Preferences and Linguistic Choices in the Multra Machine Translation System

Anna Sågvall Hein
Uppsala

Abstract

The work to be presented here concerns the ordering of alternatives in the Multra Machine Translation System. The Multra MT system is a fundamental part of the Multra prototype, modeling a translation work bench with user-controlled mixed mode of mechanical and human translation. The Multra system is based on transfer and unification. It includes three main modules, responsible for analysis, transfer, and generation, respectively. In addition, there is a separate preference module ordering the analysis alternatives before passing them on to the transfer component. Preferences are expressed by means of linguistic rules defined over feature structures. Alternative transfer rules are applied according to specificity; a specific rule takes precedence over a more general one. The specificity principle also governs the application of generation rules. The MT system as a whole, as well as its separate modules, can be tuned to present the best alternative only, or the complete set of alternatives in the preferred order.

1 Introduction

The work to be presented here was carried out in the project *Multilingual Support for Translation and Writing, Multra* (Sågvall Hein 1993a). It concerns the ordering of alternatives in the Multra Machine Translation system. The Multra MT system is a fundamental part of the Multra prototype, modeling a translation work bench with user-controlled mixed mode of mechanical and human translation. In its present version, Multra supports the translation of car maintenance manuals from Swedish to German and English.

The Multra system is based on transfer and unification. It includes three main modules, responsible for analysis (Sågvall Hein 1987 and in preparation), transfer (Beskow 1993a), and generation (Beskow 1993b). In addition, there is a separate preference module ordering the analysis alternatives before passing them on to the transfer component. Preferences are expressed by means of linguistic rules defined over feature structures. Alternative transfer rules are applied according to specificity; a specific rule takes precedence over a more general one. The specificity principle also governs the application of generation rules. The preference rules along with the specificity principle of the transfer and generation processes constitute the Multra preference machinery.

