HANNE FERSØE

Representational Issues within Eurotra

Abstract

Machine translation of subordinate clauses, whether finite or non-finite, poses a series of problems within representational issues. In **EUROTRA** this subject has recently been investigated for complements as well as for modifiers to verbal, nominal, adjectival, and prepositional heads.

Some of the problems encountered during these investigations will be addressed in this paper, with a particular focus on the representation of finite and non-finite modifiers.

The claim is made that neutralization of surface differences is necessary. This neutralization of the surface structure is shown to be a prerequisite to simple transfer. The concept of simple transfer is briefly introduced.

Various example representations are shown and discussed.

1 Introduction

Machine translation of subordinate or dependent clauses, whether finite or non-finite, poses a series of problems and challenges within representational issues. In **EUROTRA** this subject has recently been investigated for all types of sentential complements and modifiers, finite as well as non-finite, to verbal, nominal, adjectival, and prepositional heads. (PO-22, 1989)

This paper presents some of the results of the subset of the work, which has been carried out by the Danish language group, and within this, in particular the special research topic dealt with by my colleague Patrizia Paggio and myself.

We have been particularly interested in the definition of a representation which neutralizes arbitrary surface variation found in dependent clauses in the nine EEC languages. Of the many possible structures encountered in the above mentioned syntactic functions we have primarily concentrated on finite and non-finite adverbial modifiers to verbs and nouns. The specifications, which we have proposed as a result of our work, are now part of the Eurotra Reference Manual, which has 'legislative' status in the project. (R.M. 6.0).

In the following quite a few examples of such structures will be shown and discussed. The following list of examples serves to clarify, which types of constructions are involved.

Proceedings of NODALIDA 1989, pages 157-169

Complements to verbal head, non-finite

Industrien prøver AT OVERTAGE FINANSIERINGEN

La Commission veut RENVERSER CETTE TENDANCE

Complements to verbal head, finite

EF ønsker, AT INDUSTRIEN SKAL OVERTAGE FINANSIERINGEN

The Commission wonders, WHETHER INDUSTRY WILL TAKE OVER

Modifiers to verbal head, non-finite

EVERYBODY HAVING AGREED ON THE DATE, the meeting was adjourned

AL SER CONSCIENTE DEL PROBLEMA, la Comisión formuló una propuesta

Modifiers to verbal head, finite

EFTER AT ALLE VAR BLEVET ENIGE OM DATOEN, blev mødet hævet

DA DEN BLEV OPMÆRKSOM PÅ PROBLEMET, formulerede Kommissionen et forslag

ALTHOUGH LABOUR IS EXPENSIVE Europe employs many people

Modifiers to nominal head, non-finite

The actions TAKEN by industry have not reversed the trend

Modifiers to nominal head, finite

de forholdsregler, SOM INDUSTRIEN HAR TAGET, har ikke vendt tendensen

2 Simple Transfer and IS

One very fundamental issue in the **EUROTRA** translation model is the concept of simple transfer. Poul Andersen has given an introduction to this concept, its implications, and provided exemplification in Andersen (1989).

Simple transfer is achieved through a very deep analysis of the input string such that the 'meaning' of e.g. a clause is represented as an annotated tree structure. Simple transfer then consists of substituting the lexical units of the source language tree, also called the **IS** representation, with the appropriate lexical units of the target language. For the translation to be successful, however, many conditions have to be fulfilled, one of them is successful lexical disambiguation, which is a non-trivial problem. The condition in focus here, though, is not lexical, but representational. Simple transfer requires that the annotated tree structure must be transferred unmodified.

The **EUROTRA IS** theory attempts to give a detailed definition of the IS representation from which simple transfer is to take place. One important aspect

of the structures defined at IS is that arbitrary surface differences, which exist between the languages, should be neutralized precisely so that structural transfer needs not take place. Structural transfer means modification of the source language IS structure during transfer to the target language IS structure.

Such surface differences can be observed in the examples listed above. They show examples of both noun and sentence modifiers, where a direct structural mapping from one language to another, i.e. simple transfer, is not possible. Our investigations have shown evidence of great variation w.r.t. surface realization of the so called sentential complements and modifiers across the nine EEC languages.

