Anna Sågvall Hein, Department of Linguistics, Uppsala University

From natural to formal dictionaries

ABSTRACT: Possibilities and problems in the development of machine dictionaries from human dictionaries are discussed. In particular, the concrete results of our current work on a formalization of a monolingual Swedish dictionary are presented: a machine-tractable morphological description, and data on the coverage of the dictionary with regard to the definitions. Most problems encountered in the formalization of dictionaries are due to their use of natural, or semi-natural, language, without providing for the implied background knowledge. To identify and fill this gap is a major task in the development of machine dictionaries.

A well-known obstacle to large-scale application of computational language models is the shortage of comprehensive, machine-tractable dictionaries. During the last decade, much work has been devoted to the development of strategies for automatic or semi-automatic extraction of lexical knowledge from text (see e.g. Church & Hanks 1990), and to the development of machine dictionaries based on human dictionaries (see e.g. Byrd et al. 1987; Boguraev et al. 1989). Here some aspects of the dictionary alternative will be examined with regard to our work on a formalization of Svensk Ordbok [A Dictionary of Swedish] (1986), SOB.

Computational language models, as opposed to traditional language descriptions, are well-defined with respect to linguistic behaviour. They are directed towards limited aspects of comprehension, production, and translation, and, accordingly, make precise demands on linguistic competence. Examples of commonly recognized comprehension aspects are the recognition of words, the recognition of syntactic and semantic structure, and the linking of the linguistic objects thus identified to referents in an interpretation phase. Typically, a formalization effort is directed towards a specific language model with its distinct demands in terms of linguistic aspect, coverage, and representation format; in our case the Uppsala Chart Processor, UCP, (Sågvall Hein 1987), the computational framework of a parser for Swedish[1].

The information stated in machine dictionaries must be exhaustive and explicit, in order to be fully exploitable by computational models, comprising, typically, operations of search and inferencing. This is not the case with human dictionaries which rely heavily on unspecified background knowledge. A fundamental problem, thus, in the development of machine dictionaries from human dictionaries lies in identifying this background knowledge, and spelling it out in a formal and exhaustive way.

SOB is a monolingual Swedish definition dictionary of almost 60,000 entries, lemmas (lexical words, see Allén 1970: XXXV). The lemmas are described with regard to spelling, part of speech, inflection, and, in some cases, pronunciation. Each lemma is accompanied by the lexemes (main senses, see Allén 1981: 382), that it may realize. The meanings of the lexemes are expressed by definitions, semantic relatives, and examples.

1 A formalization of the inflectional model of SOB

Inflection in SOB is described in terms of stem, part of speech, and significant endings, e.g.

(1a) fest [party] subst. [noun] –en –er.

(1a) is a compact representation of the inflectional potential of the noun, spelled out by the full paradigm presented in (1b).

(1b)	*singular*		*plural*	
	indefinite	*definite*	*indefinite*	*definite*
basic case	fest	festen	fester	festerna
genitive	fests	festens	festers	festernas

To infer (1b) from (1a) is no problem to a linguistically trained dictionary user, mastering the Swedish inflectional system. The inference is based upon knowledge about the inflectional categories of Swedish nouns (number, case, species), about their expressions (-, -s, -en, -er, -na etc.) and about their order of appearance. Further, the user must know about the lexicographic convention of presenting the paradigm via significant endings, and, additionally, that the crucial endings of Swedish nouns are those of the definite singular form (decisive for gender), and the indefinite plural (decisive for declension type). This is quite an impressive amount of knowledge, partly taken for granted by the lexicographer, partly presented in the introduction to the dictionary (in natural language). In order to make the SOB inflectional pattern usable to a morphological analyser, we must spell this implicit knowledge out, or express inflection in a totally different way. One alternative, even though uneconomical, would be to present it by means of the full paradigms (cf. 1b). Our approach has been to represent the information conveyed by the paradigms by means of rules, i.e. recognition rules tuned for UCP. In (1c) we present the UCP format of the rule corresponding to the inflectional pattern of (1a) with its implicit knowledge, or to (1b).