Proceedings of NODALIDA 1993, pages 267-276

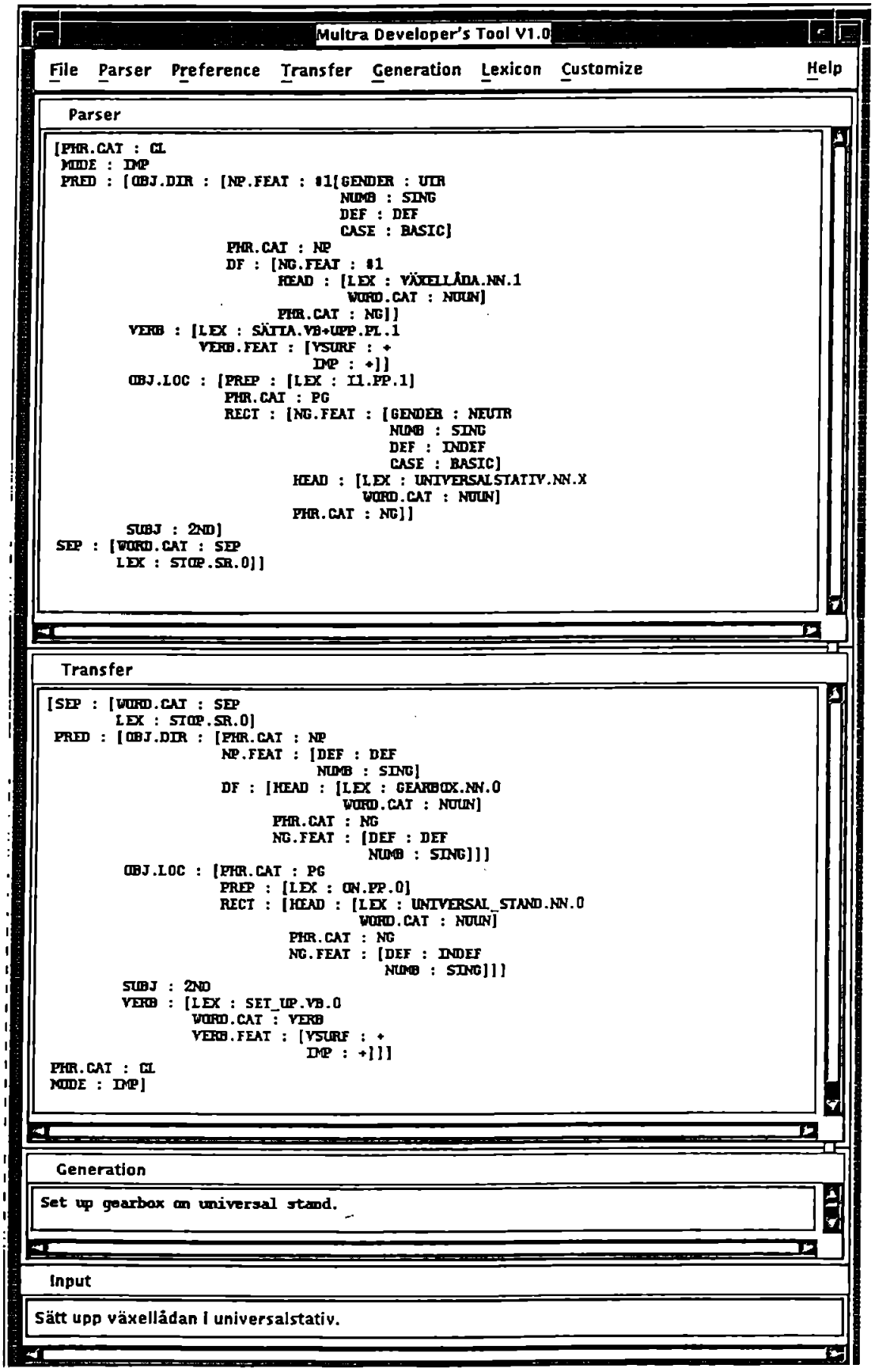

Figure 1: First-best translation of *Sätt upp växellådan i universalstativ.*

The MT system as a whole, as well as its separate modules, can be tuned to present the best alternative only, or the complete set of alternatives in the preferred order. For the design and testing of translation rules, a special environment, Multra Developer's Tool, MDT (Beskow 1992), has been developed, and we will start our presentation of the Multra MT system and preference machinery in this environment.

2 An example

Fig. 1 presents the first-best translation of the Swedish sentence *Sätt upp växellådan i universalstativ.* (see **Input** window) into English, *Set up gearbox on universal stand.* (see **Generation** window). The sentence is analyzed as an imperative clause consisting of a predication (a verb with its complements), and a separator (see **Parser** window). The predication is made up by the phrasal verb *sätta upp [set up]* (lexeme SÄTTA.VB+UPP.PL.1) and its (elliptic) subject, direct object, and locational object. Recursively and in parallel lexical and structural transfer rules apply to analysis structure yielding the English transfer structure displayed in the **Transfer** window.

The target transfer structure is (functionally) isomorphic to the source analysis structure, and the translation process may seem trivial. It does, however, include three kinds of phenomena that call for the preference machinery. They relate to the analysis phase, the transfer phase, and the generation phase, respectively, as will be demonstrated below.

The Swedish noun *universalstativ* (the head of the locational object) doesn't distinguish formally between its singular and plural forms. The intended reading in this example is singular, but a plural reading, eventhough rare, cannot be excluded in this type of contexts. Thus both alternatives have to be accepted but priority be given to the singular form. A preference rule (i) takes care of that.

(i)
```
PREFERENCE OBJ.LOC.SING.PLUR
        <* PRED OBJ.LOC RECT NG.FEAT NUMB> = SING
PRECEDES
        <* PRED OBJ.LOC RECT DF NG.FEAT NUMB> = PLUR
```

As is the case with most prepositions, there are several translations of the Swedish preposition *i*, even though it has been recognized as denoting location in space (not in time). Its default translation into English would be *in*, but when it collocates with *universal stand*, *on* is the correct expression. In other words, the transfer component must account for a default translation, as well as for a translation in context. We introduce the Multra transfer rule format (Beskow 1993a) by presenting the simple lexical rule accounting for the default translation of the preposition (ii).

