

[Proceedings of a Workshop on Machine Translation, July 1990, UMIST]

Complex Transfer in EUROTRA

Lorna Balkan

Department of Language and Linguistics

University of Essex

Wivenhoe Park

Colchester, UK

July, 1990

COMPLEX TRANSFER IN EUROTRA

Lorna Balkan

Introduction

This paper discusses some translation problems for Eurotra, the transfer-based machine translation project of the European Economic Community. While some of the problems and the solutions adopted are specific to the Eurotra formalism, other problems are common to all transfer-based (as opposed to interlingual) systems. The paper is in four parts

1. Overview of Eurotra
2. Definition of complex transfer in Eurotra
3. General strategies for dealing with complex transfer in Eurotra
4. Reasons for complex transfer in Eurotra

1. Overview of Eurotra

We began by noting that Eurotra is a transfer-based machine translation system. In transfer systems a source language sentence is analysed into a source language representation called an interface structure (IS) which is then mapped onto a target language IS representation, from which the target language sentence is generated:

source text...source IS -> target IS...target text

In interlingual systems, on the other hand, the target sentence is generated directly from the source language IS.

There are three stages to the translation process in transfer-based systems: analysis, transfer and synthesis. Transfer is where the source IS is mapped onto the target IS. In Eurotra, analysis and generation are performed in a series of steps since it is difficult to go from surface structure to IS in one single operation. The levels are as follows:

EMS: Eurotra Morphological Structure

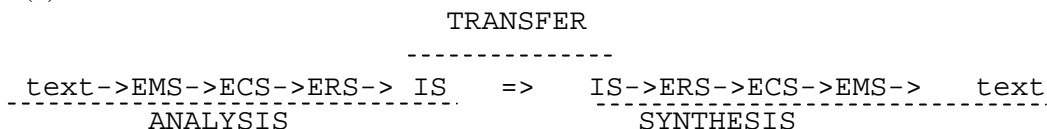
ECS: Eurotra Configurational Structure

ERS: Eurotra Relational Structure

IS: Interface Structure

This can be shown diagrammatically:

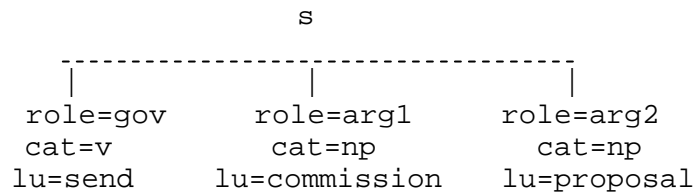
(1)



Both ERS and IS are lowered-governor dependency grammars. There are two sorts of dependents - arguments or complements and modifiers. The complements at ERS represent surface syntactic functions (subject, object, etc) and the arguments at IS represent deep syntactic functions. Importantly for the discussion that follows, syntactic category information is preserved at IS.

At each level of Eurotra linguistic information is represented as a tree structure, where the nodes of the tree are feature bundles and contain information in the form of attribute value pairs. For example the (simplified) IS representation for the sentence *the commission sent the proposal* looks as follows (where “lu” stands for “lexical unit”):

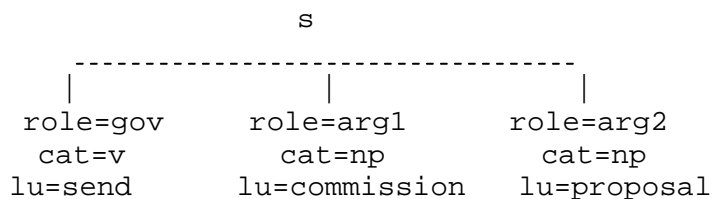
(2)



To go from one representation to another, a tree transducer is applied which either keeps the structure of the source tree or maps the source structure onto a different target structure by, for example, inserting or deleting elements. (Notice that this transduction takes the source tree to the target tree in a single step, unlike "transformational" systems, where a sequence of transductions may apply, taking input from earlier transductions and putting output to later transductions). The information that is contained in the nodes can also be copied or changed. In transfer, the interface between two languages, it is important to keep the transformations as simple as possible (since there is a total of nine languages in the project, which makes 72 transfer modules). In the ideal case, transfer will consist just in exchanging the “lu” value on the leaf nodes, as in the example below:

