

Anki Mattisson

The Swedish Academy Dictionary – from A to Ö in 135 years

The offices of the Swedish Academy's editorial staff are located in the center of Lund, enclosed by a park. The academy acquired the building in 1998. Before that time, the editorial staff had leased office space from the Lund University Library. Having a dedicated editorial building with a garden might seem like a small matter, but I argue that the environmental conditions for editorial work should not be overlooked. The editorial staff consists of 24 people. Like other dictionary editorial staffs, we have some difficulty with the recruitment of editors who are well-versed in language history; the few persons who possess the relevant competence hold research positions at the universities. Our responsible authority, the Swedish Academy, has its headquarters in Stockholm, roughly 650 km from Lund. The work on the dictionary is financed by a donation from King Gustaf III, which is used for language projects. King Gustaf III established the Swedish Academy in 1786.

The Swedish Academy Dictionary is a historical dictionary but also a so-called diachronic contemporary dictionary. The reason for the latter description is that there is no final date for the material. Each volume contains linguistic material from 1521 up until the year of print, which means that the first volume covers the period between 1521 and 1897. Hence, volume 38 will cover the period between 1521 and 2017, a span of 120 years longer than that of the first volume. Nowadays, the Swedish Academy Dictionary, or SAOB, is a descriptive dictionary, as opposed to a normative one. It describes Swedish language usage in written material. In the beginning it was said also to provide an overview of the speech of educated Swedes. The previous, normative aspect was realized through comments such as “formal language”, and “colloquial language”.

The collection of materials was initiated in 1885, following careful planning of which sources were to be excerpted. In the spirit of the times, the material was of utmost importance, and the greater the quantity, the better. The initiative was taken by an Academy commissioner, who was assisted in the planning by K. F. Söderwall, professor of Scandinavian languages and renowned for his dictionary of medieval Swedish. Söderwall's dictionary covers old Swedish up to the 1520s, and the SAOB modern Swedish from 1521. Around 50 people were hired for excerpting, mostly priests, teachers, and senior citizens with good general knowledge. The most essential early source material came from bible translations from 1526 and 1541. King Gustaf I, also known as Gustav Vasa, left an archive behind, which was published early on. It covers the period between 1521 and 1560, and is invaluable for use in a historical dictionary. The archives of Gustav Vasa comprise 29 printed volumes.

Our list of references include around 21 000 sources, one of which may consist of, for example, editions of a newspaper spanning 150 years, a collection of materials like the archives of Gustav Vasa, or a single book. The fact that the planning of the dictionary took place in a different period of time than the finalizing of the volumes has had numerous consequences. Though the work on the dictionary was certainly non-political, the source material reveals that some of the issues that characterized the end of the 19th century, and the beginning of the 20th, have been given more space than is justifiable. One such issue is that of

national defense and an industrialization aimed at weapons. Sweden and Norway were in union with one another but the union was on the verge of breakdown, and at the same time the Swedes were concerned about the “Russification” of Finland. War history, military sources and magazines, as well as weapons’ literature make up a disproportionately great part of the material. Nevertheless, the issue of voting rights, in particular women’s suffrage, is nowhere to be found in the material. One explanation for this may possibly be the fact that there were only male editors among the editorial staff between 1884 and 1980.

At present, our complete card index contains over 9 million excerpts, or slips, in boxes of around 400. We are currently working on the letter V, but as of yet, we have not finished the printing of the letter U. After V, the letters W, X, Y and Z remain (consisting of a very modest sample), as well as Å, Ä and Ö. This year, we will publish volume 36, leaving two volumes for print. Our deadline is in 2017.

The material is still in the form of quotation slips, though we also use a large newspaper corpus and a few smaller concordances. The editor is presented with approximately 3 000 to 5 000 slips for editing at a time, and there is a (fairly) stable relation between the number of slips and the completed printed columns. The editor writes the text in a word processing program, using a template which is typographically designed like a printed page. The number of language samples per section is limited due to restraints on space and time. We estimate that barely 10 % of our material will be present in the print.

Since about 75 years ago, our work organization has consisted of one group in charge of preparing the material, editors for the actual editorial work, and one group for checking the quotations. A small administrative unit handles staff concerns, the financials, and make sure that the scripts are correct before they are sent off to be printed. In later years, an IT-administrator has joined the organization, with the specific task of working towards a user-friendly and correct website, as well as making sure that the editorial staff is provided with an editing system which is useful for updates. In my opinion, the editing of a dictionary requires teamwork, and our organization remains highly functional as long as the dictionary is unfinished. The organization will probably undergo drastic changes following its completion in 2017.

At the completion of the SAOB in about five years, the dictionary will be divisible into five parts: volumes 1–7, 8–16, 17–23, 24–32, and the final six volumes. According to the original plan, the SAOB was to be completed in approximately 25 years, but 20 years after the publication of the first volume, only three volumes had gone to print. In the first seven volumes, derivations and some compounds were entered as headwords, since the structure of the SAOB was almost exclusively alphabetical. From the volume 8, compounds and derivations were entered alphabetically under the headword (in accordance with the so-called affinity principle). Currently, the SAOB is a niching dictionary, which to a certain extent will cause problems with its conversion to a well-functioning digital product.