(ii) LABEL
I1
SOURCE
<* LEX> = I1.PP.1
TARGET
<* LEX> = IN.PP.0
TRANSFER

Lexical transfer rules in Multra define translation relations between lexemes, or basic senses (Allén 1981). The rule in (ii) relates the (Swedish) source lexeme I1.PP.1[1] to the (English) target lexeme IN.PP.0. Analogous to (ii) is the lexical transfer rule UNIVERSALSTATIV presented in (iii).

(iii) LABEL
UNIVERSALSTATIV
SOURCE
<* LEX> = UNIVERSALSTATIV.NN.X
TARGET
<* LEX> = UNIVERSAL_STAND.NN.0
TRANSFER

The translation of *i* in context is handled by a transfer rule covering the preposition along with the noun that it governs (iv). The rule applies to a prepositional group, PG, consisting of the preposition I1.PP.1 and a nominal expression with UNIVERSALSTATIV.NN.X as its head. Further, the whole of the nominal expression governed by the preposition, its rection, is assigned to the variable ?RECT1. Corresponding to the source structure of (iv) the rule defines a target prepositional group introduced by the preposition ON.PP.0. Further a target language attribute, RECT, is defined with the variables ?RECT2 as its value. Finally, ?RECT2 will be bound to ?RECT1 via the TRANSFER relation; recursively and in parallel, transfer rules will be applied to ?RECT1, concluding with the application of the lexical rule UNIVERSALSTATIV (iii).

Both (ii) and (iv) are applicable to our example. However, (iv), or rather its source part, is more specific than that of (ii), and consequently, (iv) will be preferred. (Being more specific means specifying a greater number of identity relations, more specific identity relations, or a greater number of transfer relations, see further Beskow 1993b).

In the transfer process of the example, no shift (cf. Ingo 1990) of function, structure, category, or feature takes place. For instance, the

[1]A lexeme is represented by the basic form of its lemma, followed by a part of speech marker, and a lexeme number. The Swedish lexeme numbers accord with those given in Svensk Ordbok (1986). Lexemes outside the scope of Svensk Ordbok are assigned lexeme number X. If the basic forms of two lemmas coincide, numbers keep them apart, as in our preposition example. As for target lexemes, they are, so far, assigned a zero lexeme number.

Swedish direct a object in the definite form is transfered as such into English.[1] However, in accordance with the English model translation,[2] the direct object will appear in its indefinite form in the resulting translation, viz. *Set up gear box on universal stand.* Thus a shift of definiteness will take place in the generation phase, as will be explained below.

(iv)
```
LABEL
    I_UNIVERSALSTATIV
SOURCE
    <* PHR.CAT> = PG
    <* PREP LEX> = I1.PP.1
    <* RECT HEAD LEX> = UNIVERSALSTATIV.NN.X
    <* RECT> = ?RECT1
TARGET
    <* PHR.CAT> = PG
    <* PREP LEX> = ON.PP.0
    <* RECT> = ?RECT2
TRANSFER
    ?RECT1 <=> ?RECT2
```

The standard rule for generating the predication of an English imperative clause with a direct object and a locational object is presented in (v) below. The rule is formulated in a PATR like style (Beskow 1993a). It comprises three parts, i.e., a label, a sequence of constituents (variables) to be generated, and a number of identity equations, binding the variables to path expressions in the transfer structure and expressing constraints upon this structure.

(v)
```
LABEL PRED3a
X1 --> X2 X3 X4 :
<X1 PRED SUBJ> = 2ND
<X1 PRED VERB> = <X2>
<X1 PRED OBJ.DIR > = <X3>
<X1 PRED OBJ.LOC> = <X4>
```

In (v), X1 refers to an imperative predication of a transfer structure, and the first equation identifies it as such. (The value of the implied SUBJ attribute is set to 2ND in imperative clauses.) The value of the verb attribute will be assigned to X2, the value of the direct object attribute to X3 etc.

