Sven-Göran Malmgren, University of Göteborg

From Svensk ordbok ('A dictionary of Swedish') to Nationalencyklopediens ordbok ('The Dictionary of the National Encyclopedia')

ABSTRACT: The paper deals with two defining dictionaries of modern Swedish, "Svensk ordbok", "SOB" ('A dictionary of Swedish'), and "Nationalencyklopediens ordbok", "NEO" ('The dictionary of the National Encyclopedia'). The former was published in 1986; the latter, to a large extent based on the former, will be published in 1995. – As for the "SOB", the semantic description and the examples are accounted for. The most important improvements of the "NEO", compared to the "SOB", concern information on: 1) pronunciation, 2) etymology, 3) word formation, 4) synonymy, hyponymy, etc.

In 1986, a defining dictionary of Swedish, "Svensk ordbok", "SOB"('A dictionary of Swedish'), was published. This dictionary, comprising almost 60,000 lemmas, was the biggest monolingual Swedish dictionary since the 1950's, and, with regard to semantic descriptions, the most ambitious ever to appear in Sweden, except for the great historical dictionary, "Ordbok över svenska språket utgiven av Svenska Akademien", the "SAOB" ('A dictionary of Swedish edited by the Swedish Academy'). However, partly due to lack of time, not all the original ambitions were realized. In 1990, possibilities were created for a continuation of the work, when the publishing house producing the new Swedish national encyclopedia, Bra Böcker ('Good Books') decided to support a project aiming at publishing a dictionary ("Nationalencyklopediens ordbok", "NEO", 'The Dictionary of the National Encyclopedia') alongside of the encyclopedia. The director of this new project, as well as of the original one, is Professor Sture Allén. The "SOB" dictionary was produced at the Department of Computational linguistics, University of Gothenburg. Geographically, the "NEO" dictionary will be produced at the same place, but the former Department of Computational Linguistics is now part of the Department of Swedish Language. The "NEO" dictionary will be published in 1995.

In this paper, I will give a brief presentation of the "SOB", including a few words on its reception. (For a fuller account, see Malmgren 1988 and Gellerstam & Malmgren, forthcoming.) I will then proceed to the most important characteristics of the "NEO" dictionary, especially in relation to the SOB. In this connection, some general theoretical problems will be discussed.

1. The SOB dictionary and its reception

1.1. Macro-structure

The establishment of lexical units (lemmas), i.e., roughly, the solution of the polysemy-homonymy problem, in the "SOB", is based on the so-called lemma-lexeme model (see Allén 1981 for details; a LEXEME is roughly, in our terminology, a numbered sense in the dictionary). Basically, this implies that no two lemmas belonging to the same part of speech can have the same spelling, pronunciation and inflection. For instance, the ambiguous word FJÄLL ('high mountain'; 'scale (of a fish)') is treated as *one* lemma, although the two senses are totally unrelated, etymologically as well as psychologically. The reason is that the sense difference is not accompanied by any difference in inflection or pronunciation. The advantage of this principle is, of course, that here is no need for subjective judgements on the lemma level. – Another consequence of the lemma-lexeme model is that the dictionary is strictly alphabetically ordered. There is no grouping of morphologically related lemmas; no lemma has a higher priority than any other lemma.

1.2. Micro-structure

As already mentioned, the ambitions concerning the semantic descriptions of the "SOB" were high. Basically, every word should be provided with a true (Aristotelian, if possible) definition; only in exceptional cases was it considered legitimate to give mere synonyms. Also, the requirement of syntactic substitutability of the definitions was, in principle, imperative. For instance, transitive verbs had to be defined by verb phrases with transitive verbs as kernels, the objects not being included in the definition proper. These requirements, of course, sometimes led to definitions that might seem a bit, perhaps unnecessarily, complicated to the layman. A very important principle was the careful separation of different senses and subsenses. Basically, every sense, and every subsense, was provided with a semantic description, normally followed by an example. The relations between kernel senses and subsenses were often accounted for by means of meta-lexicographical terms like ÖVERFÖRT ('metaphorically'), UTVIDGAT ('extended'), and so on.

There are about 90,000 syntactic examples in the "SOB". Most of them are authentic or 'semi-authentic', i.e., they are quotations or modified quotations from the corpora collected by our department. The dictionary also provides morphological examples, even compounds with the defined word at the end.