(1c)
```
(define sve.gram-entry pattern.film
   #u noun('-er, 'utr, 'cons);
```

The name of the rule is pattern.film, whose final part is a model word of the paradigm type. Pattern.film calls a subrule, noun, defining the morphotax of the nouns. The noun rule is invoked with three parameters, where the first one holds the plural ending, the second one gender, and the third one type of stem-final segment. The gender parameter, in combination with the stem-final one, determines the choice of ending in the definite singular form. Letting the noun rule parameters vary over the possible values, we account for 26 paradigms. (1c), thus, presents the standard format of a pattern rule describing a paradigm whith no stem alternation. For the description of stem alternations, the pattern rules are further elaborated. In (1.1) we present the general noun rule (in a slightly simplified version).

```
(1.1) (define sve.gram-entry noun
       #u <& gd word.cat>:=:'noun,
          <& prop decl>:=:?1,
          <& prop gender>:=:?2,
          <& prop final>:=:?3,
          (advance,
           (no.noun.flex/
            process(case), process(numb,<prop>),
            process(form, <prop>))
           //advance(plur.suff));)
```

This is not the place to go into the details of the formalism and its procedural nature (Sågvall Hein, ibid.). Suffice it to point out that it includes means for guiding the analysis (initiating dictionary search ((process(numb,<prop>) etc.) and for invoking and processing grammar rules (no.noun.flex, plur.suff etc.). Basically, the noun rule invokes search in the affix dictionaries and controls the morphotactic structure of the noun.

In table 1 we illustrate the relations between the SOB inflectional format and that of our machine dictionary. Word class is explicit in SOB, whereas the stem is represented via the lemma. It can be identical to the lemma (1, 3-4), to the lemma without its homograph number (2 in table 1)[2], or to an initial string of the lemma (delimited by a slash as in 5 and 7). In cases of umlaut or other types of stem alternation that cannot be expressed by the lemma in an obvious way, all the stem alternants are exposed via representative inflectional forms (6: *hammar(e)n* [the hammer], *hamrar* [hammers]; 8: *händer* [hands]). In the SMU[3] dictionary, stem and lemma are systematically kept apart and represented explicitly, and the stem is the entry to the lemma[4]. The inflection of each stem is given via the model word of its pattern rule. For instance, the pattern rule of the model word akademi recognizes the forms of the -er declension (cf. pattern.film) and, in addition, handles the variation poesin/poesien in the definite singular form (2 in table 1).

As regards the stem, SOB exclusively uses the technical stem, "that part of the word which is common to all inflectional forms" (Hellberg 1978: 13), whereas SMU distinguishes between technical stem (dictionary entry) and linguistic stem (technical stem + (optionally) stem extension), hereby avoiding hybrid affixes including stem elements. Thus SOB and SMU differ in the treatment of words with stem alternation, see, for example, the noun *ovän* (4) with a geminated stem consonant in the definite and plural forms. SOB refers the twin consonant to the ending, whereas the SMU grammar treats it as an extension of the dictionary stem, recognized by pattern.vän. The possibility of recognizing stem extensions in the procedural pattern rules is frequently used in the SMU grammar (see also 5 and 6 in table 1). 6 is an example of a lemma with unsettled declension; the zero plural ending ('=' in SOB), as well as, the -ar plural ending are used. Such cases of free variation are also handled by means of pattern rules, recognizing the different alternatives.

7 gives one more example of departure from the technical stem in SMU. It concerns the treatment of the secondary e-vowel, a phenomenon which is handled by a general

SOB: lemma		wc	ending	SMU: stem	lemma	inflection
1	fest	subst.	–en –er	fest	fest.nn	film
2	^{2}gång	subst	–en –er	gång	gång2.nn	film
3	poesi	subst.	–(e)n –er	poesi	poesi.nn	akademi
4	ovän	subst.	–nen –ner	ovän	ovän.nn	vän
5	studi/um	subst.	–et –er	studi	studium.nn	decennium
6	hammare	subst.	hammar(e)n, hamrar el. [or] =, best. plur. [def. plur.] hamrarna	ham	hammare.nn	kammare
7	reg/el	subst.	–eln –ler	reg(e)l	regel.nn	möbel
8	hand	subst.	–en händer	hand händer	hand.nn hand.nn	strand stränder

Table 1.
Inflection in SOB and in the SMU dictionary. Examples from the -er declension and the -ar declension.

rewriting rule, defined in the grammar and invoked in connection with the dictionary search. Thanks to the operation of this rule, the recognition of both stem alternants can be based on a common dictionary stem, the stem in its non-vowel stage.