In (vi) we present a generation rule that implies a shift of definiteness. It generates a direct object in the indefinite form from a direct object in the

---

[1]Working in a multilingual translation environment, we aim at a transfer component as simple and general as possible, referring the target language specific features to the generation components, see also Sågvall Hein 1993b.

[2]From the English version of our experimental text, a maintenance manual for trucks from Saab-Scania.

definite form, picking up the (unquantified) description field (DF) of the transfered object (Sågvall Hein, in prep.). (vi) being more specific than (v) will be preferred.

If we tune the parser, the transfer component, and the generation component towards all alternatives, six English translations will be generated and presented in the preferred order:

(vi) LABEL PRED3b
X1 --> X2 X3 X4 :
<X1 PRED SUBJ> = 2ND
<X1 PRED VERB> = <X2>
<X1 PRED OBJ.DIR PHR.CAT> = NP
<X1 PRED OBJ.DIR DF > = <X3>
<X1 PRED OBJ.LOC> = <X4>

Set up gear box on universal stand.
Set up the gear box on universal stand.
Set up gear box in universal stand.
Set up the gear box in universal stand.
Set up gear box in universal stands.
Set up the gear box in universal stands.

3 Preferences among source ambiguities

In Multra, the number of analysis alternatives is restricted as far as possible by maximal use of valency information; there is, for instance, no general PP-attachment rule. All PPs, modifying NPs, are attached by valency-rules. For instance *Ta bort luckan för kraftuttagshuset. [Remove the cover of the power take-off housing.]* gets only one analysis, according to which *för kraftuttagshuset [of the power take-off housing]* expresses appurtenance[1] in relation to *luckan [the cover].* Another example: The verb *sätta på* in *Sätt på lyftverktyget 87 792 på växlingsförarhusets plats. [Attach lifting tool 87 792 in position of gear selector housing cover.]* requires a locational object; thus, there will be no interpretation of *på växlingsförarhusets plats [in position of gear selector housing cover]* as a sentence adverbial. There are, however, cases, where the interpretation of a postposed PP as an adverbial cannot be excluded, and in those cases, a preference rule will give priority to the valency bound interpretation.

Number ambiguity is a common phenomenon in the Swedish source text, and even though singular is to be preferred in most cases, there are cases when the plural reading is the intended one. An example of such a case is the headline of a table, see for instance (vii).

[1] According to a suggestion made by Jarmila Panevova.

(vii) Specialverktyg [Special tools]

Fig [Fig]	*Nummer* [No]	*Benämning* [Name]
1	*79 046*	*Dorn* [Drift]
...		

The headline of a table is analyzed as a special kind of sentence fragment, a table name, and priority to the plural reading in such cases is given by a preference rule of the type presented in (i) above. Quite a number of contexts have to be specified in preference rules in order to account for number ambiguity.

Still another type of ambiguity to be handled by preference rules is due to elliptic coordination, see e.g. (viii).

(viii) *Ta bort de fyra skruvarna för locket och kopplingshävarmen.*
[Remove the four bolts of the cover and the clutch lever.]
a) Ta bort de fyra skruvarna för locket och (för) kopplingshävarmen.
b) Ta bort de fyra skruvarna för locket och (ta bort) kopplingshävarmen.

According to a) *lock* and *kopplingshävarm* are coordinated, according to b) the two imperative clauses. a) is to be preferred, and a preference rule may express view. By means of the examples presented above, we hope to have demonstrated that the machinery of preference rules is well apt for ordering structural ambiguities; slightly extended, it can apply to lexical ambiguities as well. The strategy of referring the ordering of source language ambiguities to a separate module contributes to the portability of an MT system; the generality of a standard parser can be maintained, whereas the preference module is tuned to the needs of the individual user and his specific types of text. Defining the preference rules will be an important part of the customization process.

