1986

Harbucke, Annette og Hanne Ruus: Nordiske seminar om maskin-oversættelse side 97-104.

THE DISAMBIGUATION MACHINERY IN EUROTRA TRANSFER

Ole Togeby, Eurotra-DK

In the following I shall describe the disambiguation machinery as it has actually been used in a trasfer experiment between Danish and English. We have in some cases had to make decisions in fields where the Eurotra legislation has not been fixed yet. So the description does not necessarily in all aspects reflect official Eurotra thinking.

A. THE TRANSLATION PROCEDURE

In the present version of Eurotra framework we parse (and go through some intermediate stages) to a source language representation with syntactic, semantic and lexical information. We then transfer the syntactic and the lexical information by translation rules, but we do not transfer any semantic information. In the target language we then generate a representation of the translation of the sentence with syntactic, semantic and lexical information. And from that representation we synthesize (through the intermediate stages) the surface text in the target language. The procedure is shown in the following diagram:

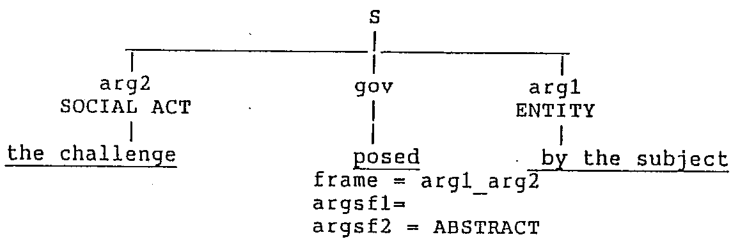
```
source text
    ->
 ________________
|                |
| analysis       |
|________________|

     ->
 ________________        ______________        ________________
|syntactic       |      |Syntactic     |      |syntactic       |
|semantic        |  ->  |              |  ->  |semantic        |
|lexical         |      |lexical       |      |lexical         |
|representation  |      |transfer      |      |representation  |
|________________|      |______________|      |________________|

                                                     ->
                                               ________________
                                              |                |
                                              | synthesis      |
                                              |________________|
                                                     ->
                                               target surface
                                                    text
```

B. THE REPRESENTATIONS

The representation in both source language and in target language is a syntactic dependency tree with *lowered govenor and arguments with semantic decorations* (in CAPITAL LETTERS): Semantic features on the argument nodes raised from the lexical unit, and semantic frame restrictictions on the lexical terminal symbol of the govenor, and lexical terminal symbols (small letters underlined) , eg:



(In this paper I will disregard the problem of word order and passive because it will be solved on the intermediate stages mentioned.)

C. THE SEMANTIC FEATURES

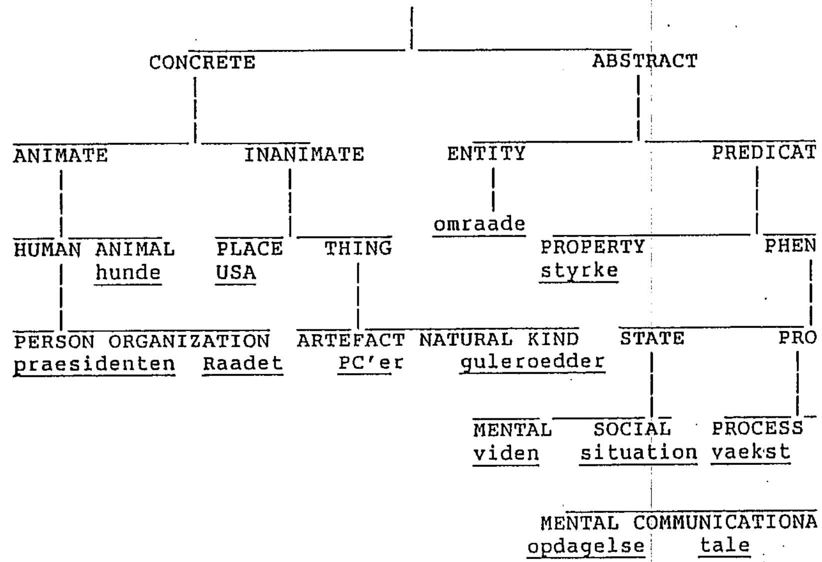
For every lexical entry for nouns, we write the semantic feature of the noun, and for every verb, adjective and for some nouns (viz. the nouns with an argument structure) we write in the lexical entry the frame of the word, viz. semantic feature restrictions on its arguments, e.g. the verb 'pose' has two arguments: arg1 and arg 2, and the second argument has to be ABSTRACT. The lexical description reflects the fact that it is possible to say: 'They posed the question', but not: 'they posed the chair'. It is described:

<u>pose</u>

frame = arg1_arg2,
arg1sf= ,
arg2sf=ABSTRACT.

The semantic decoration in the representation is chosen among a set of semantic features. The following diagram is my own imagination of a proposal for the set of semantic features, but the whole question about semantic features is under discussion in Eurotra for the time being.

SEMANTIC FEATURES FOR NOMINALS



This structure in the feature system means that every noun has to be marked only for one semantic feature, viz one of the bottom features, eg. PERSON or MENTAL ACT; and in the argsfs of the frame of the govenor, it is possible to use all the features in the structure, eg. ENTITY or PHENOMENON. I imagine that the number of bottom features will be twenty or so. The system will in implementation be defined by a set of implication rules of the following sort:

MENTAL ACT --> ACT
COMMUNICATIONAL ACT --> ACT
SOCIAL ACT --> ACT
ACT --> PROPOSITION
PROCESS --> PROPOSITION
and so forth.