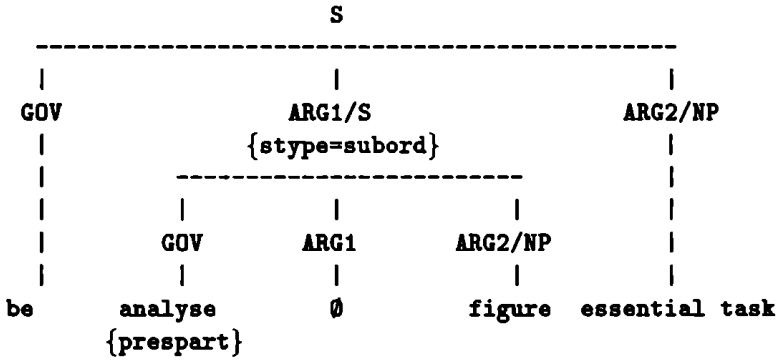
Up until recently the **EUROTRA IS** theory had not been extended to cover all these phenomena. This means that no so called 'legislation' was available for them, except for infinitives and finite clauses in object position to verbs. The filling of such legislative holes, i.e. development, extension, and refinement of the IS theory, is contributed to by the language groups through their experiments with possible representations. Consequently quite different structures have so far been implemented to support these constructions by the different language groups.

3 Example Representations

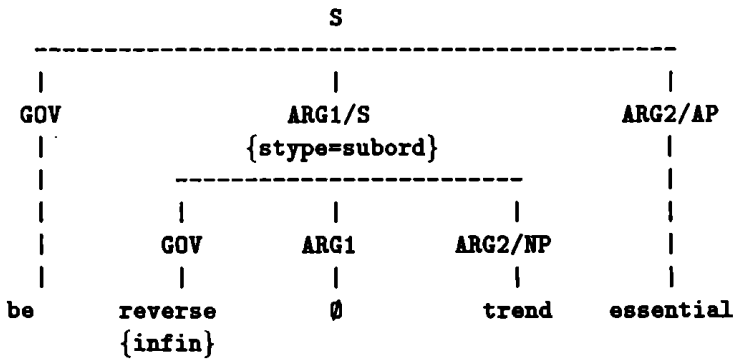
In the following a series of example structures for the clauses in question will be discussed. These examples are authentic, i.e. they have been produced by the different analysis modules developed within **EUROTRA**. The structures, however, have been simplified here for the purpose of highlighting the issue in question. Some of the examples have already been listed above, but here they will be discussed in more detail. The following labels are used in the representations:

```
GOV   - governor, head of a construction
ARG1  - argument 1, deep subject
ARG2  - argument 2, deep object
ARG3  - argument 3, indirect object
MOD   - modifier
TRANC - transconstructional
SBAR  - dependent clause
S     - sentence
NP    - noun phrase
AP    - adjective phrase
PP    - prepositional phrase
Ø     - empty element
{..}  - annotations, features, decorations
```

