The work started with the collecting of a large corpus of authentic quotations from all kinds of Swedish literature. A large part of this work was done before the editorial work began, but the collecting continued until last year. Then it was decided that the staff should discontinue the collecting of quotations in order to accelerate the work. The text of the SAOB is based on these roughly eight million quotations selected from more than 20 000 titles (Hast 1983: 160, Lundbladh 1996: 94). During compilation work the editors also consult existing dictionaries and encyclopedias. The present editors furthermore take advantage of modern computer-based corpora, mainly from newspapers and novels, to describe modern language use (cf. Stille and Wendt 2000). Since 1991, all text production within the project has been computerized.

From the start of the project the editorial staff has resided in the university town of Lund in southern Sweden. There are about twenty staff members working on the project.

The first volume of the dictionary appeared in 1898, and the latest (vol. 32, *stå an-svåpa*) in 1999. Seven, or perhaps eight, volumes still remain. The Swedish Academy has decided that the dictionary must be completed by the year 2017. It will then probably comprise nearly 640 000 lemmas, and will be one of the largest dictionaries in any language (cf. Allén et al. 1996: 8).

The content of an entry

The SAOB describes the vocabulary of written Swedish from 1521 until the present — up to the time of the editing of an article. The lexicographic description emphasizes the semantics of a word including shifts in meaning, but the

account of other aspects of word history is also detailed and precise. It is, however, much more condensed. At the beginning of an entry (the "head") various categories are dealt with in a comparatively small space, whilst the semantic description is allowed to take up much more space, not least because of the illustrative text samples. Initially the following are given: by-forms (if there are any), pronunciation, flexion, variation of word forms and spelling, and finally etymology.

That pronunciation is given in a dictionary of the written language only is something of an anomaly. It is due to the fact that at the start of the project the first editors did not systematically exclude spoken language as object of description — it only became the natural consequence of the circumstance that the empirical material only consisted of written texts, which in the course of time was turned into an explicit principle. It is the medium itself that is the conclusive criterion. There is no (further) stylistic restriction beyond the exclusion of spoken language — no matter how commonplace or colloquial an utterance in a novel may be, in the view of the SAOB it is still considered as written language, i.e. language recorded in writing.

As far as the etymology of Indo-European words, whether inherited words or loanwords from other Indo-European languages, is concerned, it is the object of the SAOB to trace them back to their original stage, although no reconstructed forms are normally given. For non-Indo-European words the SAOB is content with stating the words in the language from which they originate.

With these quite concisely described categories the initial part of the entry is complete, and the semantic description begins. If there are different sections, these are arranged etymologically — as far as the course of the semantic development is known or possible to reconstruct. The main sections are numbered 1, 2, 3 etc., and the first always deals with the sense that equals or comes closest to the etymologically original one. Here the SAOB maintains the principle that this should apply to the whole history of the word, even if it is a loanword in Swedish with the original sense recorded later than another one having entered the language at an earlier stage. In some, although rare, cases section 1 can therefore deal with a sense not recorded until recently, while the oldest sense from a purely Swedish perspective is given in a later section. This is an heritage of the comparative and neogrammarian all-Indo-European view of linguistic material. In practice, however, it very seldom has the consequences sketched above — and moreover, the semantic chronology of the word in Swedish is clear enough from the dates of the text samples. In every section the oldest attested example of the sense (within the SAOB material) is given as the first of the text samples. If the present-day usage is restricted, the most recently attested sample is also given (as the last one).

At the end of the entry, after the completed semantic description, follows an extensive list of compounds in which the word in question forms the initial component, as well as of derivatives of the word. In an article about a verb, there is also a section of what in the SAOB is called "separate combinations"

(Sw. *särskilda förbindelser*), which are roughly to be understood as verb-particle collocations, e.g. *springa före* 'run ahead', *komma undan* 'get away'. These compounds, derivatives and "separate combinations" are in a way little entries in their own right, although subordinate and within a bigger entry, but they are very summary and include (mostly) only a semantic description, although derivatives are described somewhat more extensively. Flexion and other kinds of form variation of the latter component of compounds are to be found in the entry for the simplex word. As Swedish, like most other Germanic languages, admits the formation of compounds to an almost endless degree, the number of compounds can sometimes be quite large, so that this part of the entry occupies far more space than does the account of the simplex word. Yet, the SAOB, at least in its current editorial practice, records only the well-established and more or less lexicalized compounds; the occasional compounds (which can be formed in almost indefinite numbers as the need arises) are normally not accounted for, although they may be attested in the SAOB material.