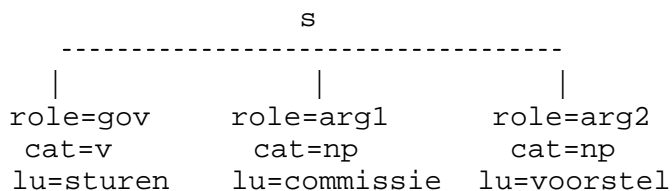
(3) The commission sent the proposal

Source IS:



=>

Target IS(Dutch):



2. Complex transfer in Eurotra

Transfer in Eurotra is characterised as follows (see Schmidt 1990)

1. structural: transfer of the tree without the terminals("lu"s)
2. lexical: transfer of the leaves

Both structural and lexical transfer can be simple or complex: (again, these definitions are according to Schmidt 1990)

1. simple: transfer is simple if a structural representation is copied without performance of a tree transformation or if a node in a tree which is a terminal is substituted by another terminal, providing that in either case attribute/value pairs (other than lu) are simply copied.
2. complex: transfer is complex if either (i) an IS representation has to undergo a tree transformation or if a leaf in this representation has to be expanded into a structure or structure has to be condensed into a leaf; OR (ii) if attribute/ value pairs have to be changed.

We shall look below at some of the causes of complex transfer in Eurotra, and consider some suggestions for how they can be avoided, simplifying transfer. First of all, however, we shall note some general strategies for solving complex transfer in Eurotra.

3. General strategies for dealing with complex transfer in Eurotra.

There are essentially two main approaches to solving complex transfer in Eurotra:

1. Give linguistic phenomena a structural representation such that simple structural transfer can take place
2. Give phenomena a semantic (interlingual) labelling such that only features have to be copied.

4. Causes of complex transfer in Eurotra

Before considering examples of complex transfer in Eurotra let us look at the conditions which must hold for simple lexical transferee, the exchange of lu values only) to take place. Arnold (1989) describes these conditions as follows:

The most straightforward of "simple" [transfer] cases involve items where the translation relation is unambiguous, and the items concerned have equivalent meanings, distributions and internal structures".

Now a lexical entry in Eurotra contains features relating not just to the lexeme itself but to any arguments it may have. It states, for example, the number and syntactic realisation of its arguments. For instance, the (simplified) lexical entry for the verb "send" is:

(4) {lu= send, cat= v, frames=arg12, cat_of_arg2= np}.

where "frame" indicates how many arguments the lexeme has (in this case two), and where "cat_of_argX" indicates the category of argX.

Complex transfer can arise where there is:

- 1) a change in the value of a feature relating to the lexeme itself, eg. a change in the "number" feature in:

(5) telecommunicatie(sing) (Dutch) -> telecommunications(pl)(Eng)

OR

2) a change in value of a feature relating to one or more of the lexeme's arguments (where "pform_of_arg1" indicates the preposition which the adjective governs).

(6) content(pform_of_arg1= de)(Fren) -> happy(pform_of_arg1= with)(Eng)

Although these are instances of complex transfer they present relatively little problem, since they can be stated in the bilingual dictionary where correspondences between lexemes are expressed. There is, however, one exception, and that is where there is a change in the category feature between lexemes.

(7) tal(cat=demonstrative)(Port) -> such(cat=adjective)(Eng)

This involves structural transfer, because not only does a demonstrative have to be transformed into an adjective, but a demonstrative phrase has to be transformed into an adjectival phrase.

Below is a list of some other causes of complex transfer in Eurotra.

4.1. Difference in Argument Structure

Differences can occur in:

- a. argument number
- b. indexation of arguments
- c. the syntactic realisation of arguments

a. difference in argument number.

This is fairly rare, since the number of arguments can usually be decided "euroversally".

For example, whether the word "well" in "projects sell well" should be interpreted as an argument or a modifier can be decided uniformly within the project.

A case where there appears to be a difference in argument structure is "intrinsically pronominal verbs".

