AN APPROACH TO SENTENCE-LEVEL ANAPHORA IN MACHINE TRANSLATION

Gertjan van Noord, Joke Dorrepaal, Doug Arnold
Steven Krauwer, Louisa Sadler, Louis des Tombe

Foundation of Language Technology
State University of Utrecht
Trans 10 3512 JK Utrecht

Dept of Language and Linguistics
University of Essex, Wivenhoe Park,
Colchester, C04 3SQ, UK.

November 25, 1994

Abstract

Theoretical research in the area of machine translation usually involves the search for and creation of an appropriate formalism. An important issue in this respect is the way in which the compositionality of translation is to be defined. In this paper, we will introduce the anaphoric component of the Mimo formalism. It makes the definition and translation of anaphoric relations possible, relations which are usually problematic for systems that adhere to strict compositionality. In Mimo, the translation of anaphoric relations is compositional. The anaphoric component is used to define linguistic phenomena such as wh-movement, the passive and the binding of reflexives and pronouns monolingually. The actual working of the component will be shown in this paper by means of a detailed discussion of wh-movement.

Introduction

Theoretical research as part of machine translation often aims at finding an appropriate formalism. One of the main issues involved is whether the formalism does full justice to the idea that the translation of a whole is built from the translation of its parts on the one hand and whether it leaves enough room for the treatment of exceptions on the other hand. In other words, the question is in what way the idea of compositionality is to be defined within a particular formalism. An answer to this question from an interlingual perspective is given in the literature on the Rosetta system (e.g. Landsbergen 1985). The CAT framework (e.g. Arnold et al. 1986) was meant to be an answer to the same question, this time for a transfer system, viz. the Eurotra system. The MiMo formalism is a reaction to the CAT framework and tries to solve several translation problems by formulating an alternative definition of compositionality. Phenomena involving anaphora [1] such as wh-movement and the coindexation of pronominals often cause problems for strictly compositional systems since translation of one word depends on (the translation of) another word, one which can be quite far away in the sentence. Rosetta tackles this problem by distinguishing between rules that are significant with respect to the compositionality of translation, so-called meaningful rules, and rules that are not, referred to as transformations (Appelo et al. 1987); in this way the system is not compositional in the strict sense anymore. The notion of compositionality MiMo adheres to is defined in such a way that anaphoric relations can be translated compositionally as well. In this paper we will introduce the anaphoric component of the MiMo formalism. It is used to define linguistic phenomena such as wh-movement, the binding of reflexives and pronouns, the passive and control phenomena monolingually. The formalism will be discussed by means of an extensive description of a possible analysis of wh-movement.

In the next section, we will first discuss and motivate some of the more fundamental characteristics of the MiMo translation system. Section two will sketch the MiMo formalism as far as necessary for understanding what will follow. The component that deals with the treatment of anaphora will be discussed in section 3. In the fourth section the actual working of the component will be shown by an elaborate discussion of wh-movement. Finally, the translation of anaphoric relations will be defined and some idea will be given of the kind of problems that remain and that will have to be subject to further research.

1 MiMo

The MiMo formalism tries to come up with an answer to the question what compositional translation should imply. Strictly compositional systems have to deal with several translation problems. As to what these problems exactly are depends on the nature of the definition of the notion compositionality. In general, two kinds of problems can be distinguished. First, there are the problems that arise when languages do not really match. Second, the problems that occur when translations of two constructions depend on one another.

The former type of problem is caused by lexical and structural holes. It means that source and target representation do not really match. Lexical holes occur

[1] In this paper the term 'anaphoric' should be interpreted in the broadest sense, as opposed to Chomsky 1981 in which only A-traces and reflexives are called anaphoric.

when a language lacks words equivalent to the ones in the source language. In the case of structural holes, the target language lacks an equivalent construction rather than a word. A description of the concept will have to be used in these cases. For an example of a lexical hole, compare sentence (1) and its translation into English (2).

(1) Jan zwemt graag
(2) John likes to swim

Unlike sentences with an adverb like 'vandaag', (1) cannot be translated compositionally in the strictest sense. The translation of (1) is not simply the translation of the parts of which the constituent is composed of. This problem has been solved in the CAT framework by liberalizing the definition of compositionality in such a way that it will be possible to render (1) directly into (2), by means of a rule like (3).