Hierarchic levels in the semantic description

As already mentioned, the main sections are numbered from 1 onwards. Sometimes, if there are many of these, they are semantically grouped under **A**, **B**, etc. If necessary, a superordinate level of **I**, **II** etc. is also introduced, often reflecting a more grammatical division, e.g. into separate parts of speech or into transitive and intransitive uses of a verb. More extensive articles are often subdivided into an intricate system of subordinate sections below the levels of 1, 2 etc. (labelled *a*, *b* etc.), sub-subordinate sections (*α*, *β* etc.) and even sub-sub-subordinate sections (*α'*, *β'* etc.). (When there is only one subordinate section, no introductory letter is used.) A fundamentally important difference in the status of the subordinate sections is whether they form an exhaustive account of subordinate senses or whether they are mere examples of certain (e.g. especially frequent, peculiar or lexicalized) instances of a superordinate sense. In the former case the semantic description of the main section (or superordinate subordinate section) is not followed by any text samples; instead, the subordinate sections appear immediately after the superior definition.

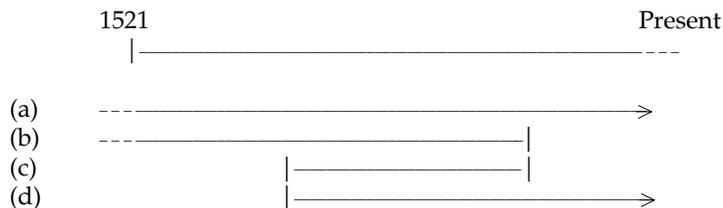
If the subordinate sections are only special instances of the superordinate sense, the definition of this sense (with its own text samples) is followed by "especially" (Sw. *särskilt*, abbr. *särsk.*). This indicates a sense which for some reason or other the editor has chosen to distinguish — it may be because it is a very frequent, perhaps even the most frequent sense, but it may equally be because it is the most unusual (and perhaps equally rare) one. The meaning of *särsk.* in the SAOB thus roughly equals "among other things". (This SAOB use of *särsk.* may be compared to those cases where *The Oxford English Dictionary* labels its first subordinate section with a *b* instead of an *a*.) Normally a superior section that ends with *särsk.* has text samples of its own, but sometimes its definition may be followed immediately by *särsk.*, whereby the editor has chosen to ignore other (attested or just theoretically imaginable) usages than those ac-

counted for in the subordinate sections. This *särsk.* might as a matter of fact be used almost anywhere in the descriptive language, not least within definitions, denoting in the same way that something has been mentioned separately without any assertion about its frequency.

Both an historical and a contemporary dictionary

Because the SAOB is simultaneously an historical and a contemporary dictionary, it always has to ascribe usage restrictions (from a modern point of view) to each of its lexicographic categories. Naturally there may be stylistic, social or regional restrictions in the use of a form or a sense. It can also be a matter of the degree of lexicalization (not least for compounds which, although attested in earlier language use, may at present not exist otherwise than as occasionally formed). Most importantly, however, the SAOB is concerned with usage — whether a form or a sense is in current use or whether it is more or less obsolete (within a three-graded scale: less used (Sw. *mindre bruklig*), seldom used (Sw. *föga bruklig*) and unused/obsolete (indicated by †). The lack of a label of this kind is then to be understood as: commonly used (at the time of publication).