The SMU grammar comprises 135 (stem) pattern rules for the nouns (incl. proper nouns), 39 for the adjectives, 65 for the verbs, 1 for the articles, 30 for the pronouns, 5 for the numerals, 9 for the adverbs, 2 for the conjunctions, and one each for the prepositions, the interjections, and the infinitive marker. In most cases, the inflectional analysis is based on one (dictionary) stem, and the stem with its pattern rule is then a sufficient characterization of the inflectional behaviour of the lemma. If, on the other hand, the lemma is represented by more than one stem in the dictionary (cf. 8 in table 1), the set of stems involved determine the inflection of the lemma, its paradigm type. The concept of paradigm type is based on stem type, as expressed by the unique pattern rules. For instance, to the same paradigm type we refer nouns of the -er declension with umlaut: *strand – stränder, hand – händer* etc.

All in all, there are 163 paradigm types of nouns, 42 of adjectives, and 106 of verbs[5]. These figures reflect a generalization of the inflectional classification emerging from a mechanical sorting of the SOB inflectional patterns (Sjögreen forthc.). Within the productive lexical categories, a large number of types was arrived at, viz. 676 for nouns, 162 for adjectives, and 546 for verbs. Among the nominal inflectional types, 403 (60%) have only one representative, among the adjectival types 107 (66%), and among the verbal types 361 (66%). The large proportion of one-word types is due to the mechanical nature of the sorting process. All inflectional patterns comprising a full form (cf. 6 and 8 in table 1) emerged as individual paradigms.

The SMU dictionary[6] and grammar together with UCP constitute the SMU analyser. It assigns morphological descriptions to the words covered by SOB. The morphological description resulting from the analysis of the noun *festernas* may illustrate such a description:

festernas	LEMMA=FEST.NN, WORD.CLASS=NOUN, INFL=FILM, GENDER=UTR, FORM=DEF, NUMB=PLUR, CASE=GEN

For an overview of the descriptive attributes assigned by the SMU analyzer, see Sågvall Hein (forthc.).

SMU is well-defined via its relation to SOB: prior to any processing we may consult SOB to find out for any word form whether it will get an analysis or not; the dictionary provides an intuitive format through which we may explore the competence of the analyzer without any prior knowledge of its formalisms or operation. SMU is also well-defined in another sense: via the SMU lemma, a direct link can be established to the SOB lemma with its lexemes and definitions.

Neither SOB nor SMU (in its present shape) comprize any word formation rules, and words outside their common scope are left without an analysis. Consequently, SMU can be used for an investigation of the coverage of SOB. With this aim, it has been applied to the SOB definitions, as well as to other text materials of substantial size. Here we will concentrate on the definitions.

2 The coverage of SOB with respect to the definitions

In spite of the large number of entries in SOB, many general language words are left out, in particular numerical expressions, abbreviations, proper nouns, derivatives (incl. phrasal verbs) and compounds[7]. How can we account for words of these types in our machine dictionary? As regards numerical expressions, proper nouns, and abbreviations, the actions to be taken are rather obvious; numerical expressions will have to be represented by means of rules (a task for which the UCP formalism is well adapted); proper nouns and abbreviations be included in the dictionary, as called for by the application at hand.

With derivatives and compounds the problem is much more difficult. The derivational and compounding disposition of the lexemes is illustrated by morphological examples, but for the rest, the lexicographers trade on implicit knowledge about the Swedish word formation system. This is a kind of knowledge that has to be accounted for in the formalization of the dictionary[8]. Thus we have to identify the derivatives and compounds that were left out (those with a meaning implicitly covered by the definitions of their parts, see further SOB: VI), find out how they can be represented in a rule-based manner, and, in particular, determine the prerequisites for deriving their meanings from their constituents.

The definitions comprize 43,934 types and 360,144 tokens, and as a result of the SMU analysis, 12,584 missing types were identified. Approximately 70% of them were found to be nouns, and among them 7,914 compounded ones.

3 Prerequisites for a semantic calculus of the meaning of uncovered derivatives and compounds

A rule-based representation of derivatives and compounds presupposes a semantic calculus. The meaning of the derived or compounded words should be computable from the meanings of their constituents. In other words, a first condition for a semantic calculus of the derived words is access to the "mother" definitions by formal means.