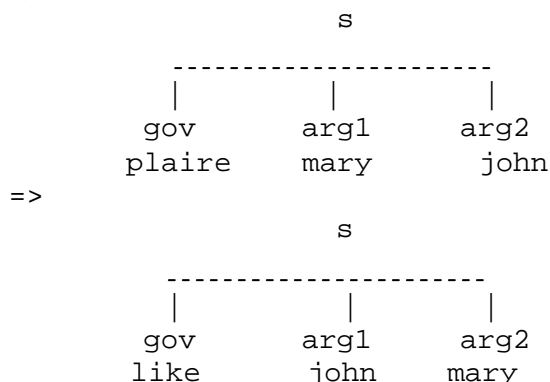
(8) s'evanouir(Fren) => faint(Eng)

("S'evanouir" looks like a 2-argument verb, while "faint" looks like a 1-argument verb). They are handled by treating the reflexive pronoun as part of the lu at IS, ie. "se_evanouir", so that complex transfer can be avoided.

b. difference in indexation of arguments.

An example is the well known pair "plaire/like", where the arg1 of "plaire" becomes the arg2 of "like":

(9)



The problem here lies with the syntactic nature of the Eurotra IS structure. A more semantic IS, where the arguments would represent semantic roles, rather than deep syntactic functions, could overcome this problem. Semantic relations for arguments are a research topic within the project (see Steiner et al. 1989). For the time being, however, cases such as these remain a source, albeit a fairly rare source, of complex transfer.

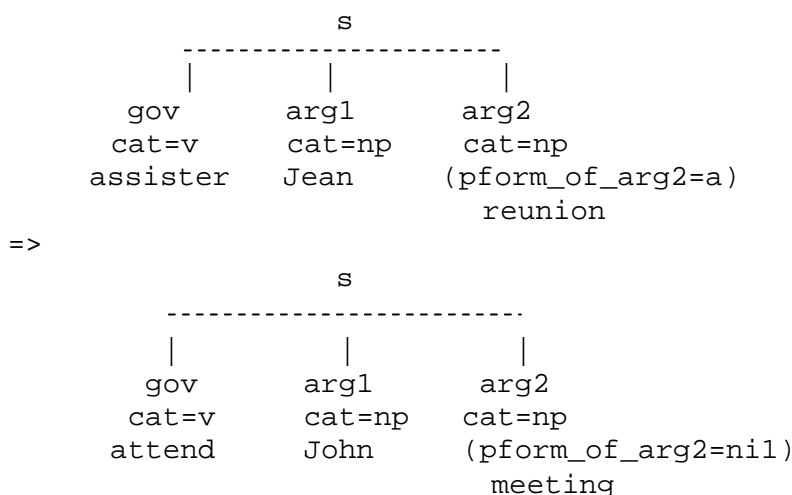
c difference in syntactic realisation of arguments.

For example, a prepositional verb in one language translates into an ordinary verb in another language

(10) assister a + obj(Fren) => attend + obj(Eng)
V + PP => V + NP

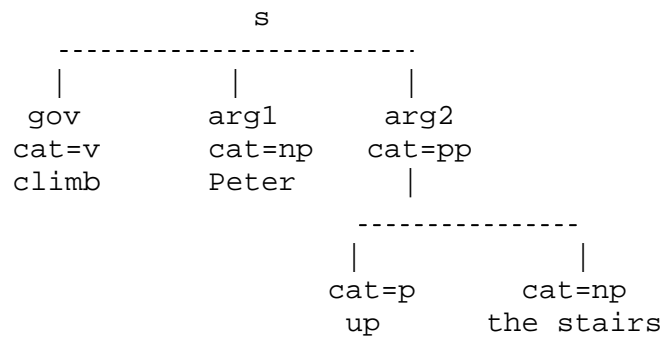
Here the syntactic difference between the arguments can be neutralised by elevating the preposition and in effect converting the PP into an NP. This renders transfer less complex (there is still the problem that the "pform_of_arg2" value will be different), i.e.