(1) ANALYSING FIGURES was an essential task



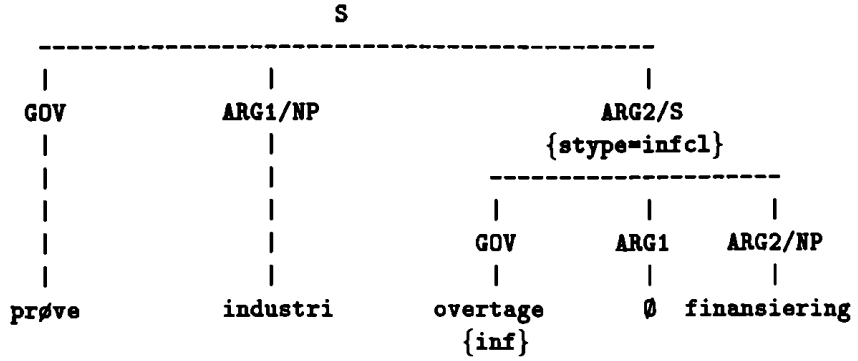
(2) TO REVERSE THE TREND is essential



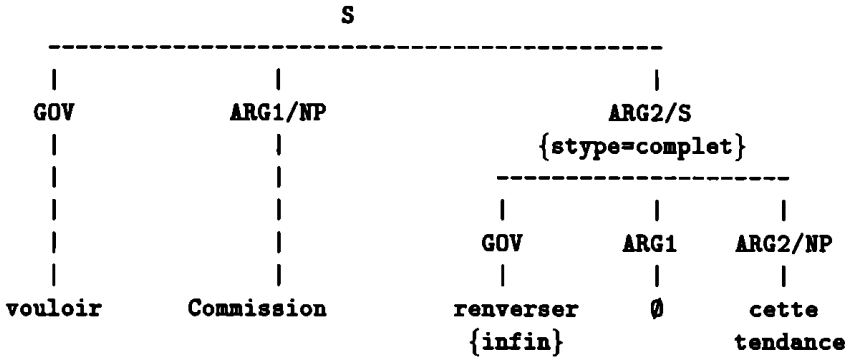
In both (1) and (2) the ARG1's, although they are internally different (-ing clause and infinitive clause), have been assigned the same structure, S, they only differ in morphological typing of the verbal head.

In the following three examples, (3), (4), and (5), the ARG2's are all infinitive constructions, and they are all represented as S structures with an empty element inserted as the logical subject of the infinitive. Again we see, that they differ in features, but not in structure.

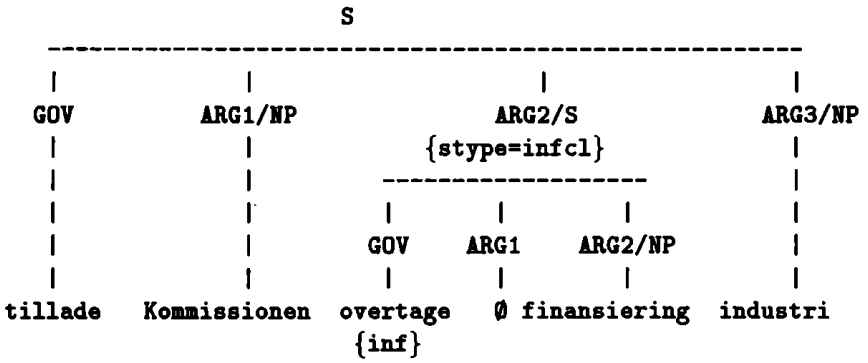
(3) Industrien prøver AT OVERTAGE FINANSIERINGEN



(4) La Commission veut RENVERSER CETTE TENDANCE



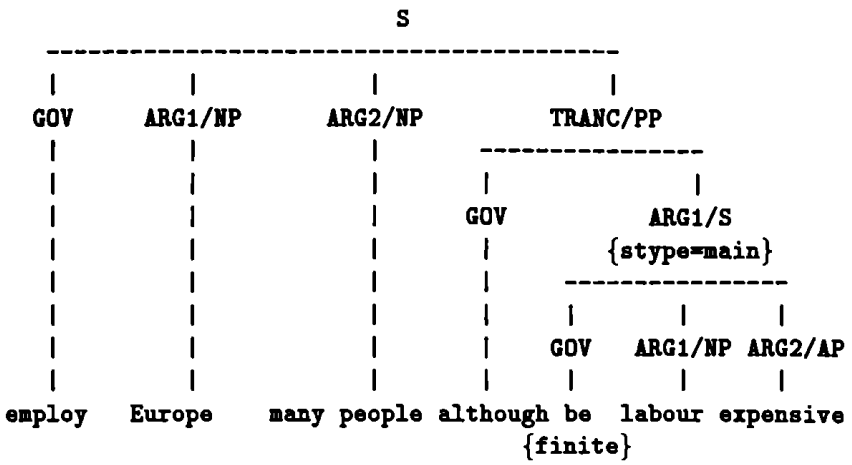
(5) Kommissionen tillader industrien AT OVERTAGE FINANSIERINGEN



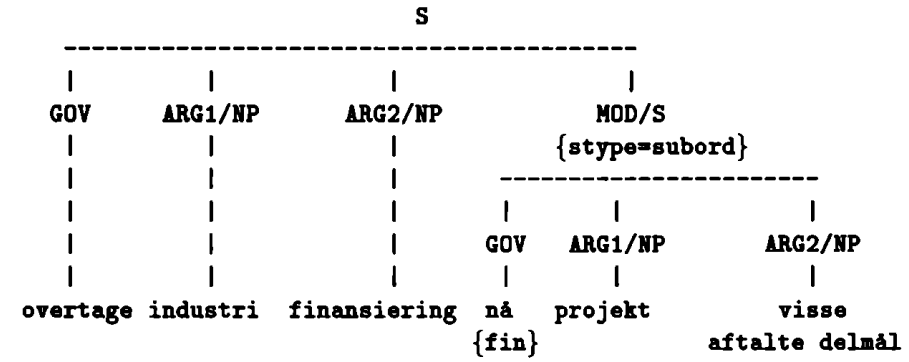
The preceding five examples all show complements—subjects or objects. Their structural representation is in accordance with the IS legislation, which already existed before our investigations started, i.e. S with an inserted empty element in the non-finite clauses.

