On the Translation of Nominal Expressions in a Multilingual Unification-Based Setting

Anna Sågvall Hein
Uppsala university, Uppsala

The work presented is carried out in the project *Multilingual Support for Translation and Writing*, or *Multra* for short. The project aims at the design and implementation of a computerized translation and writing tool (Sågvall Hein 1992). The tool includes an MT function supporting the translation of Swedish instruction text to English, German, and Russian. The Multra MT function is handled by four components, responsible for *analysis, preference, transfer*, and *synthesis*, respectively. The analyser assigns grammatical descriptions to sentence fragments (head lines) and sentences. The preference component inspects them and hands them over to the transfer component, one by one by in a preferred order, if there are alternatives. The transfer component builds target language equivalents of the source language grammatical descriptions by means of unification-based transfer rules (Beskow 1992a,b). Alternative transfer rules are applied according to specificty; a more specific rule takes precedence over a more general one (ibid.). The synthesis component provides the target language specifications in terms of synthesis grammars and dictionaries (formulated in a PATR-like style, see Beskow, forthcom.) and builds the resulting target strings.

A crucial issue in the design of a machine translation system is the *distribution of linguistic knowledge* and *share of labor over* its components. For instance, there is a well-known interrelation between the depth of the analysis and the complexity of the transfer process; the deeper the analysis, the simpler the transfer component, and vice versa. The same kind of interrelation can be found between transfer and synthesis; the more elaborate the transfer structures are, the less the work that will be left over to the synthesis, and vice versa.

In our *multilingual setting*, with a transfer component for each language pair (as a starting point), priority is given to the *simplicity and generality of the transfer components*. Whenever a task or a specification can be referred to the language specific synthesis components, this will be done. In this presentation I will dwell on the feasibility and effects of applying such a strategy to the translation of nouns and NPs. The examples are taken from the instructive text in focus of Multra.

Among the nominal categories, gender and case are internal, language specific, whereas number and definiteness, basically, are external. We may express these views by means of a general transfer rule, (App.: 1) with the effect that the values of the number and definiteness categories will be transfered, whereas source gender and case will be neglected. This is an example of a rule, that will hold for all languages in the multilingual setting; it belongs to the common (general) part of the transfer components.

The expressions of transfered definiteness and number values, as well as the choice of case, are tasks to be handled by the synthesis component.

Gender is inherent to the lexical units, and they are transfered by means of lexical rules. Lexical as well as structural transfer rules are expressions of equivalence relations, established between the lexemes (basic senses) (see e.g. App.: 2). Target gender, being irrelevant to the transfer process, is provided by the synthesis dictionary (see e.g. App.: 3).

Target language *case* values are determined by grammatical function, valency, and syntactic structure. The *top level transfer rules* of Multra specify relations between (orthographic) sentence types (sentences and sentence fragments), and primary grammatical functions.

For instance, the nominal expression *Mätning av axialspel* [En. Measuring of end float], appearing as a head line, will be analyzed as a sentence fragment:

```
['*' =
  ['PHR.CAT' = 'SENT.FRAG',
   'SENT.FRAG' =
    ['NG.FEAT' = ['GENDER' = 'UTR',
```

```
            'NUMB' = 'SING',
            'DEF' = 'INDEF',
            'CASE' = 'BASIC'],
    'HEAD' = ['LEX' = 'MÄTNING.NN.1',
           'WORD.CAT' = 'NOUN'],
    'PHR.CAT' = 'NG',
    'POST.ATTR' =
     ['NOUN.OBJ' =
       ['PREP' = ['LEX' = 'AV1.PP.4'],
        'PHR.CAT' = 'PG',
        'RECT' = ['NG.FEAT' = ['GENDER' = 'NEUTR',
                               'NUMB' = 'SING',
                               'DEF' = 'INDEF',
                               'CASE' = 'BASIC'],
                               'HEAD' =
                                 ['LEX' = 'AXIALSPEL.NN.X',
                                  'WORD.CAT' = 'NOUN'],
                                  'PHR.CAT' = 'NG']]]]]].
```

and transfered as such by means of the general sentence fragment (App.: 4) and phrase category (App.: 5) transfer rules, according to which a sentence fragment is transfered as a sentence fragment, and a source phrase category to the same kind of target category.
The general phrase category rule operates in parallell with the sentence fragment rule and transfers the source language nominal group (NG) into a target language nominal group. The target language case value of the NG corresponding to basic case of the source will be determined by the sentence fragment synthesis rules (En. basic, Ge. nominative, Ru. nominative). (An alternative treatment would have been to assign target case in the transfer rules; if so, two transfer rules would have been required: one for Swedish - English, one for Swedish - German or Russian).

In a similar manner we can handle transfer and synthesis of the post attribute of (i), translating it into English (ii), German (iii), and Russian (iv), respectivly.