(3) r1(s1,s2,graag) $\Rightarrow$ r2(t(s1),r3(like,t(s2)))

By (3) a construction composed of three daughters, s1, s2 and 'graag' will be translated into a construction having two daughters, viz. the translation of s1 and a construction that again has two daughters, that is, the verb 'like' and the translation of s2. The main disadvantage of this approach is the fact that combinations of exceptions have to be described explicitly again, see (4) and (5).

```
(4) Jan zwom gewoonlijk
    John used to swim
(5) Jan zwom gewoonlijk graag
    John used to like to swim
```

The translation of 'gewoonlijk' requires a rule similar to (3). However, a combination of 'graag' and 'gewoonlijk' appears to be possible as well. An additional rule will have to account for this. This will lead to an enormous explosion of the number of rules. It is one of the main reasons for an alternative definition of compositionality within the MiMo system. The nature of the definition allows the translation of both 'gewoonlijk' and 'graag' in case they cooccur. A translation rule separates a constituent into an ordinary part and an exceptional part. Both parts are then translated separately and finally, in the target language, the two translated parts are joined again. In the case of a sentence consisting of both 'graag' and 'gewoonlijk', the sentence is separated into an exceptional part, 'graag' for example, and an ordinary part, the rest of the sentence. This rest again is separated into an exceptional, 'gewoonlijk', and an ordinary part. The latter is again that which is left behind after extraction of the exceptional part. In the end, all these parts are joined and will make up a construction in the target language. So, in MiMo not all daughters are translated in one shot

but part of a constituent is translated while the rules can still work on the rest of the constituent. An extensive discussion of problems like these is to be found in Arnold e.a (1988).
The second type of problems w.r.t compositionality in translation involves translation of phrases that are mutually dependent. Examples hereof are translations of phrases that are anaphorically linked. Translation requires that these relations are established. Examples are to be found in (6). In (6), the relation between the subject and the reflexive pronominal is necessary to arrive at the correct form of the reflexive pronominal in French. In (7), knowledge of the functional status of the wh-word is relevant to be able to generate the right case in German.

(6) the women think of themselves $\Rightarrow$
les femmes pensent a elles-memes/*ils-memes
(7) who did you see $\Rightarrow$ wen/*wer/*wem sahest du

In this paper we will examine the component of the MiMo formalism that has been developed to enable the formulation of anaphoric relations on the one hand and compositional translation on the other. The system distinguishes itself from other systems in the field of computational linguistics, such as GPSG (Gazdar et al. 1985), PATR (see e.g. Shieber 1986) and DCG (Pereira and Warren 1980) for its central notion of modularity. The formalism enables the writer of rules to express generalizations in a simple and declarative way. This will be exemplified in section 4. In an MT context, it is however not enough to establish anaphoric relations monolingually. The question is what the behaviour of these relations in translation is. In MiMo, it is possible to translate the relations compositionally. This will be discussed in section 5.

2 The basic model

In this section an overview of the MiMo system will be given as far as is relevant for the rest of this paper. The system's architecture is as in (8). In (8) it is

```
(8) source text
        |         analyse
    source-I
        |         transfer
    target-I
        |         synthese
    target text
```

indicated that a text in a source language is parsed into an interface structure (I). This I-structure, in its turn, is translated into an interface structure in the target language. From this structure the target language text can then be generated. In this paper, mainly the construction of I-structures, through analysis and through transfer, will be focused on, hence the importance of understanding what these structures look like in MiMo terms.
An I-structure is a tree. The mother node consists of the lexical identifier (LI, the name of the lexical element), possibly provided with a set of features, and a number of slots. Slots can be filled with other I-structures that meet the requirements specified by the slots. (9) is an example of an I-structure. The

```
(9)  kiss(verb)       (10) kiss(verb)
      /     \               .[subj=john(n),
   subj(n) obj(n)              obj=mary(n)]
    |        |
john(n) mary(n)
```