The three following examples, (6), (7), and (8) show representations of finite adverbial modifiers to a verbal head.

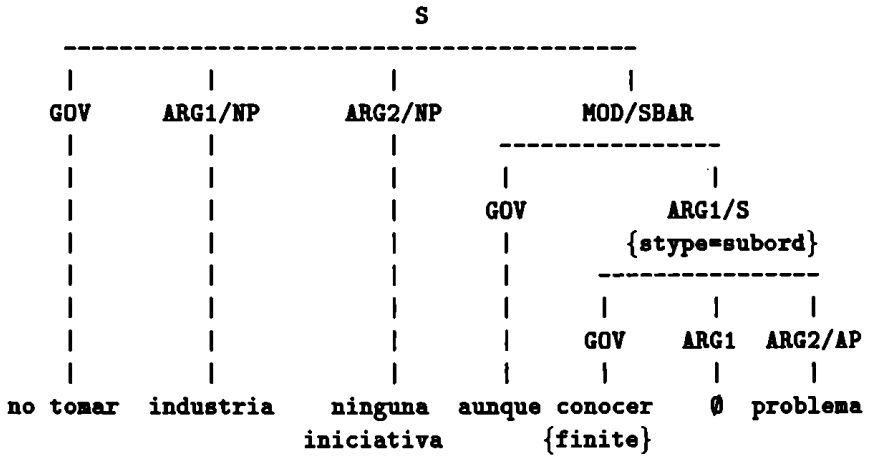
(6) ALTHOUGH LABOUR IS EXPENSIVE Europe employs many people



(7) Industrien overtager finansieringen, EFTER AT PROJEKTET HAR NÅET VISSE AFTALTE DELMÅL



(8) La industria no tomó ninguna iniciativa, AUNQUE EL PROBLEMA ERA CONOCIDO



As the examples show there is great variation w.r.t to the structures and the labelling chosen for these modifiers.

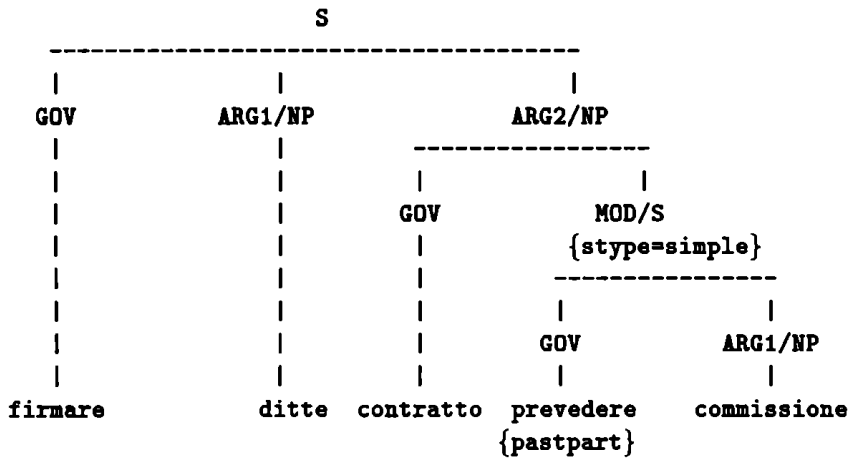
The English analysis, (6), shows the modifier labelled TRANC assigned the category PP. This approach makes the subordinating conjunction, here called a preposition, the head or governor of the entire modifier.

A similar structural approach is chosen in the Spanish analysis (8), but the labelling is different. MOD is chosen to label the modifier, but the category is SBAR with a preposition as governor of the clause. Linguistically it does not seem particularly well motivated to have a preposition as the head of a clause. The head role to clauses is usually reserved for verbs.

The Danish analysis, (7), is more flat, the subordinate clause is attached directly to the main clause, and the subordinating conjunction is elevated, thus leaving the governor position of the sentence to the verb. Here it is not quite clear, how the conjunction should be retrieved.