(ii) Measuring of end float
(iii) Messung des axialspiel
(iv) Izmerenie osevogo biejnija

A general post attribute rule (App.: 6) transfers the Swedish post attribute into a target post attribute. Further, the transfer rule for handling objects of nouns (App.: 7) is formulated in a manner that holds for the three target languages involved. In the source the object of the noun is expressed by means of a prepositional phrase. The nominal group of this PP (the rection of the preposition) is transfered into the target languages as the value of the noun object attribute. Nothing is said about the structural realization of the object. These specification are left to the Noun object synthesis rules of the three target languages. (En. *of* + noun; Ge. definite NP in the genitive case; Ru. NP in the genitive case).

In the translation of objects of nouns, as in many other cases, there are alternatives. For instance,

(v) Sw. *stukning av fyra (ev sju) kuggar i lamellhuset*
(vi) Ge. *Anstauchen von vier (oder sieben) Zähnen im Scheibengehäuse*
(vii) En. *upsetting four (or seven) teeth in disc housing*

The German translation into a prepositonal phrase (vi) with the preposition *von* is much more rare in our text than the genitive alternative. It seems to be favoured in cases with a "heavy" object or with an object where the genitive has no explicit marking. This translation alternative is isomorph to the Swedish expression and can be formulated by means of a general rule (App.: 8). The application of this general rule has the effect that the prepostional phrase constituting the noun object will be translated "out of context", i.e. by means of the general preposition (App.: 9) and rection (App.: 10) rules. A lexical rule (App.: 11) determines the choice of the preposition. A synthesis dictionary entry presents the case that it governs (App.: 12).

Both (7) and (8) will be applicable to (i) and (vi). However, due to the transfer specificity principle of Multra, (7) being more specific than (8) will take precedence. If we want (8) to be exclusively applied to constructions with heavy objects, this can be achieved by specifying its source part to that effect. If, on the other hand, we want it to apply only to cases with no distinct target language genitive marking, these constraints have to be formulated in the noun object synthesis rule; the synthesis rule will sometimes fail, and control be returned to the next transfer alternative.

In the English direct object translation alternative (vii) of the Swedish noun object (v) the verbal character of the (process) noun seems to have been decisive. We can base these alternatives on parallell lexical rules (13) and (14), one transfering the Swedish noun to an English noun, and one to an English verbal noun.

The lexical transfer rules are, of course, language specific. Examples of problems to be handled are a) source compound versus target loose compound, b) source genitive construction versus target compound or versus c) loose compound etc.

Source		Target
a)	axialspel	Ru. osevoe bienie En. end float
	hydraulmatare	Ge. hydraulische Vorshubeinrichtung En. hydraulic feeder
b)	slangarnas sträckning	Ge. Schlauchstreckung
c)	huvudaxelns axialspel	En. main shaft end float

Lexical as well as structural transfer rules are expressed in the same formalism, consequently, the kind of structural differences illustrated by a) can be easily expressed in the transfer dictionary, for instance:

(14)

```
LABEL
        AXIALSPEL
SOURCE
        <* HEAD LEX> = AXIALSPEL.NN.X
        <* HEAD WORD.CAT> = NOUN
TARGET
        <* ATTR LEX> = OSEVOE.AV.?
        <* ATTR WORD.CAT> = ADJ
        <* HEAD LEX> = BIENIE.NN.?
        <* HEAD WORD.CAT> = NOUN
TRANSFER
```

The structural relation between the attribute and the head of the target (Russian) are transfered to the synthesis component. Gender will be provided by the synthesis dictionary, and constraints such as gender, case, and number agreement be formulated in the synthesis grammar.

For German it is just as important as for Russian that the attribute - head relation is transfered, because also in German a preposed attribute agrees with its head, and thus may be inflected. For English, on the other hand, this is not the case. A simple transfer rule (15) would, in principle, do.

(15)

```
LABEL
        AXIALSPEL
SOURCE
        <* WORD.CAT> = NOUN
        <* LEX> = AXIALSPEL.NN.X
TARGET
        <* WORD.CAT> = NOUN
```

<* LEX> = END FLOAT.NN.?
TRANSFER

However, if we want to use distinctions such as light and heavy nominal expressions for choosing between synthesis alternatives (cf. *measuring of* and *measuring* above), then it is motivated to keep the structural pattern of the nominal for English too.

In the general case, a source NPs consisting of a genitive attribute followed by a head NP translates into an isomorph target construction. However, examples such as b) and c) call for a means for an alternative translation into a compound (Ge.) or a loose compound (En.). In both cases, there is a choice between an individul and a rulebased treatment.