I-structure (9) has an LI 'kiss' and two slots, an object slot and a subject slot. Fillers of these slots will have to be nominal. The subject slot has been filled by an I-structure that has 'john' as LI, the object slot by the I-structure with LI 'mary'. We will abbreviate structures like these as in (10) henceforth. So, an I-structure consists of a certain LI, a feature bundle in parenthesis and a number of slots in square brackets preceded by a dot. A slot is made up of the name followed by the equal sign and the I-structure that fills it.
Possible I-structures are defined in the lexicon. Distinct (phrase structure) rules that define I-structures are not needed, all structures are specified in the lexicon. Generalizations should be expressed in the lexicon as well. The advantage of this approach is the possibility of defining all subcategorization phenomena directly. So, only coherent structures in the sense of LFG (Bresnan 1982) are built. In the lexicon, the slots have not yet been filled by other I-structures. The I-structure for 'kiss' looks like (11) in the lexicon, the question marks indicate that the slot are still empty. In (12) the lexical representation of 'john' is given, which has no slots. When an I-structure fills the slot of some other I-structure, the features

```
(11)  kiss(v)            (12) john(n).[]
        .[subj = ?(n),
           obj = ?(n) ]
```

of the slot and those of the I-structure are unified (see e.g. Shieber 1987). The I-structures represented so far were simplified for the sake of readability. In reality, there is the possibility of indicating whether slots are optional or obligatory. Slots can also be marked with the Kleene star. The effect of this operator is that the slot is copied when an I-structure fills the slot. The I-structure will fill the copy and the original slot remains as it was. The slot can be filled several times by I-structures in this way. The slot for modifiers is in fact marked with the Kleene star[2] . An I-structure for (13a) looks like (13b)[3] .

```
(13) a. De mooie vrouw ontmoet mannen op zondag
        The nice woman meets men on sunday
     b. ontmoet(v,present)
        .[subj = vrouw(n,definite)
                 .[mod = mooi(adj).[],
                   *mod = ?()],
          obj = man(n,plural)
                .[*mod = ?()],
          mod = op(prep)
                .[obj = zondag(n),
                  *mod = ?()]
         *mod = ?()]
```

Some words in the lexicon can have the special feature 'anaphor'. I-structures having this feature will have to be bound by an antecedent in the end. Examples of these are pronouns and reflexives. This requirement also holds for empty slots. They are considered anaphoric and will have to be bound as well unless we deal with optional slots. Binding of I-structures happens through anaphoric rules. In the next section we will show the way these rules are formulated. The final structure of (13a) will be (14). In (14), a relation between the topic (I1) and the embedded subject position (I2)[4] is established[5] . The subordinate complementizer is also regarded as a lexical word. Even sentences that do not show a complementizer at surface are assigned one. This is not in any way intrinsic to MiMo but makes a uniform account of several phenomena possible. This type of complementizer has two slots: an optional slot for topics or wh-words and a slot for a verb construction.

[2]This results in a flat structure for modifiers. This is perhaps not correct from a linguistic point of view. However, translation is often much simpler this way. The representation of modifiers is a field in MT that deserves further attention.

[3]Note that the order of slots is quite arbitrary. Surface order is not related to the order of slots in I-structures in any way.

[4]I1 and I2 are unique names which are automatically assigned to every I-structure. We will indicate them henceforth as capitalized words. Names to which no further reference is made will be omitted for clarity's sake. An I-structure consists of a tree and a set of annotations that denote the anaphoric relations within the tree. The tree annotated with this set will be called I-object henceforth.

[5]Note that we will usually leave out optional slots that are not filled

```
(14)
dat(comp)
.[spec = vrouw(n,definite,I1)
    .[ mod = mooi(adj).[]],
  compl = ontmoet(v,present)
    .[subj = ?(n,I2),
       obj = man(n,plural).[],
       mod = op(p)
         .[obj = zondag(n).[]]]],
{ topic_trace(I1,I2)}
```