(11)



This solution is only adopted where the preposition is said to be strongly bound (i.e. idiomatic or semantically empty). In cases where the preposition is semantically contentful, the preposition is explicitly represented

(12) Peter climbed up the stairs



In this case, complex transfer cannot always be avoided, because it may be that another language expresses the same concept as the prepositional verb with a simplex verb

(13) climb up(Eng) -> monter(Fren)

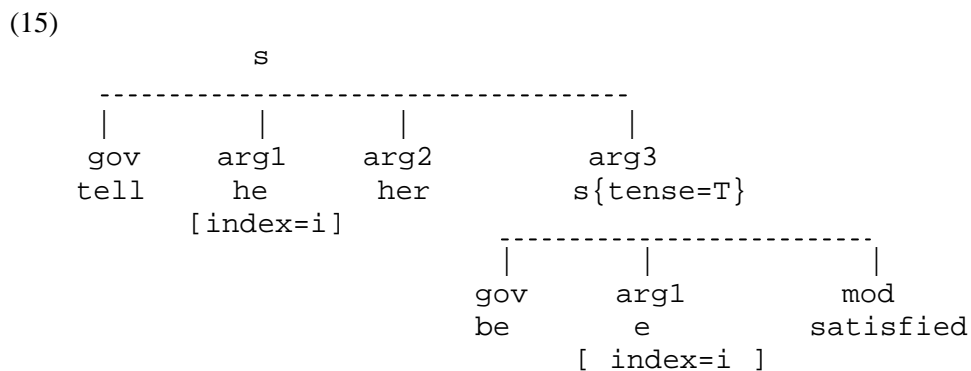
Other cases of different syntactic realisation of arguments are not so straightforward:

(14) il lui dit etre satisfait (Fren)
 he tells her that he is satisfied (Eng)

An infinitive clause in French is transformed into a tensed subordinate clause in English. Note that synthesis should insert:

- a. the appropriate pronoun
- b. the correct tense of the subordinate clause

There is currently no solution to this case of complex transfer within the project, but Durand et al. (1990a) give a sketch of what a solution might look like:



4.2. Lexical Holes

This usually entails a single lexical item being translated into a multi word unit.

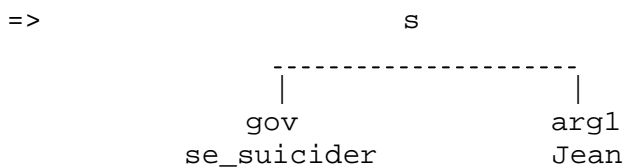
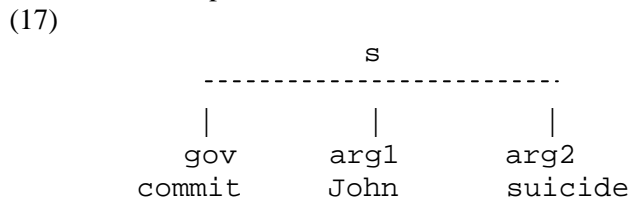
4.2.1. verbs

There are six subcases.

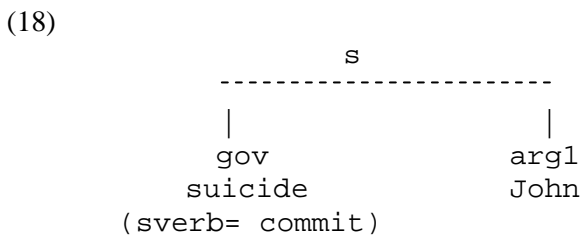
- (1) Where a simple verb in one language has no equivalent simple verb in another language it is often possible to translate it into a so-called support verb construction:

(16) V -> Sverb + N
 se_suicider(Fren) -> commit suicide(Eng)

This causes complex transfer:



A solution being considered within Eurotra is to elevate the support verb and make the noun into the governor(see Danlos 1990) Below is a very simplified representation of their proposal)



Justification for this approach is the fact that the support verb is relatively empty of meaning, while it is the following noun that carries the bulk of the meaning and determines the nature of the arguments of the construction. Notice that transfer would still be complex, since a verb would map onto a noun, but at least no structural transfer would be involved. (Research is also going on within the project on the possibility of eliminating categories at IS).

- (2) Simple verbs often also translate into a copular construction:

(19) V-> Be + Adj
 bastar(Span) -> be enough(Eng)

Investigation is currently being carried out within Eurotra as to whether the treatment proposed for support verbs could be extended to the copula, since the verb "be" is, like the support verb, relatively bereft of meaning in these cases. However, for the time being, they are still a source of complex transfer.