The five parts coincide, to some degree, with different stages of the dictionary’s life. These stages, in turn, are related to one or several supervisors. Ebbe Tuneld, who headed the dictionary in 1920, made sure that the work was streamlined, e.g., by excluding unusual jargon and limiting the number of language samples. The work progressed rapidly and in 1939 Swedish newspapers proclaimed: the dictionary of the Academy is half-way to completion – it should be finished by 1980. Yet, this would not be the case. At the 200-year anniversary of the Academy, in 1986, the dictionary had reached the entry for SPÅNTA. The

letter S, which is the most extensive letter of the Swedish alphabet, took 50 years to edit, and the final part of S was printed in 2002. The letters T–Ö should be finished in 20 years.

The SAOB, like the *Deutsches Wörterbuch* and the OED, is characterized by the ideals and principles of Jungian grammar. The historical perspective is firmly established in the spirit of the times: the key to understanding the language of today lies in the study of the language of yesterday. The development of meaning was essential. The OED like SAOB presents spelling and grammatical forms very carefully. The amount of material is substantial, and orthographic principles were not present in Swedish until the end of the 19th century. At the opening of the 16th century, the earlier system of paradigms in the Swedish language was being phased out – the remains are meticulously accounted for in the SAOB. Work on the actual presentation of forms can potentially take much longer than the editing of an entry. In the discussion of the upcoming online version, this information is part of what users should be able to hide with a click of the mouse.

While the headword remains visually and actually unchanged, the structure of the entries has altered in many other ways. Nowadays, all language samples are listed in chronological order following the definition.

The hierarchal structure has also changed throughout the years. The supply of section marks, which have in fact remained the same, contains five levels: I, II, III etc.; 1, 2, 3 etc.; a, b, c, etc.; α , β , γ etc.; α' , β' , γ' etc. However, they have been used differently and to varying degrees. One of my co-workers has studied the frequencies of γ' in the lowest level of the semantic hierarchy, in entries written in two periods of 20 years. It appears that this level of the structure was made use of to a greater extent under the letter S, in volumes 24–32, than in other volumes. The increase in relation to previous volumes was nearly 100 %. Current editorial guidelines advocate restriction on the issue of hierarchal subdivision.

Use markers and style labels have also undergone changes throughout the years. Unfortunately, in some respects, the OED system for use and style labeling did not constitute a template for the SAOB. Rather than allowing the general vocabulary to remain unmarked, and just label the archaic, the uncommon, the colloquial or any style or jargon, the SAOB contains about 1 300 variants of originally 30 use markers and style labels and many of them vary during the very long time of production.

In the 1980s, OCR-scanning of the SAOB was commenced at Språkdata in Gothenburg. The primary aim for the scanning was to improve access to the material for scholars of the Swedish language. The dictionary text, with frequent changes between fonts, proved to be too complex for the technology of the time, and the results were intermittently very unsatisfying. Proof-reading cost large sums of money, but was never completed. Despite all its shortages, the SAOB database was published online for free use in 1997. The database is very well tagged in some cases, but not so in other. Due to failings in the original scanning of the material, the answer to an inquiry from a user may be unintelligible, or the text may be difficult to read, with consistently strong or bold letters. The SAOB online is currently better than nothing, but it should not be used for quotation, especially not in a scientific or scholarly text. The website has many visitors: around 1 million per year.

The website of a dictionary is the front of today's editorial staff and publisher. Since there is a major interest in the online dictionary, the material which is publicly accessible should be accurate. In this case we have been heard by the Academy, and there is currently a re-

digitizing project of the SAOB in progress in China, through cooperation with the Kompetenzzentrum in Trier. The typographically challenging text poses some problems for the Chinese agency as well, but we maintain faith and hope to have the new text delivered next year. When the files from Trier reach us, they are reviewed and processed by hand into valid XML by the editorial staff.

The re-digitizing will be used for a new online version with an updated search function. We are currently in contact with a Norwegian company with ties to Microsoft, which seems to have a sufficiently powerful system. Much like other large, old national works (I assume), each volume must be thoroughly reviewed, and each structural element of significance must be found and tagged. The project aim is to create the right set of conditions for establishing a new website for the presentation of the SAOB, and hopefully deliver data in a form that enables the editing staff to find a system of editing that can be adapted according to our needs. In accordance with our time schedule, we expect the re-digitizing in China to be completed during 2013, and the search functions during 2014. In 2015, we hope to launch a website with a layout that is appropriate for linguists as well as the general public.

This workshop concerns topical issues for us, being in a phase where our work is not yet completed and at the same time planning for the future. These issues are:

- How to update a work that has been in progress for nearly 125 years when the work is uneven in quality, and has a different structure depending on which part of the work is referenced.
- How to make a scientifically and scholarly weighty dictionary useful online without a major loss of information.

For those of you who have already been through the stages of the work-process that we are facing, we welcome all forms of feedback from you. We have no desire to re-invent the wheel.