D. THE FILTER DEVICE

In the generation of the target language we then have a filter rule, a so called strict a-rule which controls the matching of semantic features of the verb-frame and the arguments, of noun-frame and the arguments, and of the adjective-frame and the arguments.

```
:a/strict:
sfmatch1 = ?.[(gov, {argsf1=SF1, argsf2=SF2, argsf3=SF3}),
              ^(arg1, {sf=SF1}),
              ^(arg2, {sf=SF2}),
              ^(arg3, {sf=SF3}),
              *],
```

^ denotes optionality, * denotes optionality and iteration

The proposed a-rules, the dictionaries and the implication rules will together be sufficient to filter out all the non acceptable constructions .in the target language for which it has been possible to define the distinctive linguistic properties. How it works will be shown in the following:

E. TRANSFER

The transfer rules will be of two sorts: syntactic transfer rules and lexical transfer rules. I here presuppose (but it is not the case) that the syntactic transfer rules only will be a simple mapping from one representation to another representation with the same geometry, what we call simple transfer, e.g.:

```
source:                target:
arg1 gov arg2  ==> arg1 gov arg2
arg1 mod gov arg2 ==> arg1 mod gov arg2
```

The lexical transfer rules cannot possibly be one to one for all words. For the word 'pose' in English, we will get both 'fremsaette' and 'frembyde' in Danish; and the two translations are not interchangeable in all contexts:

formanden fremsatte forslaget
*formanden fremboed forslaget

forureningen fremboed en trussel
*forureningen fremsatte en trussel
Here * denotes not acceptable

So the problem is how to filter out the wrong translations in their respective contexts. It can be done in the generation of the Danish representation. In the Danish dictionary the words are described as follows:

```
fremsaette
     frame= arg1_arg2
     argsf1= HUM
```



```
     argsf2= SOCIAL ACT

frembyde
     frame: arg1_arg2
     argsf1 =
     argsf2 = ABSTRACT
```

On the basis of the material from the transfer rules there are two representational objects which the target language generator tries to generate.

```
                              S
                              |
          ---------------------------------------
          |                   |                  |
        arg2                arg1                gov
     SOCIAL ACT            ENTITY                |
          |                   |                  |
   den udfordring           emnet            frembyder
                                                 frame = arg1_arg2
                                                 argsf1= ,
                                                 argsf2 = ABSTRACT
```

```
                              S
                              |
          ---------------------------------------
          |                   |                  |
        arg2                arg1                gov
     SOCIAL ACT            ENTITY                |
          |                   |                  |
   den udfordring           emnet            fremsaetter
                                                 frame = arg1_arg2
                                                 argsf1 = HUMAN,
                                                 argsf2=SOCIAL ACT
```

But because of the lexical entries and the filter rule one of them is excluded, viz. 'den udfordring som emnet fremsaetter'. The semantic feature restriction in the frame of the verb: argsf1=HUMAN, does not match with the semantic feature of the arg1: sf= SOCIAL ACT.

F. ANOTHER EXAMPLE

This mechanism can be shown on another example where it is the frame of a noun that is the filtering device.

1. REPRESENTATION IN SOURCE LANGUAGE.

```
                              S
                              |
             _________________|________________
            |                 |                |
           arg1              gov              arg2
           ACT                |           ORGANIZATION
            |                 |                |
     _______|______           |                |
    |       |      |          |                |
   mod     gov    arg2        |                |
    |       |   SOCIAL ACT    |                |
    |       |      |          |                |
the main drive2for such a  has come      from industri
                programme
```

2. LEXICAL ENTRIES IN THE SOURCE LANGUAGE.

drive1

frame = arg1_arg2
argsf1 = HUMAN (eg. our)
argsf2 = ARTEFACT (eg. car)
example: 'After a few hours drive in our new car we arrived.'

drive2

frame = arg1_arg2_arg3
argsf1 = HUMAN (eg. industry)
argsf2 = PROPOSITION (eg. efficiency)
argsf3 = HUMAN (eg. the Commission)
example: 'The Commission's drive for efficiency came from industry.'

drive3

frame = arg1_arg2 (only in compounds)
argsf1 = HUMAN (eg. Britain's)
argsf2 = PROPOSITION (eg. export)
example: 'Britain's export drive'

3. TRANSFER DICTIONARY:

drive1 ==> koersel
drive2 ==> initiativ
drive2 ==> tilskyndelse
drive3 ==> fremstoed

Notice that in this example two of the possible translations are filtered out already on the source language side because the word 'drive' is described as 3 different lexical units depending on the syntactic frames. In this example the only possibility in English is 'drive2'.

3. TARGET LANGUAGE DICTIONARY:

initiativ

frame = arg1_arg2
argsf1 = HUMAN
argsf2 = ACT

tilskyndelse

frame = arg1_arg2_arg3
argsf1 =
argsf2 = ARTEFACT
argsf3 = HUMAN

4. TRANSLATIONS:

Following the filter rule and the lexical entries in the target language only two translations will be produced.

```
                                   S
                                   |
               ____________________|___________________
              |                    |                   |
             arg1                 gov                 arg2
             ACT                   |                 HUMAN
              |                    |                   |
       _______|_______             |                   |
      |       |       |            |                   |
     mod     gov     arg2          |                   |
      |       |   SOCIAL ACT       |                   |
      |       |       |            |                   |
initiativet til et saadant er kommet  fra industrien
                   program
tilskyndelsen til et          er kommet fra industrien
                saadant program
```

Of these two translations only the first seems to be correct. It has not been possible to define the distinctive linguistic properties for the two translations, and it is yet unclear how it should be done. So in this example the machine will produce both translations.