- (3) Other simple verbs require a full paraphrase:

(20)

madrugar(Span) => get up early(Eng)
stomp(Eng) => andar con paso fuerte(Span)
drive(Eng) => aller en voiture(Fren)

Verbs of motion fall into this category:

(21)

run into(Eng) => entrer en courant(Fren)
stomp into(Eng) => entrer d'un pas lourd(Fren)

- (4) Sometimes a simple verb in one language has a multi word translation in another language because an affix in one language has no corresponding affix in the other language. This may be because a morphological process in one language is more productive than in another language

(22)

misunderstand(Eng) -> missverstehen(Ger)
but miscalculate(Eng) -> *missberechnen(Ger)
= falsch berechnen(Ger)

in which case there is a "morphological hole" in one language. Or it may be that the morphological process simply does not exist in the other language.

(23)

fortschreiben(Ger) -> continue writing(Eng)
zurueckgeben(Ger) -> give back(Eng)

Either way, it is a source of complex transfer.

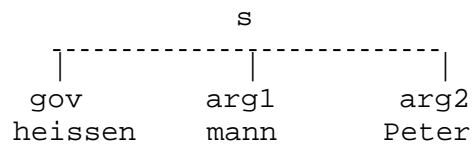
- (5) Another pattern sometimes found is when a verb translates into an adverb

(24) he likes swimming(Eng) -> er schwimmt gerne(Ger)

Again this is a current source of complex transfer.

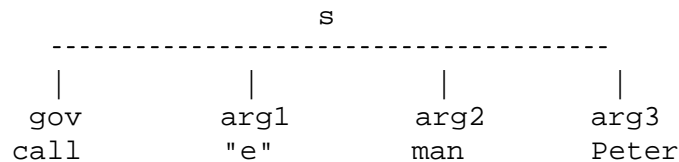
- (6) A case of a lexical hole that does not have a multi word translation is the German verb "heissen" which has no equivalent in English, and which has to be translated as "be called". This causes complex transfer, since the voice of the verb changes from active to passive, and in Eurotra, the empty subject of a passive has to be represented explicitly:

(25) "Der Mann heisst Peter"



=>

"The man is called Peter"



(where "e" is the empty node)

4.2.2. Other Categories.

nouns:

(26) N -> N + Adj
 focus(Eng) -> punto focale(Ital)

As with verbs, the reason that a noun receives a multi word translation may lie with a non-parallelism of morphological processes. For example, English has no equivalent of the German prefix "Ur" which generally translates as "original"

(27) Urtext(Ger) -> original text(Eng)

adverbs:

(28) Adv -> PP
 hierzulande(Ger) -> in this country(Eng)
 repeatedly(Eng) -> a plusieurs reprises(Fren)

adjectives:

(29) Adj -> PP
 lebenswichtig(Ger) -> di importanza vitale(Ital)

Again, the source of complex transfer may be a morphological hole (30), or or the total lack of a morphological process(31):

(30) reproachfully(Eng) -> *reprobateurement
 (d'un air de reproche)

(31) cordless(Eng) -> sans fiche(Fren)

(French has no analogue of the suffix "less")

prepositions:

- (32)
- | | | |
|----------------------|----|-------------------------|
| Prep | -> | PP |
| conformement a(Fren) | -> | in accordance with(Eng) |

All the above cases are examples of complex transfer. Note, however, that the last case can be dealt with by regarding the PP as a complex preposition that is lexically one unit at IS, as for example "in_accordance_with", so that transfer in this case would be simple.

There are a couple of points to note about lexical holes. Firstly, as Arnold (1989) notes, cases of lexical hole present a problem for transfer systems, not only because translation equivalents have different internal structures, but because they have different "external" properties or distributions. Thus it cannot be guaranteed that if "x" translates as "y" in context A it will translate as "y" in context B. For example compare (33a) with (33b) and (34a) with (34b).