3 The definition of anaphoric relations

Anaphoric relations are defined by a type of rule that is quite different from the ordinary rules. This distinguishes the system from, for example DCG. With PATR and DCG the possibility of percolation from, say topic to trace, influences all the other rules. MiMo's approach, a separate type of rule for the anaphoric component, has the advantage of leaving the other rules, i.e the lexical I-structures, as they are. Modularity is one of MiMo's qualities. This quality is also considered important in GPSG (Gazdar et al. 1985) where it is realized by the use of metarules that multiply the number of rules. This would be undesirable in MiMo since every lexical word is its own rule. So then even the number of words would have to be multiplied.
The use of a different rule type is also motivated by the process of translating anaphoric relations. If we only used feature percolation to encode anaphoric relations, the relations established would not be explicit anymore. Annotations in MiMo are clearly distinguishable from the rest of the representation and as such make it possible to define a compositional translation of them in transfer. Besides being modular, the system also proves to be declarative. Both qualities, modularity and declarativity, enhance the workability for the user. Changes and extensions are quite easily achieved and rules can be defined in a general way. An anaphoric component written for one particular language can often be used for another language with minor changes.
Anaphoric rules create anaphoric relations within I-structures. This has two consequences in our system. In the first place, some of the features of antecedent and anaphor are unified. These features are called 'transparent'. This, for example, makes it possible to define agreement phenomena. The linguist defines which features are transparent with respect to a certain rule. The motivation for this approach is discussed at length in Krauwer et al. (1987). The main point is that identity of some but not all features is required in an antecedent-anaphor relation. In the second place, the I-structure is augmented with an annotation that specifies the binding. This annotation consists of the name of the relation

and the unique names of the nodes between which the relation exists. The definition of anaphoric relations makes use of these annotations (see also section 5). A relation cannot be created unless the correct structural relation between antecedent and anaphor exists. So the grammar writer defines for each relation:

1) the name of the relation
2) the transparent features
3) the structural relation

An example of an anaphoric rule is the one that establishes a relation between a wh-element and an open slot. The rule looks like (15)[6] MiMo[7] .

```
(15) wh_trace : c_command( {wh}, {open} )-
                {agreement,case}
```

The wh-trace relation is established when the structural relation c_command holds between a wh-constituent and an open slot. The agreement features and the case feature are unified if possible, if not, the relation will not be established. The structural relation itself, c_command in this case, is defined by the user as well. Either a simple structural relation is defined or a complex structural relation. The latter is composed of a regular expression over structural relations[8] . An example of a simple structural relation is the sister-relation, defined in (16).

```
(16) sister(ANT,ANA) :
      ?() .[ ? = ?(ANT),
             ? = ?(ANA) ]
```

```
(17) c_command:
       sister +
       ancestor
```

The structural relation sister holds between the I-structures ANT and ANA if there exists an I-structure in which both ANT and ANA fill slots. The exact nature of the LIs is not important nor are the features or the names of the slots, hence their representation as question marks in (16)[9] . A complex structural relation is defined by means of a regular expression over structural relations. The regular expressions make use of the operators '^', indicating optionality, ';' for disjunction, '*' for iterativity (0, 1 or more times) and '+'. The latter has a special meaning which can best be explained by means of the definition of the c_command relation mentioned in (17). The '+' operator indicates that the sister relation should hold between the antecedent and some intermediate

[6]In fact, the wh-trace relation is subject to more restrictions than c-commandment. We will return to this in section 4.

[7]A special feature 'open' is used to refer to open slots. All slots have this feature by default as long as they are not filled. So, 'open' can be regarded as a feature of the trace since slots not (yet) filled can be considered potential traces.

[8]This idea is partly based on LFG's notion of functional uncertainty. See Kaplan et al. 1987.