That the SAOB in a way has a double mission, at the same time describing an historical development and the present-day language, has been extensively discussed by the former editor-in-chief Hans Jonsson (1983), and the following exposition is partly based on his views. Within the description span of the SAOB (1521–today) the following different extensions of a certain sense or usage (or for that matter form) are possible:



A certain sense of a word may have existed from the very beginning of the SAOB time span, that is to say, it may be attested already in the medieval period, in Old Swedish, and be current till this day. (Even for a sense of undisputable Old Swedish origin the SAOB gives the first record in its own — Modern Swedish — material, from whatever year it comes. That the word already existed in Old Swedish can only be seen in the etymological parenthesis, where only less exact specifications of the different senses are given.) In this case, as in (a) above, we therefore have a sense that is (from the point of view of the description period) original and current. A sense could likewise be older than 1521 but could have become unused during the SAOB period: original, but at present obsolete, as in (b) above. The sense may also have appeared for the first time — through spontaneous sense development or foreign impact — after

1521, whether early in the sixteenth century or quite recently, perhaps just a few years before the time of publication, and thereafter has either disappeared again or is still in use: recent and obsolete as in (c) or recent and current as in (d) above.

Synchronic and diachronic perspectives

When it is a question of distinct senses, these naturally appear in separate main sections, each with its own usage label. In principle this account gives a purely synchronic description in separate layers: a nowadays obsolete sense **1** is described in its own right, assigned to an older era of the language (with the boundaries given (or more accurately: indicated) by the oldest and the most recent text sample), then a still current sense **2** (whether original or new) is described in its own right and assigned to the prevailing state of the language (with the past boundary at (or indicated by) the oldest text sample). The senses separated into different sections are thus synchronic descriptions of one time period each, although usually partly overlapping each other. The development from sense **1** to sense **2**, the diachronic dimension, is often only implicit. Sometimes, however, it can be made explicit through information in an introductory etymological parenthesis in the latter section, e.g. "**2** [properly metaphoric use of **1**] ...". On rather rare occasions, it can be part of the semantic description proper, e.g. "**2** in metaphoric use of **1**, about ...". In general, however, the historic and the present-day dictionary are in these cases combined by merging synchronically described layers, in so far as they are not common for the objects of both dictionaries.

The combination becomes more complicated when the changes of usage concern nuances of a sense which are hardly divisible into separate sections — when an historic state which is somewhat differentiated from the present-day state is described in the same section as the latter. This opposition within one and the same section can be handled in two fundamentally different ways: either with an introductory usage label or with an indication of restricted usage later in the definition text. In both cases there is a diachronic dimension in the manner of description, but most evidently so in the latter.

In the first case an explicit usage label may be lacking (equal to full usage), and then older, somewhat extended or special uses of the sense are accounted for after the formula "earlier also". Here the synchronic point of view is still dominating: a still prevailing state is described as it is, and then a more peripheral addition is also made to cover a somewhat wider usage that existed in earlier times. The description becomes more truly diachronic when restricted use is introduced by specifying explicit exceptions of common use in certain detached subordinate sections, i.e. of the type "**1** († except in *b-d*)", then followed by the general sense which existed earlier in the language but is nowadays only found in those sense nuances, usages with lexical or syntactic restrictions or more or less lexicalized phrases that are listed in *b*, *c* and *d* (which in a case such as this are all introduced with the label "commonly

used"). These isolated present-day usages of a word are thereby anchored in an older general and once unrestrictedly productive sense. They are put into a diachronic context within which they are present-day remainders of an earlier state.

The diachronic perspective becomes even more evident when the present-day restrictions are given only further on in the section, as an integrated part of the definition text. Without any introductory usage label an older and more general semantic description is given, then followed by "nowadays only" and the restriction that applies to this sense of the word in present-day language, whether concerning sense nuances, types of referents, lexical or syntactic contexts or something else. Thus, in this case, there is an exception to the principle that sections in the SAOB without any introductory usage label implicitly indicate unrestricted present-day usage; the whole of the definition has to be read to be certain of that. In this case, the definition itself includes the sense development that has taken place: from a more general former state to a more restricted modern state: "this was the meaning of the word originally (within the SAOB period of description), but nowadays it only applies to the following cases".