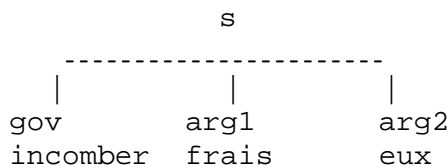
- (33)
- a. the man is hungry(Eng) -> l'homme a faim(Fren)
 - b. the hungry man(Eng) -> l'homme qui a faim(Fren)

(because adjectives (eg. "hungry") have different distributions to nouns (eg. "faim"))

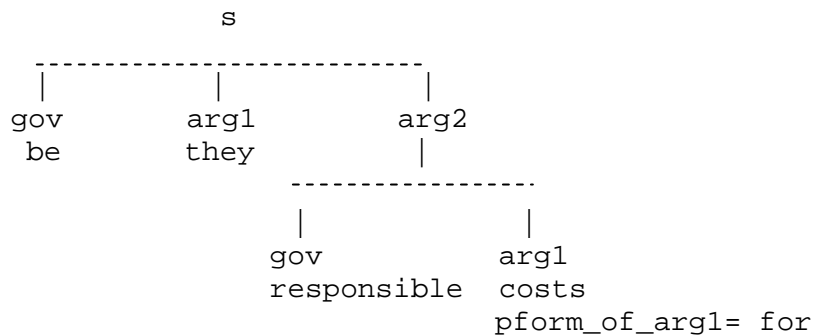
- (34)
- a. he grew tall(Eng) -> il a grandi(Fren)
 - b. he grew very tall(Eng) -> *il a tres grandi(Fren)

Secondly, there are cases where complexity of transfer is due not only to a lexical hole, but to different indexation of arguments.

- (35)
- "les frais leur incombent"(Fren)



=> "they are responsible for costs"(Eng)



4.3. Structural Holes

This is where a grammatical construction in one language has no equivalent in another language. These cases have received little attention in Eurotra, and the examples quoted below are from Arnold (1989).

(36) "people who I am bigger than"

has no direct translation in French:

(37)

*les gens que qui je suis plus grand

*les gens qui je suis plus grand que

but must be translated as follows:

(38) les gens qui sont moins petit que moi.

For resultatives, which have no equivalent in French, Arnold (1989) suggests a general solution. The specific example

(39) he hammered the metal flat(Eng) ->
il martela le metal jusqu'a ce qu' il est devenu plat(Fren)

can be generalised to

(40)

N1 V N2 Adj -> N1 V N2 jusqu'a ce que N2< devenir> Adj

Again, the tense of "devenir" must be worked out.

4.4. Difference in Word Structure

Under this category is the problem of compounds. While German, for example generally concatenates words to build larger words, English prefers to juxtapose them, so that the following two patterns are frequently found in German -> English translation:

(41)

N -> N + N

Wiederaufbauphase -> reconstruction period

(42)

N -> Adj + N

Nachkriegsphase -> postwar period

Romance languages, on the other hand, usually make the relationship between the elements more explicit by means of prepositions:

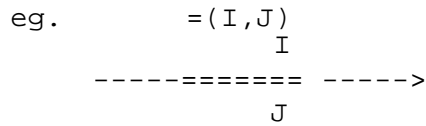
(43)

N -> N + Prep + N

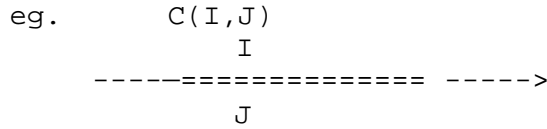
Kooperationsprogram -> programa de cooperacion

A suggestion has been made within Eurotra to represent all compounds structurally, (see Badia et al, 1990) For example, the expression "car production" could receive the following representation:

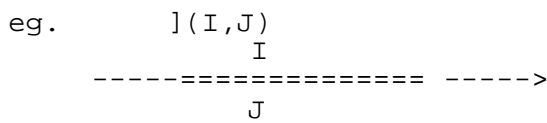
= identity



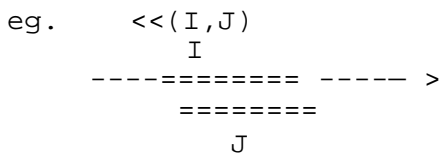
C is a part of



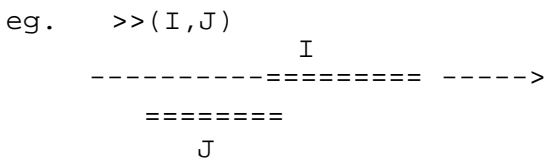
] contains



<< left overlap



>> right overlap



Since the time of speech is taken to be a moment of time, rather than an interval, the number of possible values for tense reduces to three:

- simultaneity =(R,S)
- anteriority <(R,S)
- posteriority >(R,S)

where "R" stands for time of reference and "S" stands for time of speech.