[9]Note that the order of ANT w.r.t ANA is not relevant since the order of the slots is not in any way related to word order in the sentence.

node and the ancestor-relation between this intermediate node and the anaphor. The Prolog-variant of (17) is (18). So, the c_command relation holds between the I-structures ANT and ANA when one of ANT's sisters is ANA's ancestor. The MiMo definition of 'ancestor' is given in (19a). The relation is defined in

```
(18) c_command(Ant,Ana) :-
      sister(Ant,X), ancestor(X,Ana).
```

terms of the simple relation 'mother'. The structural relation of the latter is in (19b)[10] . Features can be added to the structural pattern to restrict the range

```
(19) a. ancestor :  * mother
     b. mother(ANT,ANA) :
           ?(ANT).[ ? = ?(ANA) ]
```

of possible relations further. This will be illustrated in the fourth section when we discuss a possible way of treating wh-movement. To conclude this section, we give an example of an I-structure to which (15) applies. (20b) shows the structure before and (20c) after application of (15).

```
(20) a. wat ziet John (what does John see)
     b. dat(comp)
         .[spec = wat(wh).[],
           compl = zien(v)
            .[subj = john(n,third,sing,masc),
               obj = ?(open,acc)]]
     c. dat(comp)
        .[spec = wat(wh,acc,I1).[],
          compl = zien(v)
          .[subj = john(n,third,sing,masc,I2),
             obj = ?(open,acc,I3)]],
        {wh_trace(I1,I3}
```

[10]All I-structures are also their own ancestor according to the definition in (19a). This is the correct result when used in the c_command definition since sisters do c_command one another. In case this is undesirable however, the relation could be defined as follows :
ancestor : mother + * mother
Generally, the correct definition of a relation like c_command depends of course on the use it's being made of in anaphoric rules and on the make up of the I-structures used. The definition above should merely be regarded as an exemplification of the mechanism.

4 WH-Movement

In this section, the actual working of the anaphoric component will be discussed. We will do this by showing how a linguistic phenomenon like wh-movement could be implemented. Note that none of the linguistics in this section follows from the system. The aim of the discussion is to give an idea of the power of the anaphoric component and of the kinds of linguistics that can be put to use. We will first introduce the linguistic environment and present some data from Spanish that reflect some of the surface phenomena caused by the presence of anaphoric relations. The section on the implementation of the wh-relation will argue that and show how surface phenomena of this nature can be handled deterministically.

4.1 Introduction

The wh-trace relation seems the most interesting one because it shows both how general and powerful the mechanism is and how restrictive the rules should be to account for the data. At least the data show in (21) should be accounted for. In

```
(21) a. why do you think John left (ambiguous)
     b. who do you think Bill told me ... Susan
        said _ was ill (unbounded dependency)
     c. *who do you believe the claim that Bill
        saw _ (violation complex NP constraint)
     d. *who do you know whether _ left
        (violation wh-island constraint)
     e. *who did you whisper _ came
        (non-bridge verb)
```

the GB framework (e.g. Chomsky 1981), wh-movement is seen as an instance of the transformation 'move alpha', which respects the subjacency principle. The subjacency principle claims that no rule can relate X and Y in the following structure (22):

```
(22) ..X..[...[.Y..]..] ..
          a   b      where a, b bounding nodes
(23) who [ .. [t [Bill told me [t [Susan saw t
     |   s     ||s             ||s          |
     |_________||_______________||___________|
```

For English, S and NP are assumed to be bounding nodes. Wh-movement takes place cyclically via the comp-positions of the intermediate clauses, leaving behind traces (the so-called comp-to-comp movement). As such, it does not cross more than one bounding node at a time in a structure like (23).

Our discussion of wh-movement in the next section is in accordance with the comp-to-comp movement. Although other approaches, such as direct movement, are feasible too, we will adhere to the comp-to-comp approach. Data from Spanish (Torrego 1984) also seem to support the preference for actual movement from complementizer to complementizer.

(24) Que [dice Juan [que [creian los dos [que [habia pensado Pedro [que [habia aplazado el grupo [el grupo habia aplazado
What says John that thought the two that believed Peter had postponed the group ; that the group had postponed

According to Torrego, inversion is obligatory in all clauses except the lowest. In the lowest clause, inversion is optional. The GB theory accounts for this by claiming that for Spanish S-bar, instead of S, is the bounding node. This predicts that movement in the lowest cycle can take place in two ways as shown in (25). Neither of the two violates subjacency. Assuming that a wh-constituent,

```
(25) a. .... [ wh [        [ o [      t
             s'   s        s'  s
               |______________________|
     b. .... [ wh [        [ o [      t
             s'   s        s'  s
                |____________||_______|
```

or its trace, in comp triggers inversion, the variation in Spanish word-order in the lowest cycle is accounted for.
We will return to these data in the next section. We will argue that these data can be handled by the MiMo-mechanism as well, given the correct rules for the binding of complementizers.