Let us now take a look at an article in the "SOB" dictionary: TAK ('roof; ceiling', or rather 'roof+ceiling'); see table 1. The entry starts out with information on word-class membership and inflection. The main definition immediately after the raised dot '(inner or outer side of) the uppermost, delimiting part of a building or a room (or similar space, e.g. a car)' is followed by several morphological and syntactic examples. A small square after the last syntactic example (on line 11) indicates a slightly extended sense. The second square, some lines further down, indicates a slightly symbolic usage: the ceiling (tak) may represent the whole house. The next square ("äv. överfört") indicates a clearly figurative usage ('upper limit'). Finally, idioms containing TAK are accounted for (after

the asterisk). As is clearly seen, every subsense is illustrated by syntactic, and sometimes morphological, examples. Basically, every entry in the "SOB" has this "cyclic" structure.

Table 1.
The entry TAK
in the "SOB"

tak subst. ~*et* = * (insida eller utsida av) den översta, skyddande delen av byggnad eller rum el. annat utrymme där man kan vistas, t. ex. fordon. *takbeivsning; takfönster; taklampa; taklucka; takmålning; takteget; innertak; plåttak; sadeltak; yttertak.* i ~*et* hängde en praktfull ljuskrona: *de låg på ~et och solade.* *snön ligger vit på ~en; husen har branta ~; bilen hamnade på ~et* ≥ ₁ tekn. sammanhang av. innefattande delar i takkonstruktionen: *ett ~ består av taklag och taktäckning* ≥

ofta som symbol för hela byggnaden: *samla alla institutioner under ett ~: få ~ över huvudet* ≥ av. överfört övre begränsning särsk. av utgifter o. d.: *lånetak: pristak: chefen ville sätta ett ~ både för löner och frånvaro* ☞ *det är högt i ~ (i partiet)* även avvikande åsikter får komma till uttryck (i partiet); *glädjen stod högt i ~* stämningen var glad och uppsluppen: *ingen fara på ~et* ingenting (har hänt) som ger anledning till oro: *kors i ~et!* fantastiskt! uttr. för häpen beundran e. d.; *slå klackarna i ~et* festa under uppsluppna former

1.3. On the reception of the "SOB"

Written responses to the "SOB" (reviews, letters to the staff, and so on) have been systematically collected. Generally, reactions from linguists were positive, and newspaper/journalist reactions were relatively positive – admittedly, with two exceptions, concerning, on the one hand, misprints, grammatical errors and the like, and, on the other hand, factual errors, especially in the articles on technical terms. (Incidentally, the former kind of criticism is similar to that raised against the "Duden" dictionary by Timothy Buck (Buck 1984).) Reactions from private users were on the whole positive, too. Some frequent points in the letters to the staff involve suggestions concerning new words, and questions whether there are plans to give etymologies in the next edition. As will be clear, all these suggestions and reactions, positive as well as negative, will prove extremely useful in the work on the "NEO" dictionary.

2. Some characteristics of the "NEO" dictionary

All the principles of the "SOB" accounted for so far will be adhered to in the "NEO". Apart from additions of new words and careful checking of the definitions of technical terms, the most important improvements in the "NEO", compared to the "SOB", will concern information on:

1. Pronunciation,

2. Etymology,

3. Word formation,

4. Synonyms, hyponyms, and so on.

Each of these points entails interesting problems, also from a general metalexicographic point of view. I will comment briefly on all of them.

2.1. Pronunciation

The strictly alphabetical order principle implies, among other things, that there is no grouping of morphologically related lemmas in the "SOB". For instance, words like ACCEPT ('accept(ation)'), ACCEPTABEL ('acceptable'), and ACCEPTERA ('to accept') are all treated as lemmas in their own right. There is a pronunciation difficulty common to these words: the first syllable should be pronounced [aks-]. (Many Swedes have difficulties in pronouncing these words.) Now, in older Swedish lexicography, it was customary to group morphologically related words together, and to mark the groups typographically (only the basic lemma was placed at the beginning of a new line). In *THIS* case, it was natural to give the pronunciation of only the basic word. Unfortunately, the "SOB" adopted this principle, thus demanding from a user looking for information about the pronunciation of the word ACCEPTERA (a common word) that he consult the entry ACCEPT (an unusual word). In the "NEO", pronunciation will be provided at EVERY word that may be difficult to pronounce.