Since no distinction is made between identity and is_a_part_of relation in aspect, there are six values for aspect:

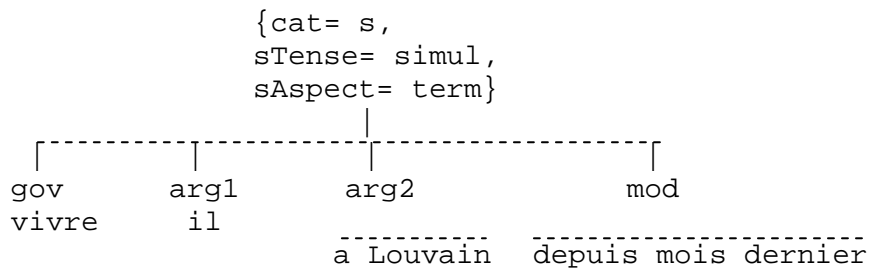
perfectivity	=C(E,R)
durativity] (E,R)
retrospectivity	<(E,R)
prospectivity	>(E,R)
terminativity	<<(E,R)
inchoativity	>> (E,R)

where "E" stands for the time of event.

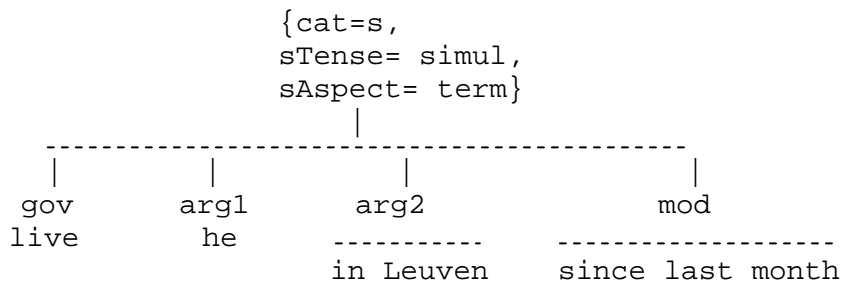
The tense and aspect features of the sentence are calculated from the tense and aspect features of the verb and any temporal modifiers, and the features and their values are copied over to the target IS, thus eliminating the need for complex transfer.

(48)

Il vit a Louvain depuis le mois dernier:



=> He has lived in Leuven since last month.



4.6. Modality

A start to giving modality an interlingual treatment also has been made within the project, but is confined as yet to modal verbs (see Durand 1990b). The problem with modal verbs is that there is not always a one to one correspondence between modal verbs in different languages:

(49) il doit faire chaud en ete(Fren)
 -> it must be hot in summer(Eng)

(50) il ne doit pas faire chaud ici en hiver(Fren)
 -> it cannot be very warm here in winter(Eng)
 *it mustn't be very warm here in winter

Two main types of modality are distinguished in Eurotra:

1. Epistemic, which can be further subdivided into:
 - necessity
 - probability
 - possibility

2. Deontic.

The value of the modal verb is calculated in analysis and copied as a feature onto the target IS. It is left to the synthesis grammar to work out the form of the target modal.

Unlike auxiliary verbs however, which disappear at IS, the modal cannot be elevated because

1. modals can be modified
2. modals can be stacked

(51) il doit pouvoir le faire

3. scope relations have to be expressed

(52) il ne peut pas ne pas le faire(Fren)
He cannot not do it(Eng)

4.7. Periphrastic Constructions

A periphrastic construction is a phrase headed by a verb with an auxiliary function and which endows the phrase with a particular modal or aspectual value.