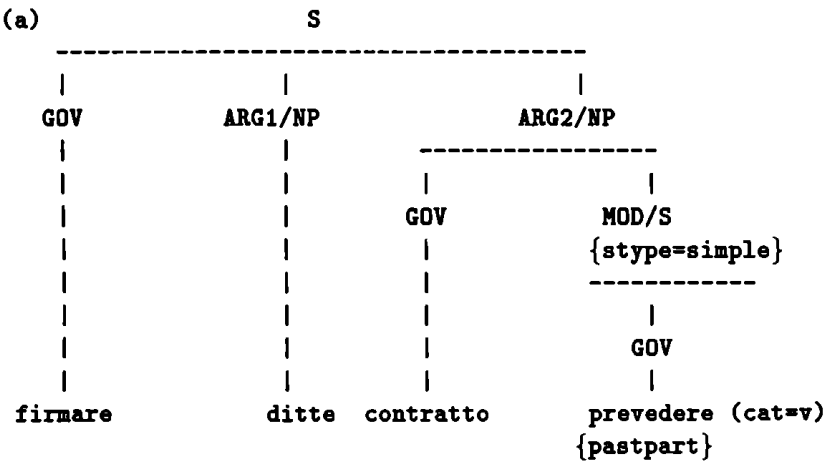
The following five examples show dependents of nominal heads.

(9) Le ditte firmeranno il contratto PREVISTO DALLA COMMISSIONE

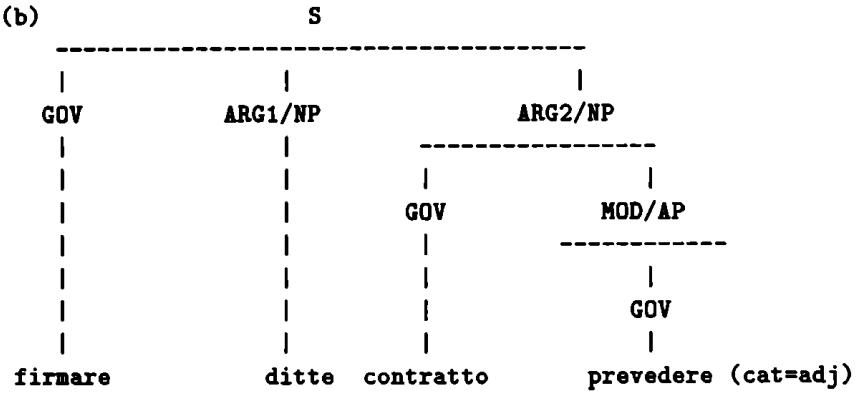


(10) Le ditte firmeranno il PREVISTO contratto

(a)

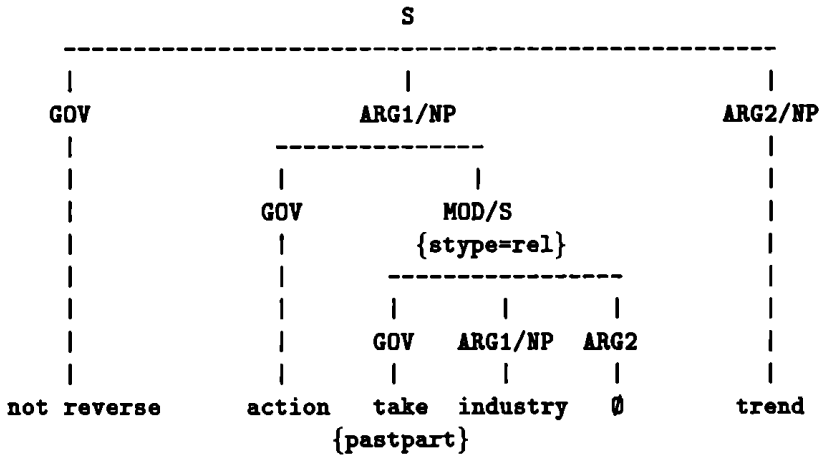


(b)