By means of structural word formation rules, the mother lemma of a derivative can be determined[9], e.g. *öppen* [open] as the mother lemma of the derived noun *öppenhet* [openness]. However, associated with the lemma *öppen* in SOB are seven lexemes, representing concrete as well as abstract senses of the word. The noun *öppenhet* appears as a morphological example of a transferred meaning[10] of the 4th lexeme. Thus SOB points out a primary mother candidate of the noun, without excluding other possibilities: there are no claims about exhaustiveness in the presentation of morphological examples. In a previous study of newspaper text, most of the uncovered derivatives were found to be derived from unambiguous lemmas (lemmas with only one lexeme), or presented as morphological examples of their typical mothers (Sågvall Hein, forthc.). In other words, a first condition for rule-based representation of (some types of) derivatives was found to be reasonably fullfilled. We assume this to be the case for the definition derivatives as well.

As regards the compounds, however, the situation is worse. In the set of uncovered compounded nouns we found 2,012 different heads, and in table 4 we present the 10 most used ones. With one exception (*område*), they can all be traced back to more than one definition (via more than one lemma, more than one lexeme, or more than one definition of one lexeme: kernel sense and transferred sense, respectively). The definition of *område* refers to the concrete sense of the word, but the possibility of an abstract sense is also mentioned, even though not explicitly defined. Thus, in this case, as in similar ones, the dictionary user should be capable of inferring an abstract sense from the definition of a concrete one. We didn't examine to what extent the individual uncovered compounds were represented by morphological examples; suffice it to note that the number of morphological examples of the head (as the head of a compound head) is, in general, smaller than its number of uses in different uncovered compounds. The idea, however, of basing the lexical choice on a list of morphological examples must be rejected; it would presuppose a large-scale extension of the dictionary, and be in conflict with the fundamental idea behind rule-based representation of new words, i.e. to account for the handling of new words, among them unforeseen ones.

From the data presented in the table 2, we conclude that the appropriate definition of the head constituent may be mechanically accessed only in a few cases. If, in addition, the potential ambiguity of the first constituent of the compound is taken into account, the

head	no. of def.	morph. ex.	diff. comp.
form [form etc.]	7	17	84
område [area]	1	13	69
anordning [device etc.]	3	3	68
växt [plant]	4	16	67
system [system]	3	17	60
tid [time]	2	3	45
medel [means]	3	4	43
redskap [tool]	2	1	43
material [material]	3	5	39
stycke [piece etc.]	8	1	38

Table 2.
The most used heads of the uncovered compounded nouns in the definition vocabulary

number of definitions to be considered in the semantic calculus of the compound will increase substantially.

The lexical choice in compounding is a fundamental problem, and it must be based on knowledge about the meaning of the words, presented in the definitions and the examples. For this purpose, the definitions have to be formalized. Such a formalization is also called for as a prerequisite to the semantic calculus as such, too.

4 A fundamental obstacle to a formalization of the definitions

The SOB definitions are expressed in, basically, unrestricted Swedish (Järborg 1988). Consequently, they display variation, ambiguity, redundancy, incompleteness, and vagueness, thereby failing to fulfill the demands on computability made by a computational language model.

Ambiguity, incompleteness, and vagueness present classical problems to computational language comprehension, which is what we are, in fact, dealing with when attempting to formalize the definitions. Most of the ambiguity problems can, in principle, be solved or reduced by systematic paraphrasing of the definitions. Aspects of incompleteness and vagueness, however, represent problems of a different nature.

For an illustration of the problems involved, we will discuss what the conditions are for basing an interpretion of the phrase *"djup snö"* [deep snow] on the definitions of its constituents. The phrase is given as an example of the first lexeme of the adjective *djup* [deep], defined as follows: "som har stort avstånd mellan yta och botten ... [having a large distance between surface and bottom ...]/my transl.]"

The lexeme *"snö"* in its kernel sense is defined: "nederbörd i form av vita, mjuka, löst sammanfogade iskristaller ... " [precipitation in the shape of white, losely connected ice crystals...]. The head of the definition, precipitation, is primarily defined as an amount without any spatial extension. In other words, the definition of *"snö"* in its kernel sense and that of its hyperonym *"nederbörd"* don't give any basis for an interpretation of a distance.