4.2 Implementation

The structural relation for wh-movement should reflect the idea that the wh-constituent may bind across one bounding node at most. Note that, before and after the crossing of this bounding node, it may theoretically cross an unlimited number of nodes that are not bounding. The structural relation that reflects this idea looks like (26b), the wh-trace relation is defined in (26a). The wh-trace relation is established by the structural relation subjacent between a wh-element and an open slot. The definition of the subjacent-relation closely resembles that of c_command. Instead of the relation 'ancestor', a relation 'subj_path' is defined that specifies a path consisting of one bounding node at most. Non-bounding nodes may invervene freely. Subjacency then is not defined as a filter, it is a positive formulation of possible relations. Note that (26) is valid both for

```
(26) a. wh_trace : subjacent(wh,open)-
                   {agreement,case}
     b. subjacent: sister + subj_path
     c. subj_path : *mother({nobounding},{})
                  + ^mother({bounding},{})
                  + *mother({nobounding},{})
```

languages in which S is a bounding node, such as English, and for languages which have S-bar as bounding node. The difference in boundedness will be expressed in the lexicon and the bindings will be established according to the definition of subjacency and given the boundedness of particular nodes[11] .

As has been shown in (25a) and (25b), the trace can always be bound in two ways in languages that have S-bar as a bounding node, provided there are at least two clauses in between the antecedent and the trace. We can make good use of this in MiMo. The Spanish synthesis component can check whether the comp-position of a clause is either filled or bound. If so, the clause is inverted. In this way, the variation in word order in Spanish wh-questions will be quite naturally accounted for.

This leaves us to show that our definition of wh-trace indeed establishes a relation in two different ways between the antecedent and the open position. (27b) shows the MiMo version of the structure in (27a). (27c) indicates the way in which the relation is found without binding the complementizer in the embedded clause. The relation 'sister' holds between the antecedent and the node 'pensado'. Ths node in its turn binds the open position I3, through mother-relations. The movement involves the crossing of one bounding node. (27d) indicates the relation found.

(28) shows that two relations can be found. The GB structure and the MiMo structure are shown in (28a) and (28b) respectively. In (28c1), the relation between I1 and I2 is found and (28c2) shows the one between I2 and I3. Both relations are mentioned in (28d).

In (28), the intermediate empty complementizer-position is bound, hence inversion will take place. In (27) the complementizer is neither filled nor bound, so no inversion in this case. The data are accounted for in quite a natural and linguistically sound way. They are the direct consequence of the definitions of structural relations and they do not have to be generated by some kind of arbitrary inversion mechanism.

[11] The difference between bridge verbs and other verbs is also encoded in the lexicon. Only bridge verbs allow comp-to-comp movement. The generalization might be expressed by assigning the feature bounding to sbar complements and modifiers in all other cases. Like this, sbar is a bounding node in some cases too.

```
(27)
a. [s'  wh [s     [s'  o  [s    t
        |______________________|
b. [que t [pensado P. [que t [aplazado grupo t
      (I1)               (I2)             (I3)
       |__________________________________|
c. wh_trace: subjacent(wh,open) -
         { person,number,gender,case}
   subjacent: sister(open(wh,I1),pensado)
   subj_path: mother(pensado(nobounding),que())
        + mother(que(bounding),aplazado())
        + mother(aplazado(nobounding),open(I3))
d.  { wh_trace(I1,I3) }
```

The translation of anaphoric relations

In this section, we intend to give an impression of the usefulness of coindex relations in translation and the translation of the relations themselves. In linguistics, a monolingual account of coindexation is quite an achievement. In machine translation, the most important part of research deals with the translation of the relations that were established monolingually.
The I-object to be translated consists of an I-structure annotated with anaphoric relations. An I-object is the result of the application of certain anaphoric relations (denoted by the annotations) to a particular I-structure. The compositional translation of an I-object is the result of the application of the translated annotations to the translated I-structure. We hold the view that anaphoric relations are universal in MiMo. The translation of a relation between the I-structures I and J is that same relation between the translations of I and J. This is summarized in (29).