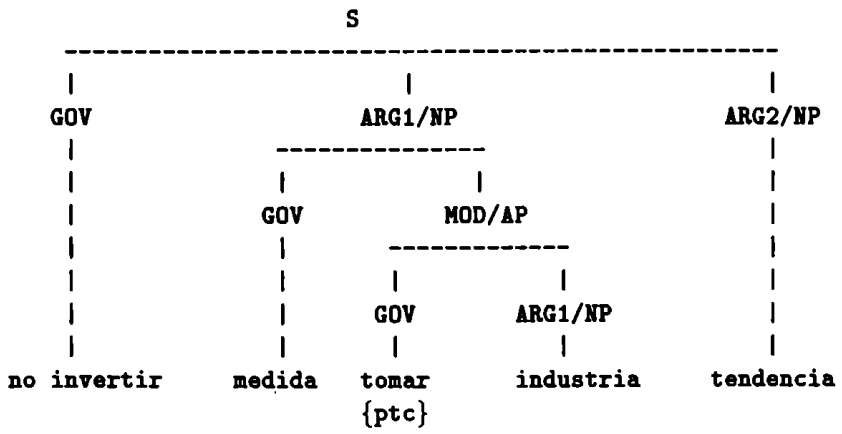


The most interesting observation about these Italian examples is that when the participle has an explicit by-object, as 'dalla commissione' in (9), it is unambiguously considered an S, whereas (10), where no explicit by-object is present, is considered ambiguous between S and AP, and thus the grammar yields two results. This, of course, is unacceptable, since ultimately only one translation should be produced.

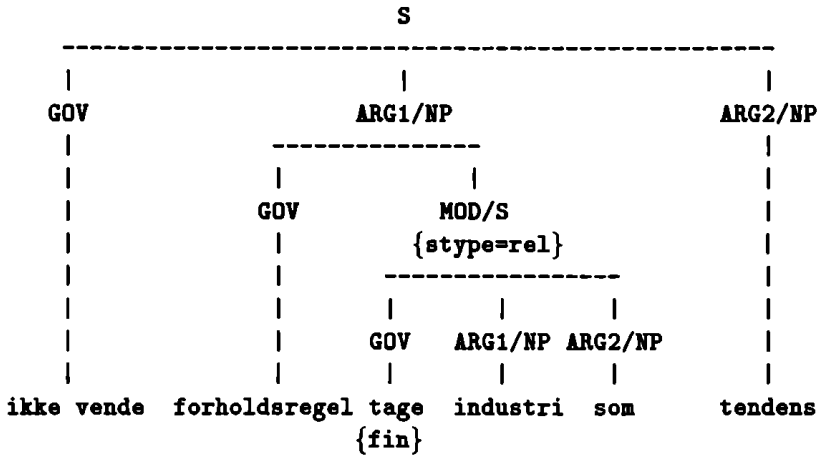
(11) The actions TAKEN BY INDUSTRY have not reversed the trend



(12) Las medidas TOMADAS POR LA INDUSTRIA no invirtieron la tendencia



(13) De forholdsregler, SOM INDUSTRIEN HAR TAGET, har ikke vendt tendensen



(13) as compared to (11) and (12) shows that for Danish a finite relative clause is required to express the meaning conveyed by the non-finite participle in English and Spanish.

A number of observations have been made here concerning the variations in structure and labelling of these representations. One observation, which holds for all of them though, is that the structure assignment is linked to the category assignment to the topmost node of the subordinate clause. The IS theory specifies what constitutes an S, an NP, an AP etc., and therefore the decision taken about the category of the top node implies the structure of the local tree. The following table lists our observations in short form.

3.1 Complements

Category assignment to top node of finites as well as non-finites: **cat=s**.

Stype assignment to top node:

stype=subord, **stype=complet**, **stype=infcl**.

Category assignment to head:

cat=v.

Morphological verbform assignment to head: **prespart**, **infin**, **inf**

Not surprisingly, the category assignment and therefore the structure is identical for all the complements—they follow the IS specifications. The feature differences, however, do cause some problems. Similar problems arise in connection with the modifiers, which will be in focus in the remainder of this paper.

3.2 Modifiers

Category assignment to top node:	
non-finite sentence modifiers:	**cat=s, cat=sbar, cat=pp**
non-finite np modifiers:	**cat=s, cat=ap**
finite modifiers, all:	**cat=s**
Stype assignment to top node:	
sentence modifiers:	**stype=main, stype=subord**
np modifiers:	**stype=simple, stype=rel**
Category assignment to head:	
sentence modifiers:	**cat=v**
np modifiers:	**cat=v, cat=ptc, cat=adj**
Morphological verbform assignment to head:	
sentence modifiers:	**fin, finite, infin**
np modifiers:	**fin, pastpart**

All these representational and classificational differences reflect a purely monolingual analysis underlying the representation provided by the different analysis modules. What we have aimed at in our work is a multilingual analysis as a point of departure for a euroversal definition of the representation.