An individual treatment may be realized in two ways; either we transfer via a rule relating the source NP structure to a target lexical element, or transfer is carried out by means of regular transfer rules relating the source NP to an isomorph target NP. In the latter case, synthesis rules account for the generation of the compound alternative. A rulebased treatment of this alternative amounts to introducing synthesis rules transforming the NPs to compounds regardless of their lexical values. An optimal treatment of the phenomenon should be based on such rules in addition to individual synthesis rules, and regular rules for the generation of isomorph constructions. Precedence should be given to the individual rules, whereas so far we lack the evidence to guide an ordering between the compound rules and the isomoprh ones. Handling the phenomenon by the transfer component, either individually or in a rulebased manner is in conflict with our strategy aiming at the generality of the transfer component.

Lexical number differences between source and target, e.g. target language plurale tantum, cause no problems to the transfer component. Source number is transfered to the target, and, accordingly, a translation generated. If the target language equivalent is plurale tantum (or singulare tantum), inherent number is given in the synthesis dictionary; further, number in the synthesis dictionary is given precedence to transfered number. In other words, the problem is handled by the synthesis component and doesn't affect the transfer process.

How to handle other kinds of number shifts as well as shift of definiteness are problems outside the scope of this presentation.

Conclusion.
xxx

References

Beskow, Björn. 1992a. Unifieringsbaserad transfer [Unification-Based Transfer]. Examensarbete. Datalingvistlinjen. xxx

Beskow, Björn. 1992b. Machine Translation in a Unification Based Framework. Presentation vid de Nordiska Datalingvistikdagarna 1991. ?Under utgivning.

Sågvall Hein, Anna, 1992, Aims and Achievments of the Multra Project. June 1992. Department of Linguistics. Uppsala University.
1

Appendix: Examples of transfer and synthesis rules

```
(1) LABEL NG.FEAT SOURCE
        <* NG.FEAT DEF> = ?DEF
        <* NG.FEAT NUMB> = ?NUMB
        <* NG.FEAT GENDER> = ANY
   <* NG.FEAT CASE> = ANY
TARGET
        <* NG.FEAT DEF> = ?DEF
        <* NG.FEAT NUMB> = ?NUMB
TRANSFER

(2) LABEL MÄTNING
SOURCE
        <* WORD.CAT> = NOUN
        <* LEX> = MÄTNING.NN.1
TARGET
        <* WORD.CAT> = NOUN
        <* LEX> = MESSUNG.NN.1
TRANSFER

(3) WORD MESSUNG.NN.1
        <* GENDER> = FEM

(4) LABEL SENT.FRAG
SOURCE
 <* SENT.FRAG> = ?SENT.FRAG1
TARGET
 <* SENT.FRAG> = ?SENT.FRAG2
TRANSFER
  ?SENT.FRAG1 <=> ?SENT.FRAG2

(5) LABEL PHR.CAT
SOURCE
 <* PHR.CAT> = ?PHR.CAT
TARGET
 <* PHR.CAT> = ?PHR.CAT
TRANSFER

(6) LABEL POST.ATTR
SOURCE
  <* POST.ATTR> = ?POST.ATTR1
TARGET
  <* POST.ATTR> = ?POST.ATTR2
TRANSFER
  ?POST.ATTR1 <=> ?POST.ATTR2
```

```
(7) LABEL NOUN.OBJ_PP-NP
SOURCE
  <* NOUN.OBJ PHR.CAT> = PP
  <* NOUN.OBJ PREP LEX> = ANY
  <* NOUN.OBJ RECT> = ?RECT1
TARGET
  <* NOUN.OBJ> = ?RECT2

(8) LABEL NOUN.OBJ
SOURCE
  <* NOUN.OBJ> = ?NOUN.OBJ1
TARGET
  <* NOUN.OBJ> = ?NOUN.OBJ2
TRANSFER
  ?NOUN.OBJ1 <=> ?NOUN.OBJ2

(9) LABEL PREP
SOURCE
  <* PREP> = ?PREP1
TARGET
  <* PREP> = ?PREP2
TRANSFER
  ?PREP1 <=> ?PREP2

(10) LABEL RECT
SOURCE
  <* RECT> = ?RECT1
TARGET
  <* RECT> = ?RECT2
TRANSFER
  ?RECT1 <=> ?RECT2

(11) LABEL
        AV1.PP.1
SOURCE
        <* WORD.CAT> = PREP
        <* LEX> = AV1.PP.1
TARGET
        <* WORD.CAT> = PREP
        <* LEX> = VON.PP.1
TRANSFER

(12) WORD VON.PP.1
        <* CASE> = DATIV

(13 ) LABEL STUKNING.1
SOURCE
        <* WORD.CAT> = NOUN
        <* LEX> = STUKNING.NN.1
TARGET
        <* WORD.CAT> = NOUN
        <* LEX> = UPSETTING.VN.1
TRANSFER

(14) LABEL STUKNING.2
SOURCE
        <* WORD.CAT> = NOUN
        <* LEX> = STUKNING.NN.1
TARGET
        <* WORD.CAT> = VNOUN
        <* LEX> = UPSETTING.NN.1
TRANSFER
```