(29) the translation of an I-object:
The translation of an I-object I1 is the result of the application of the translations of the annotations of I1 to the translation of I1's I-structure. The translation of an annotation R(I,J) is R(t(I),t(J)).

The final set of anaphoric relations of the target object should be equivalent to the set that existed at the source level. The following example illustrates principle (30) :

(30) Por que [dice Juan [que [los dos creian [que [Pedro habia pensado [que [el grupo habia aplazado la reunion
Why say John that the two thought that Peter believed that the group post-

```
(28)
a.  [s'  wh [s       [s' t  [s      t
              |_______________||__________|
b. [que t [ pensado P [ que t [ aplazado grupo t
      (I1)                (I2)               (I3)
       |___________________||__________________|
c.1. wh_trace: subjacent(wh,open)-
               {person,number,gender,case}
     subjacent: sister(open(wh,I1),pensado)
     subj_path: mother(pensado(nobounding),que())
              + mother(que(bounding),open(I2))
  2. wh_trace: subjacent(wh,open)-
                {person,number,gender,case}
     subjacent: sister(open(wh,I2),aplazado)
     subj_path: mother(aplazado(nobounding),
                open(I3))
d. {wh_trace(I1,I2),wh_trace(I2,I3)}
```

poned the meeting

Inversion being obligatory in all clauses except the lowest, 'por que' can only bind the modifier position in either the first or the second clause. Each relation further down is excluded as more clauses would have to show inversion then. When we ignore the bindings established at the Spanish I-level, translation into English will produce a lot of possible translations since 'that' may or may not be inserted in every complementizer position in English. However, the impact of this complementizer on possible anaphoric relations is not totally irrelevant. According to WAHL (1987), the complementizer blocks binding of 'why' to an empty position deeper down, cf. (31) and (32).

(31) why(i)/(j) do you think _(i) the boat sank _(j)
(32) why(i) do you think _(i) that the boat sank _

When we preserve the bindings from Spanish and we claim that in English 'that' may never be inserted when its modifier position is bound to an antecedent, we can deterministically arrive at the right translation :

(33) Por que [dice Juan [que [los dos creian [que [Pedro habia pensado [que [el grupo habia aplazado la reunion
(34) Why [did John say [[the two thought [that [Peter believed [(that) the group had postponed the meeting

Both are ambiguous since both can question the reason for John's 'saying it' and 'the two believing it'. Other interpretations are excluded in both Spanish and English.
Definition (29) also causes some problems. Take the following example from Italian (cf. Chomsky 1981) :

(35) l'uomo [che mi domando [chi abbia visto]]
the man(i) of whom I wonder who(j) e(i) saw e(j)

One might wonder what the English translation would have to be in the first place. In MiMo, the incorrect literal translation will not be found because the necessary anaphoric relations cannot be established. In cases like these, separate translation rules are needed to arrive at a translation of (35). It is possible to refer explicitly to anaphoric relations as long as they are restricted in depth. This is necessary in case an expression without anaphoric relations translates into one which requires a linking between an antecedent and an anaphor. An example is (36).

(36) Jan zwemt graag ⇒ John(i) likes _(i) to swim

Unboundedly deep embedded relations are however not accessible by translation rules in the transfer component. Another problem we face deals with the interaction of anaphora and other standard 'non-compositional' phenomena, such as the example of Dutch 'graag' translating as 'to like' in English (see section 1). These examples, as well as anaphora, can be handled compositionally, as we have shown. The interaction however poses some problems, see (37).

(37) Hoe graag zwom Jan ⇒ How much did John like to swim

Since 'graag' is displaced, translation of 'graag' as the exceptional part of the embedded sentence is not possible, given that the movement is not undone[12] . These cases are even noncompositional from MiMo's tolerant view on compositionality.