We have investigated the translational behaviour of a number of these constructions between a number of EEC languages. Not surprisingly, it turned out that a one-to-one mapping of category, morphological verbform and structure is not possible. Therefore a neutralization of structures was necessary, and we proposed cat=S with its implications as the neutral structure. This proposal has now been integrated into the IS 'legislation'.

4 A Neutral IS Structure

4.1 Non-finite Sentence Modifiers

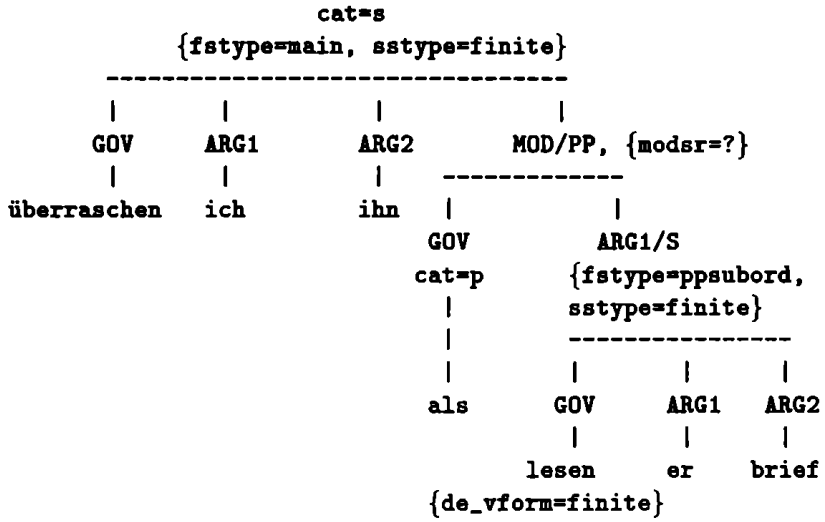
These are not possible in all languages. In the languages which have them, they can be expressed in gerunds, participles or infinitives. In the languages which do not have them, their equivalents are finite clauses preceded by a presentential particle, traditionally classified as either a preposition or a subordinating conjunction.

The annotated tree structures below are a bit more decorated than the structures (1) through (13). Detailed motivation for the presence of this additional information exceeds the aim of this paper. As a basic key to the additional features we can say that **sstype** (surface stype) contains information about the finite- or non-finiteness of the clause, **fstype** (functional stype) contains information about the location in the tree of a given subordinate clause, e.g. the value 'ssubord' is assigned to all ARGi/S and MOD/S daughters to S nodes, **modsr** ('semantic relation of modifier to head') with no values calculated in the

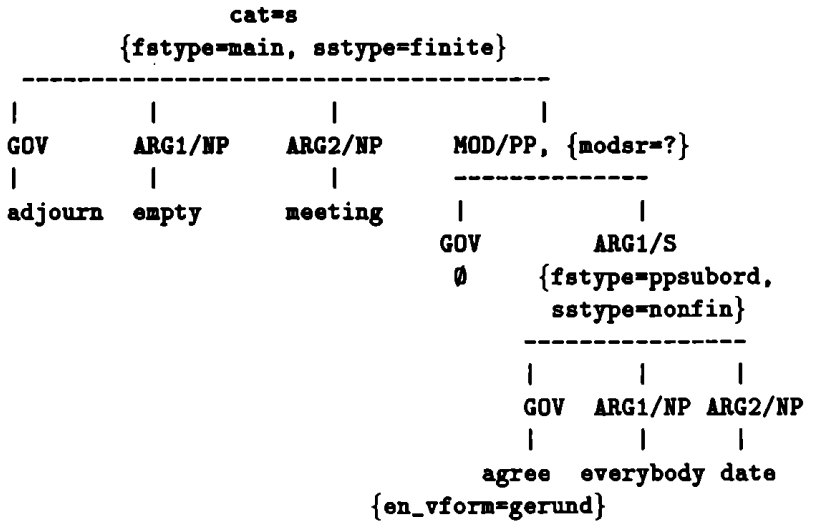
examples, and **V_form** (morphological verb form; the variable, V, is a language prefix, e.g. da) which contains information about verbal inflection, e.g. finite, infin, prespart, etc.