The translation of the verb can sometimes be an adverb:

(53) suelen venir por la tarde(Span) ->
they usually come in the evening(Eng)

(54) il a failli tomber(Fren) -> he almost fell(Eng)

(55) despues de leer tanta linguistica, llego a gustarle(Span)
-> after reading so much about linguistics, he finally liked it(Eng)

(56) le gusto tanto que llego a comprarlo (Span)->
he liked it so much that he even bought it(Eng)

(57) no volvere a verla jamas(Span) -> he will never see her again(Eng)

(58) il vient d'arriver(Fren) -> he has just arrived(Eng)

The translation of the verb can also be an aspectual variant of the copula:

(59) Juan ha pasado a ser muy puntual(Span) ->
Juan has become very punctual(Eng)

The translation can also be a simple verb:

(60) he wanted to carry on working(Eng) ->
er wollte weiterarbeiten(German)

Yet others are translated by an idiomatic expression:

- (61) estuve para ir a su casa, cuando entro(Span)
-> I was (about to go/on the point of going)
to his house when he came(Eng).

All these examples are at present cases of complex transfer, but ideally the treatment of tense, aspect and modality could be extended to include them, and this is currently under investigation in the project.

Other phenomena that require an "interlingual" treatment and which are presently the object of research within Eurotra include negation (Hoekstra 1990) and determination and quantification (Van Eynde, 1990b).

4.8. Idioms

The problem of idioms and collocations has not received much attention within the project as yet, but they are obviously a major source of complex transfer. Work has been done on the use of semantic relations for modifiers within the project (see Durand 1969). Semantic relations could help, for example, in the disambiguation of prepositions. German "nach", for instance, translates into English as "to" when followed by a noun designating a location in space, while it translates as "after" when followed by a noun designating a point in time

- (62) er faehrt nach Deutschland(Ger) -> he is travelling to Germany(Eng)
(63) er kommt nach neun(Ger) -> he is coming after nine(Eng)

However, apart from the difficulty of typing nouns in this way, the following examples taken from Collins-Robert French dictionary, show the limitations of such an approach:

- (64)
on that occasion -> a cette occasion
on rare occasions -> en de rares occasions
on just such occasions -> dans une occasion tout a fait semblable

Other more extreme cases of idiom include the well-known example "kick the bucket" which is translated by another, unrelated and equally idiomatic expression in French, "casser sa pipe" (literally, "break one's pipe").

REFERENCES

- Arnold, D. (1989) "Transfer", chapter 8 in *Theoretical and Descriptive Issues in Machine Translation*, unpublished Ph.d. Thesis, University of Essex.
- Badia, T., Blatt, A., Carulla, M., Maas, D.(eds)(1990) "Compounds and related constructions", in *Eurotra Reference Manual, 6.1*.
- Danlos, L. (ed) (1990) "Support Verbs and Predicative Nouns", in *Eurotra Reference Manual, 6.1*.
- Durand, J. (ed) (1989) "Semantic Relations for Modifiers", in *Eurotra Reference Manual, 6.0*.
- Durand, J., Bennett, P. Allegranza V., van Eynde, F., Humphreys, L., Schmidt, P., and Steiner, E. (1990a) *The Eurotra Linguistic Specifications: An Overview*, to appear in *Machine Translation*.
- Durand, J., Van Eynde, F.(eds) (1990b) "Mood and Modality", in *Eurotra Reference Manual, 6.1*.
- Hoekstra, H. (ed) (1990) "Negation", in *Eurotra Reference Manual, 6.1*.
- Johnson, M.(1981) "A unified theory of tense and aspect", in P. Tedeschi and A. Zaenen (eds) *Tense and Aspect*, New York: Academic Press.
- Raw, A., van Eynde, F., ten Hacken, P., Hoekstra, H., Vandecapelle, B. (1989) "An Introduction to the Eurotra Machine Translation System". Working Papers in Natural Language Processing, Katholieke Universiteit, Leuven.
- Reichenbach, H. (1947) *Elements of Symbolic Logic*, Berkeley: University of California Press.
- Schmidt, P. (1990) "Transfer", in *Eurotra Reference Manual, 6.1*.
- Steiner, E. and Reuther, U. (1989) "Semantic Relations for Arguments", in *Eurotra Reference Manual, 6.0*.
- Van Eynde, F. (1990a), "Tense and Aspect", in *Eurotra Reference Manual, 6.1*.
- Van Eynde, F. (ed) (1990b), "Determination and Quantification", in *Eurotra Reference Manual, 6.1*.
- Vangilbergen, L. (1989):"Complex Transfer: An Overview", Eurotra internal document.