Swedish is one of the few European languages (Norwegian is another) that have a musical accent system. The accent of one-syllable words (and many longer words, too) is usually called ACCENT 1 ('). The accent of most compounds, and many other words with more than one syllable, is consequently called ACCENT 2 (`). There are some hundred minimal pairs of this kind in Swedish, e.g. AN`DEN ('the spirit') and AN'DEN ('the duck'). For a foreigner, it is possible to speak Swedish without mastering the difference between accent 1 and accent 2, and there are dialects without accent 2. However, since the difference is phonematic in Standard Swedish, it is still somewhat surprising that many monolingual dictionaries of Swedish do not give this information, or give it only incompletely. It was, e.g., not given in the "SOB" (perhaps it is even more surprising that not a single critic pointed out that it was missing), but it will be given in the "NEO". Moreover, since it is not always possible to predict the accent of inflected word forms from the accent of the basic form, this information will be given, too, whenever necessary. For instance, compare the pronunciation of FIL`MER ('films', from FILM) with that of GRE'KER ('Greeks', from GREK). As far as I know, this information is not given in any Swedish, or Norwegian, dictionary, in any case not in any general dictionary. It seems, incidentally, that many dictionaries tend to neglect the simple fact that inflected forms are no less important to the user than the 'basic forms'.

2.2. Etymology

As hinted earlier, there is a great popular interest in etymologies in Sweden. In the "NEO", etymologies will be given in a non-technical way, similar to that of the "Duden" and "Robert" dictionaries. Moreover, like, e.g., the "Robert" dictionary, the "NEO" will give the year of the first known occurrence of every (numbered) sense of every lemma, if taking place after 1520 (otherwise, the information 'before 1520' will be given). This chronological information will be possible to give mainly thanks to the existence of the "SAOB" (see introductory paragraph) and the collections associated with this dictionary.

Now, there are some interesting theoretical problems in connection with the etymo-logical, and chronological, information. First, there is a slight conflict – theoretically somewhat embarrassing perhaps, but practically not very important – between the lemma-lexeme model, which is non-etymological, and the arrangement of the etymo-logical information. The conflict will arise in any dictionary that does not establish its lemmas strictly according to etymology, i.e., probably in most modern dictionaries. Thus, in a number of cases, more than one etymology will have to be given within a lemma (say, one etymology for lexemes 1 to 4, and another for lexemes 5 to 6). This may be a cosmetic drawback, but hardly much more than that. A related problem, much more interesting but not really of practical importance is the following: what is a lemma, from a diachroni-cal point of view? How can we say that a certain word in 20th century Swedish is 'the same' as a similar word in, say, 10th century Swedish?

The second problem concerns the identification of given modern senses with similar older ones. More specifically: if we are going to answer the question when a certain (modern) sense of a word appeared for the first time in Swedish, then we must decide what is 'the same' sense. In practice, we will certainly be forced to accept some degree of latitude. For instance, the word SAMHÄLLSKUNSKAP ('knowledge about society'), in modern Swedish, normally refers to a school subject. This is the only sense given in the "SOB", and it will probably be treated in the same way in the "NEO". According to the "SAOB", the first known occurrence of the word in Swedish took place in 1907. But at that time, there was no 'school' sense, only a more general sense. We must now decide whether this more general sense is so close to the sense given in the "NEO" that we can give the year 1907. One possibility is to give the information, say, '1907, in a slightly different sense'.

2.3. Word formation

In Swedish, most verbs may be nominalized by means of the morpheme -(N)ING and/or the morpheme -ANDE/-ENDE. There is a slight sense difference between these two kinds of nominalization; the second one often tends to express iterativity. Most of the -(N)ING and -ANDE/ENDE nominalizations are not treated as separate lemmas in dic-tionaries of Swedish. Therefore, any dictionary of a certain size will have to consider giving information on nominalization possibilities within the verb article. Such informa-tion is, to some extent, given in an earlier medium size dictionary of Swedish, "Illustrerad svensk ordbok" ('An illustrated dictionary of Swedish'), but it is not given systematically in the "SOB". In the "NEO" dictionary, it will be given.

There are some problems in connection with this information, however. Firstly, very often one of the nominalizations, either the -ANDE/ENDE one or (probably more often) the -(N)ING one, is very natural, while the other one is only theoretically possible. Should we give both in these cases? Secondly, in some cases, neither of the -(N)ING and the -ANDE/ENDE nominalizations is more than theoretically possible, some other kind of nominalization being preferred. For instance, the natural nominalization of the verb JÄMFÖRA ('compare') is JÄMFÖRELSE, while JÄMFÖR(N)ING is impossible and JÄM-FÖRANDE is hardly more than theoretically possible. Should we give JÄMFÖRANDE? And if the answer is 'yes', should we not also give the information that the normal

nominalization is JÄMFÖRELSE? (It is not possible to give all minor kinds of nominaliz-ation systematically.)