Conclusion

In this paper we showed the need for a non-standard notion of compositionality in translation. With the MiMo definition of compositionality we are able to define the translation of sentence level anaphora. In MiMo, anaphoric relations

[12]It is of course also possible to assume that all wh-movements have been undone. In Mimo, this only means a shift of problems from the transfer to the analysis and synthesis modules. Besides, the issue would still hold for other long-distance phenomena like pronouns.

are defined by a separate type of rule. This enables linguists to define anaphoric relations in a declarative and modular way. It appeared that linguistic generalizations can be defined quite naturally and generally. It is up to the linguist to decide which generalizations are to be preferred and how they can best be expressed. We chose to formulate principles in a general way. The relation 'subjacent' was meant to serve all languages. Restrictions, e.g. by semantic features, can be added freely. The definitions relate to information that is encoded in the language-specific lexicon. This produces the variations that exist across languages.
The use of a separate type of rule enables a compositional definition of the translation of anaphoric relations because the applied rules are still visible - as annotations - in the structure to be translated. The translation of an I-object was defined as the translation of the I-structure to which the translations of the anaphoric rules applied. The translation of an anaphoric rule is the target equivalent of that rule. This point of view poses problems in cases where the source language is less restrictive than the target language. In that case, special rules have to be written to assign a translation nonetheless. When a particular relation (read also : interpreation) has been established in the source language, it should be present in the target language. All interpretations should be translated of course. This is not yet possible in the current system when unboundedly deep relations need to be seen in the transfer component.

Acknowledgements

The work we report here had its beginnings in work within the Eurotra framework. MiMo however is not "the" official Eurotra system. It differs in many critical respects from e.g Bech & Nygaard (1988). MiMo is the result of the joint effort of Essex, Utrecht and Dominique Petitpierre from ISSCO, Geneve. The research reported in this paper was supported by the European Community, the DTI (Department of Trade and Industry) and the NBBI (Nederlands Bureau voor Bibliotheekwezen en Informatieverzorging).

References

L Appelo; C Fellinger; J Landsbergen, 1987: "Subgrammars, Rule Classes and Control in the Rosetta Translation System" in: *European Chapter ACL.* 1987 Copenhagen.
D Arnold; S Krauwer; M Rosner; L des Tombe; G Varile, 1986: "The CAT framework in Eurotra: A theoretically committed notation for MT". in: *Proceedings of Coling.* Bonn 1986.
D Arnold; S Krauwer; L des Tombe; L Sadler, 1988: "Relaxed compositionality in Machine Translation". in: *Second International Conference on Theo-*

retical and Methodological issues in Machine Translation of Natural Languages Carnegie Mellon Univ. Pittsburgh 1988.
A Bech; A Nygaard, 1988: The E-framework: a formalism for natural language processing". in: *Proceedings of Coling.* Boedapest 1988.
J Bresnan (ed) 1982: *The Mental Representation of Grammatical Relations.* Cambridge MIT press. 1982.
N Chomsky 1981: *Lectures on Government and Binding.* Foris Dordrecht, 1981.
G Gazdar; E Klein; G Pullum; I Sag, 1985: *Generalized Phrase Structure Grammar.* Blackwell Publishing and Cambridge Mass. 1985.
R Kaplan; J Maxwell; A Zaenen, 1987: "Functional Uncertainty". *CSLI Monthly* vol 2, no 4 january 1987.
S Krauwer; M King (eds), 1987: *The Eurotra Reference Manual 3.0*
J Landsbergen, 1985: "Isomorphic Grammars and their use in the Rosetta Translation system". in King, M (ed) *Machine Translation Today* Edinburgh university press 1985.
F Pereira ; D Warren, 1980 : "Definite Clause Grammars for Language Analysis - A Survey of the Formalism and a Comparison with Augmented Transition Networks". *Artificial Intelligence 13* F Pereira ; S Shieber, 1987: *Prolog and Natural Language Analysis.* CSLI 1987.
S Shieber, 1986: *An introduction to unification based approaches to grammar.* CSLI 1986.
E Torrego, 1984: "On Inversion in Spanish and Some of Its Effects", *Linguistic Inquiry* 15, 103-130.
WAHL, 1987: J Aoun; N Hornstein; D Lightfoot; A Weinberg: "Two types of locality" *Linguistic Inquiry* 18, 4.