Since every section has only one collection of text samples in strict chronological order, the reader, on the other hand — in the latter case — is left without any answer as to when the change(s) took place. The most recently attested sample that shows the older sense nuance or usage is not necessarily the very youngest of those in the total SAOB material. Thus, when the descriptive language of the SAOB is the most diachronic — as we maintain it is when an only gradually given restriction is explicitly introduced within the definition — the diachrony is the least unambiguous as far as text sample chronology is concerned. Instead this chronology is most unambiguously presented when the sense development is stratified into each (partial) sense described synchronically by itself. Compare e.g. section 6 with sections 7 and 8 of SIMPEL (the same word as Eng. *simple*) (in translation):

- 6) [...] as an adverb: simply [...]; nowadays only in the phrase *helt (enkelt och) simpelt*, quite simply [...]. [Text samples in chronological order, the last ones displaying the only phrase in use nowadays.] [...]
- 7) (†) straightforward, naive, unspoiled [...]. [Text samples from 1526 to 1885.]
- 8) plain, ordinary [...]. [Text samples from 1620 to 1967 (the year of publication).]

Which of the solutions sketched above is preferred in the individual case is dependent on what the empirical material actually looks like. The editor often has to accept that the most sensible and for that matter most practical and convenient solution for a certain word, does not do full justice to every aspect of the complex web of sense development.

Concluding remarks

In the SAOB — without comparison the largest project within the humanities

in Sweden (cf. Jonsson 1993) — a large number of text samples have been dealt with since the work began, and, as we hope to have shown, also a great mass of information is given in the 32 volumes published, available for those who can assimilate the often very concentrated presentation. Since its start in 1884, more than four generations of editors have been working on the project and a great deal of work still remains, but the present editors will probably see the project completed.

Postscript

The version of the dictionary on Internet can be visited at: <http://g3.spraakdata.gu.se/saob/konk/>. The modern corpora of Swedish can also be consulted on Internet at: <http://spraakbanken.gu.se/>. More information about the Swedish Academy can be found at: <http://www.svenskaakademien.se/>.

References

A. Dictionaries

- Deutsches Wörterbuch* by Jacob Grimm and Wilhelm Grimm. Leipzig 1852-1960.
Le Dictionnaire de l'Académie française dédié au Roy. Paris 1694.
Glossarium-Sveo-Gothicum Eller Swensk-Ordabok by Haqvin Spegel. Lund 1712.
Latinskt-svenskt glossarium efter Cod. Ups. C 20 ed. by Erik Neuman. Uppsala 1918-42.
A New English Dictionary on Historical Principles. Oxford 1883-1928.
Ordbok över svenska medeltids-språket by K.F. Söderwall. Lund 1884-1918.
Ordbok över svenska språket utgiven av Svenska Akademien (SAOB). Lund 1893-.
The Oxford English Dictionary. Oxford 1933-1986.
Woordenboek der Nederlandsche Taal. s'Gravenhage 1852-1998.

B. Other sources

- Allén, Sture, Yvonne Cederholm, Sofie Johansson Kokkinakis, Lena Rogström, Rudolf Rydstedt and Lars Svensson. 1996. *Om svar anhålles. Rapport från projektet OSA*. Göteborg: Department of Swedish, Göteborg University.
- Hast, Sture. 1983. Om SAOB:s material. *Nysvenska studier* 63: 159-192.
- Jonsson, Hans. 1983. Om historisk lexikografi typ SAOB. *Nysvenska studier* 63: 127-158
- Jonsson, Hans. 1993. Sveriges största humanistiska projekt. *Kungl. Vitterhets-, Historie- och Antikvitetsakademiens årsbok* 1993: 119-128.
- Lundbladh, Carl-Erik. 1996. Värdering av SAOB:s korpus och språkprovssamling. *LexicoNordica* 3: 91-103.
- Stille, Per and Bo-A. Wendt. 2000. Användning av Språkdatas konkordanser i det redaktionella arbetet med SAOB. Byrman, Gunilla, Hans Lindquist and Magnus Levin (Eds.). *Korpusar i forskning och undervisning. Rapport från ASLA:s höstsymposium i Växjö, 11-12 november 1999*: 266-272. Uppsala: ASLA.