In addition to the kernel sense of *"snö"*, SOB distinguishes a transferred sense of the word, with the definition "äv. om större mängd av (nedfallna) sammanpackade iskristaller av detta slag" [also about a large amount of (fallen down) ice crystals of this kind that are packed together]/my transl. This sense of snow is illustrated by the example *"snötäcke"* [snow cover], whose head *täcke* [cover] is concrete with a potential extension in space. This aspect is, however, not emphasized in the definition of the kernel sense of *täcke*: "större, rektangulärt, dubbelt tygstycke, stoppat med ..." [large, rectangular, double piece of cloth, stuffed with ...]. In a transferred meaning we find "äv. om naturligt skikt av ngt som breder ut sig över marken, himlen e.d.:" [also about a natural layer of something spreading over the ground, the heaven etc.] and among the morphological examples reappears *"snötäcke"*. So far, we have been able to trace *"snö"* back to *"skikt"*, which in turn is defined in terms of *"tunnare lager"* [thin layers], and *"lager"* as "utbredd massa med relativt ringa tjocklek <syn. skikt>:" [widely spread mass of a relatively small thickness <syn. layer>].

Via definitions of kernel senses, definitions of transferred senses, and examples we can find a path in the dictionary from *"snö"* to "utbredd massa med relativt ringa tjocklek", but not to the information we need in order to interpret *"djup"* in relation to *"snö"*, i.e. a basis for determining its surface and bottom, and the distance between them. This is a meaning aspect of snow which must be implied, or explicitly stated, for the definition of deep to make sense in the example. However, this is not enough. In addition, we need a norm for the depth of snow. When is snow considered to be deep? That depends on the situation. Half a meter of snow may be much for a tired person walking in the snow, but little for certain types of skiing.

Thus we conclude, that for the definitions to function as the primary vehicles of comprehension in a computational language model, they must, among other things, be augmented with extra-linguistic knowledge, based on experience. The task at hand is so gigantic, that it cannot be approached in any general scale, but only in relation to strictly delimited comprehension goals. Is formalizing those definitions of SOB on which the uncovered definition compounds are based, to the extent required for a semantic calculus of them such a realistic goal?

5 Summarizing remarks

Our work on the development of a machine dictionary from SOB includes a formalization of the inflectional model of the dictionary, studies of the coverage of the dictionary with regard to the definition vocabulary, preliminary studies of the uncovered compounds incl. prerequisites for a semantic calculus of them, and, some reflections on the feasibility of a formalization of the definitions. Inflection is expressed in a semi-formal way, and takes an inflectional grammar for granted. The intended coverage presupposes full command of the word formation system, structurally as well as semantically. The semantic calculus of the uncovered, derived words must be based on a formalization of the definitions, a task which, in addition to linguistic knowledge, requires situational knowledge based on experience.

End notes

1 Work on a formalization of SOB for UCP was initiated and, to a large extent, carried out in the project a Lexicon-oriented Parser for Swedish, at the department of Computational Linguistics, University of Göteborg. The LPS parser is responsible for handling the analysis part of the machine translation component of the project Multilingual support for translation and writing, Multra, at the department of Linguistics, Uppsala university.

2 Homograph lemmas are systematically kept apart by numbers.

3 SMU is short for Swedish Morphology in the Ucp formalism.

4 This is the case when the dictionary is engaged in morphological analysis. However, to support further lexicographic work on the dictionary, a corresponding database, with a menu-based interface, has been created. The interface allows the user to choose how to enter the base: via a stem, a lemma, or a model word.

5 Frequency data on the inflectional types of the SMU words have been presented elsewhere, see Sågvall Hein & Sjögreen 1991. The first formal morphological description of Swedish was formulated by Hellberg (1978), accounting for inflection and word formation. With respect to inflection, the Hellberg classification is somewhat less exhaustive than that of SOB and SMU. (For instance, 133 noun paradigms were established, 18 of them due to issues of word formation.) Hedelin & Huber (forthc.) claim to cover "inflection, derivation and compounding both in written and in spoken language" by means of a system of 500 paradigms (as compared to the 235 paradigms of Hellberg). It is not clear, how many of them that are due to inflection in written language. Karlsson (forthc.) presents a comprehensive description of Swedish inflection and word formation in Koskenniemi's (1983) twolevel formalism. Because of the different descriptive approaches, paradigms versus twolevel rules, a comparison of descriptive power with regard to inflection, (word formation, so far, outside the scope of SMU) has to be based upon performance rather than on competence. Karlsson (ibid.), includes a summary of previous computational approaches to Swedish morphology.

6 The work on the SMU machine dictionary was carried out in cooperation with Christian Sjögreen at the University of Göteborg (Sågvall Hein & Sjögreen 1991). It was based on a well-organized set of datafiles from which the printed version of the dictionary was drawn (Sjögreen 1988).