4 Ordering lexical translation alternatives

As an example of a translation ambiguity, we present the set of German equivalents of the Swedish verb *ta bort [remove]* that we found in our experimental text. In all, there are 10 different translations, i.e., *entfernen*, *abnehmen*, *herausnehmen*, *abbauen*, *herausschrauben*, *demontieren*, *ausbauen*, *lösen*, *herausheben*, and *herunternehmen*. The verb is transitive, and, evidently, the distribution of the target language alternatives is determined by its direct object, for instance, *Schrauben herausschrauben; Kupplungsservomechanismus, Kupplungshebel* and *Mutter abbauen*; *Ausrücklager*, *Deckel*, *Dichtung*, *Distanzstück*, *Kupplungshebel*, *O-Ring*, *Dichtring*, *Sicherungsring*, *Traghülse*, *Passscheibe*, *Planeten-getriebebeteil*, *Schaltstangengehäuse*, *Schmierleitung*, *Schraube*, *Sicherungsschraube*, *Traghülse* and *Ölpumpe entfernen*. *abnehmen* takes the same set of objects as *entfernen*, and, in addition to

that, *Kupplungsgehäuse*. These two verbs have the widest use, and hence the most neutral meaning. In all, there are 107 occurrences of *ta bort*, and 57 elliptic uses. *entfernen* covers 90 (58 + 32) cases and *abnehmen* 32 (17 + 15). *entfernen*, being more frequently used than *abnehmen*, will be considered to have the most general meaning, and hence be chosen as the default translation of the verb. Its definition (*entfernen: wegbringen, beseitigen; dafür sorgen dass jmd., etw. nicht mehr da ist*) in Duden (1989) gives further support to this decision. The default translation will be expressed by a simple lexical transfer rule (cf. ii). *abnehmen*, on the other hand, appearing as a more or less absolute synonym of *entfernen*, will be neglected and the remaining translation alternatives be given in context (cf. iv). Due to the specificity criterion, priority will be given to the contextual translations. To sum up, distribution, frequency, and definition provide the general basis for determining default translations in Multra. There is only one default translation for each translation ambiguity, and remaining alternatives are presented to the system by means of phrasal (contextual) transfer rules.

5 Ordering generation alternatives

In **2** we presented the format of the generation rules and the application of the specificity principle to generation by means of an English example, i.e., the generation of a direct object in the indefinite form to be preferred to the definite form. Here we will give one more example of the specificity principle, demonstrating its application to the generation of ellipsis in coordinated clauses in German. The Swedish sentence *Ta bort kopplingsservomekanismen och yttre kopplingshävarmen. [Remove clutch servo mechanism and outer clutch lever.]* is analyzed as a coordinated clause with an elliptic expression of the verb in the second clause, i.e., *Ta bort kopplingsservomekanismen och (ta bort) yttre kopplingshävarmen.* The verb in the first clause is marked '+ surface', and the second one '- surface'. Corresponding to the two possible translations of the verb *ta bort*, a default translation and a translation in context (cf. **4**) four German transfer structures will be presented, based on *abbauen* (the preferred translation) and/or *entfernen*. If the same verb is used in both clauses, an elliptic expression in the first German clause (cf. the Swedish ellipsis in the second clause) must be considered. This can be arranged by means of a generation rule such as the one presented in (x). (x) being more specific than (xi), the default rule for generating coordinated clauses, will be preferred.