Finite and non-finite sentence modifiers, like in (14) and (15) below, will be represented in the following way:

(14) Ich habe ihn überrascht, ALS ER DEN BRIEF LAS



(15) EVERYBODY HAVING AGREED ON THE DATE, the meeting was adjourned

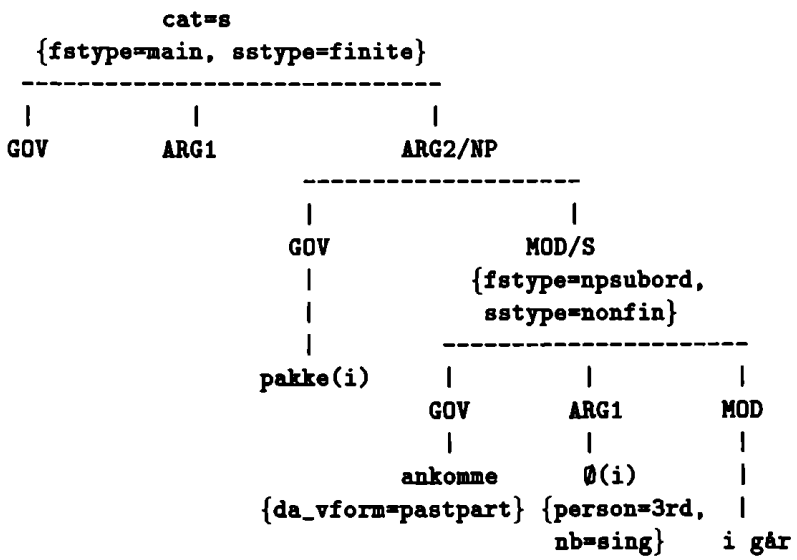


The motivation for this PP structure is translational in that it captures the finite constructions, which in some languages are the only possibility, and which are alternative possibilities in the languages which also have the non-finite constructions. Note in particular the introduction of the 'empty' governor in the structural representation of non-finite sentence modifiers.

4.2 Non-finite NP Modifiers

All the languages included in the investigation can express sentential np modification through non-finite participial constructions or finite relative clauses. But there are restrictions concerning the mappings between languages. Some languages will only accept a relative clause, where other languages will also accept a participle. Relative clauses are always acceptable as a sort of default structure, so the non-finite np modifiers will be represented as an S structure, as in (16) below:

(16) en pakke ANKOMMET i går



The **EUROTRA IS** theory supposes a full GOV - ARG representation for sentences with an obligatory ARG1 (deep subject). This implies insertion of empty nodes in non-finite constructions and a coindexation mechanism as well. The aim of this paper does not permit me to go into details about this interesting aspect of the representation, but we have dealt with it in our research report, and it is an integrated part of our proposal.

5 Concluding Remarks

A surface-neutral structural representation, as shown above, is far from sufficient to ensure successful simple transfer. As shown in the series of trees and the tables earlier, feature assignment problems also have to be solved.

Essentially what is needed to solve these problems is a semantic classification system for subordinate clauses, which is sufficiently fine grained to permit the target language to generate the appropriate surface representation—finite or non-finite—from the input, regardless of the source language. The information, which must be computable from the semantic values, concerns time, tense, mood,

modality, and diathesis for the sentence as such, plus information for the correct translation of presentential particle.

Unfortunately such a system does not exist—yet.

References

Andersen, Poul: How Close Can We Get to the Ideal of Simple Transfer in Multi-lingual Machine Translation (MT)?, This volume:103–113, Reykjavík, 1990.

PO-22: Sentential Complements and Non-finite Constructions. EUROTRA Research Report, April 1989, unpublished. Collaborative effort of EUROTRA-DK, GB, DE, IT, and EL.

The most relevant of the references herein is:

R. Quirk & S. Greenbaum. 1972. A Grammar of Contemporary English, London and New York.

R.M.6.0: *Eurotra Reference Manual version 6.0*, Luxembourg, August 1989, unpublished.

EUROTRA-DK
Københavns Universitet
Njalsgade 80
2300 København S
Danmark