7 The proportion of each type differs with the text, for instance, proper nouns and abbreviations dominate the missing words in newspaper text, numerical expressions (incl. mixtures of numbers, letters, and special signs in pharmacological text), and, finally, compounds in the definition corpus as well as in a mixed corpus of LSP text that we examined (Sågvall Hein, forthc.).

8 To include all possible compounds is an impossible strategy as applied to the general language, due to the openness of the compounding mechanism allowing sometimes for, as it seems, quite accidental formations, e.g. additionsflink [~quick at adding] found in our newspaper material. See further Blåberg 1988; Karlsson forthc.

9 The aptness of the UCP formalism for the formulation of structural word formation rules has been demonstrated elsewhere (Sågvall Hein 1987). Karlsson (forthc.) presents a comprehensive set of structural word formation rules for Swedish in Koskenniemi's (1983) two-level formalism.

10 "The lexeme is assumed to possess a primary sense, called kernel sense and, facultatively, several transferred senses (extended, specialized, metaphorical, etc.), derivable from the kernel sense by regular processes." (Järborg 1987: 143).

Bibliography

ALLÉN, S. (1970): Nusvensk frekvensordbok baserad på tidningstext. 1. Graford. Homografkomponenter. [Frequency dictionary of present-day Swedish based on newspaper material. 1. Graphic words. Homograph components.] Stockholm.

ALLÉN, S. (1981): "The Lemma-Lexeme Model of the Swedish Lexical Data Base". B. Rieger (ed.), Empirical Semantics. Bochum. Pp. 376-387.

BLÅBERG, O. 1988. A study of Swedish compounds. Umeå University. Dept. of General Linguistics. Rep. No 29.

BOGURAEV, B., BYRD, R., KLAVANS, J., and NEFF, M. (1989): "From machine-readable dictionaries to lexical databases". First lexical acquisition workshop, Zernik ed., Detroit.

BYRD, R, CALZOLARI, N., CHODOROW, M. KLAVANS, J. NEFF, M., RIZK, O., (1987): "Tools and methods for computational lexicography". Computational Linguistics, Vol. 13, Numbers 3-4.

CHURCH, K. W. & HANKS, P. (1990): "Word association norms, mutual information, and lexicography". Computational Linguistics, Vol. 16, Number 1.

HEDELIN, P. & HUBER, D. (Forthc.): A new dictionary of Swedish pronounciation. Papers from the eighth Scandinavian conference of computational linguistics.

HELLBERG, S. (1978): The morphology of present-day Swedish. Stockholm.

JÄRBORG, J. (1988): "Towards a formalized lexicon of Swedish". In: Studies in computer-aided lexicology. Stockholm. Pp. 140-158.

KARLSSON, F. (Forthc.): SWETWOL: "A comprehensive morphological analyzer for Swedish". Nordic Journal of linguistics.

KOSKENNIEMI, K. (1983): Two-level morphology: A general computational model for word-form recognition and production. University of Helsinki. Dept. of General Linguistics. Publications No. 11.

SJÖGREEN, C. (1988): "Creating a dictionary from a lexical database". In: Studies in computer-aided lexicology. Stockholm. Pp. 299-338.

SJÖGREEN, C. (Forthc.): Från SOB till LPS. En arbetsrapport från projektet En Lexikonorienterad Parser för Svenska (LPS). [From SOB to LPS. A working report from the project A Lexicon-oriented Parser for Swedish.]

SÅGVALL HEIN, A. (1987): "Parsing by means of Uppsala Chart Processor", (UCP). In: L. Bolc (ed.) Natural language parsing systems. Berlin & Heidelberg. Pp. 203-266.

SÅGVALL HEIN, A. & SJÖGREEN, C. (1991): "Ett svenskt stamlexikon för datamaskinell morfologisk Analys". En översikt. [A Swedish stem dictionary for computational morphological analysis. An overview.] In: M. Thelander et al. (eds.) Svenskans beskrivning 18. Lund. Pp. 348-360.

SÅGVALL HEIN, A. (Forthc.): "The Coverage of a Morphological Analyser based on Svensk Ordbok" [A Dictionary of Swedish]. Papers from the eighth Scandinavian conference of computational linguistics.

SVENSK ORDBOK [A Dictionary of Swedish.] (1986): Stockholm.

KEYWORDS: computational lexicography, computational morphology, Swedish morphology, dictionary coverage, lexical definitions