(x) %Coordinated clauses; two clauses with a conjunction: same verb

```
LABEL CL.COORD1
X1 ---> X2 X3 X4 X5 X6 :
    <X1 FIRST PRED VERB LEX> = <X1 SECOND PRED VERB LEX>
    <X1 PHR.CAT> = CL
    <X1 FIRST PHR.CAT> = <X1 SECOND PHR.CAT>
    <X1 FIRST MODE> = <X1 SECOND MODE>
    <X1 FIRST PRED SUBJ> = <X1 SECOND PRED SUBJ>
    <X1 FIRST PRED OBJ.DIR DF> = <X2>
    <X2 NG.FEAT CASE> = ACC
    <X1 CONJ> = <X3>
    <X1 SECOND PRED OBJ.DIR DF> = <X4>
    <X4 NG.FEAT CASE> = ACC
    <X1 SECOND PRED VERB> = <X5>
    <X1 SECOND SEP> = <X6>
```

(xi) %Coordinated clauses; two clauses with a conjunction; same or different verbs

```
LABEL CL.COORD2
  X1 ---> X2 X3 X4 :
    <X1 PHR.CAT> = CL
    <X1 FIRST> = <X2>
    <X1 SECOND> = <X4>
    <X1 CONJ> = <X3>
```

If both the transfer and the generation components are tuned for all alternatives, the following translations are generated and presented in the order of appearance below:

Kupplungsservomechanismus und äusseren Kupplungshebel abbauen.
Kupplungsservomechanismus abbauen und äusseren Kupplungshebel abbauen.
Kupplungsservomechanismus entfernen und äusseren Kupplungshebel abbauen.
Kupplungsservomechanismus abbauen und äusseren Kupplungshebel entfernen.
Kupplungsservomechanismus und äusseren Kupplungshebel entfernen.
Kupplungsservomechanismus entfernen und äusseren Kupplungshebel entfernen.

The first alternative corresponds to the model translation. Whether the order between the remaining alternatives is the best one can be discussed. The one presented, however, is the one that is generated when preferences (in terms of rule specificity) are adequatly formulated within each module, but no integration takes place between them. Integrating rule application control between the three modules of the MT system is a major undertaking. It should be motivated only if empirical data supporting a more sophisticated ordering of translation alternatives can be presented. One of the aims of the evaluation of the Multra prototype on site (Saab-Scania AB, Scania Trucks & Buses) is to examine the feasibility of such an effort.

References

Allén, Sture. 1981. *The lemma-lexeme model of the Swedish lexical database.* In B. Rieger (ed.) *Lexical semantics.* Bochum.

Beskow, Björn. 1992. *TransferTool on UNIX: An Introduction.* Dept. of Linguistics. Uppsala University.

Beskow, Björn. 1993a. *Generation in the Multra system.* Dept. of Linguistics. Uppsala University.

Beskow, Björn. 1993b. *Unification Based Transfer.* RUUL 24. Dept. of Linguistics. Uppsala University.

Duden. Deutsches Universalwörterbuch. 1989. Mannheim.

Ingo, Rune. 1990. *Från källspråk till målspråk. Introduktion i översättningsvetenskap [From source language to target language. Introduction to translation theory.]* Studentlitteratur. Lund.

Sågvall Hein, Anna. 1987. *Parsing by means of Uppsala Chart Processor UCP).* In: Bolc, L. (ed.) *Natural Language Parsing Systems.* pp. 203–266. Springer Verlag.

Sågvall Hein, Anna. 1993a, *Multilingual Support for Translation and Writing. MULTRA. Final Research Report. HSFR/NUTEK Language Technology Research Program.* Dept. of Linguistics. Uppsala University.

Sågvall Hein, A., 1993b, *On the Translation of Nominal Expressions in a Unification-Based Multilingual Setting.* In: Hajicova, E. (ed.) *Proceedings of the Conference Functional Description of Language.* Prague.

Sågvall Hein, Anna. In print. *Preference Mechanisms of the Multra Machine Translation System.* In Hall Partee, B. & P. Sgall (eds.) *Meaning and Discourse. Festschrift for Eva Hajicova.* J. Benjamins Publishing Co. Amsterdam/Philadelphia.

Sågvall Hein, Anna. In preparation. *A Computational Grammar of Swedish.*

Svensk Ordbok [A Dictionary of Swedish]. 1986. Stockholm.