2.4. Synonyms, hyperonyms, hyponyms

It is an interesting question to what extent a monolingual defining dictionary can be designed as an active dictionary, a dictionary where you can find 'the right word'. The "NEO" dictionary is intended, to some extent, to be an active dictionary. Obviously, references to synonyms are relevant in this connection. But (direct or indirect) references to co-hyponyms are perhaps still more important. However, a defining dictionary could hardly be arranged as a thesaurus, but possibly it could adopt some of the features of a thesaurus. As an example, let us have a look at names of animals' 'homes', e. g., NEST, LAIR, etc. Now, it is hardly possible to refer from every such word to all the others, but it is possible to refer from every word to the nearest hyperonym, if there is one (if not, you will have to choose a substitute). A reasonably good (pseudo-)hyperonym in this case is BO (roughly 'nest'). Then, the entry BO should contain references to all the (pseudo-)hy-ponyms. We may now imagine a user looking for the Swedish word for the bear's winter home, IDE. Either he will feel that he should look at once at the entry BO, or he can look at any word denoting an animal's home and find a reference to BO. Under the entry BO, then, he will find a reference to IDE. He may either recognize this word, or – if, e.g., he is an non-Swede – he will have to check the explanations of some of the words referred to.

2.5. A preliminary "NEO" entry

Finally, we may take a look at a preliminary NEO entry, SALTA ('to salt'; see table 2). We will concentrate on those parts of the article that are relevant to the discussion in this paper. As is clear from the first line, this word has accent 2, and it is possible to form the nominalizations SALTNING and SALTANDE. A little later, we find a reference to the hyperonym SMAKSÄTTA ('to give extra flavour to'), where the user will be provided with some co-hyponyms of SALTA, e.g., PEPPRA ('to pepper'). At the end of lexeme 1, you get the information that the word, in this sense, appeared in Swedish before 1520, followed by a brief etymology. As for the second sense (SALTA EN RÄKNING 'to present a stiff bill'), the information is given that its first known appearance in Swedish dates back to 1700.

sa͡t/a verb -*ade* -*at:* -ning -ande
1 • tillsätta (kok)salt till livsmedel. för att
förhöja smaken {▷ **smaksätta**}: ∼ *och peppra*
köttet efter behag o äv. a) utvidgat. i
konserveringssyfte: *genomsaltad: lättsaltad: -ad*
skinka: -at smör b) ytterligare utvidgat: *här i*
staden -ar man vägarna vintertid i stället för
att sanda dem c) överfört. särskilt i uttryck för
från kritik. ironi e. d.: ∼ *sitt tal med sarkasmer*
[före 1520: fornsv. *salta:* till ¹salt. ²salt]
2 • lägga oskäliga kostnader på räkning e. d.:
en -ad hotellräkning; ∼ *notan* [1700; se 1]

Table 2.
The (preliminary) entry
SALTA in the "NEO"

Bibliography

ALLÉN, Sture (1981): "The Lemma-Lexeme Model of the Swedish Lexical Data Base." In: Empirical Semantics. II. Ed. by B.B. Rieger. Pp. 376-387. Bochum.

BUCK, Timothy. (1984): "Neue Maßstäbe im deutschen Wörterbuch?" In: Der Spiegel 48. Pp. 218-220.

DUDEN. Deutsches Universalwörterbuch. (1983). Mannheim/Wien/Zürich.

GELLERSTAM, Martin & MALMGREN, Sven-Göran (forthcoming): "Recent trends in monolingual Swedish lexicography." (Paper read at the 5th Polish-Swedish linguistic symposium; to be printed in the Symposium volume.)

ILLUSTRERAD SVENSK ORDBOK. ('An illustrated dictionary of Swedish'.) 1956. Stockholm.

MALMGREN, Sven-Göran (1988): "Some characteristics of Svensk ordbok, a new defining dictionary of Swedish." In: Symposium on lexicography IV. Ed. by K. Hyldgaard-Jensen & A. Zettersten. Pp. 133-141. Tübingen.

ORDBOK ÖVER SVENSKA SPRÅKET UTGIVEN AV SVENSKA AKADEMIEN. ('A dictionary of Swedish edited by the Swedish Academy'.) 1898-. Lund.

SVENSK ORDBOK. ('A dictionary of Swedish'.) 1986. Stockholm.