[*International Conference on the State of Machine Translation in America, Asia and Europe*. Proceedings of IAI-MT86, 20-22 August 1986, Bürgerhaus, Dudweiler]

Axel Biewer, Christian Féneyrol, Johannes Ritzke,
Erwin Stegentritt

ASCOF - A modular multilevel system for French-German translation

Universität des Saarlandes
Sonderforschungsbereich 100/C
Bau 4
Im Stadtwald
6600 Saarbrücken
Tel.: 0681/ 302-3713

Name of the system: ASCOF
Status: research
Type of system: bi-lingual
Translated languages: French -> German
Speed of system: ?
Costs: ?
Type of analysis: dependency tree as output
Dictionaries: analysis: about 50000 entries
              transfer: about 10000 entries
              Synthesis: about 10000 entries
Data bases: analysis: ATN and context-free grammar
              Synthesis: transformational grammar
Implementation language: Comskee
Operating system: BS 2000
type of hardware: mainframe

ASCOF - A System for French-German Translation

Axel Biewer Christian Féneyrol Johannes Ritzke
Erwin Stegentritt

1.1

In this article we describe the ASCOF system along with its various components. Since the overall strategy and the most salient parts of ASCOF have already been elaborated upon in other articles and book publications, we offer here a general overview (Chapter 1.2), an example of the output tree of the analysis and then touch a number of features and problems of French-German translation (Chapter 1.3).

1.2

ASCOF (**A**nalysis and **S**ynthesis of **F**rench by means of **C**omskee) is a computer system for the processing of natural language with the purpose of translating written French texts into German texts. This system has been under development since 1981 at the University of the Saarland at Saarbrucken, West Germany. It will be stopped at the end of 1986 by lack of funds.

ASCOF is especially conceived for French and German. For general and detailed descriptions of ASCOF see Chapter 1.4.

1.2.1

The system is programmed in Comskee (Computing and String Keeping Language; cf. Mueller-v. Brochowski et al. (1981), Messerschmidt (1984).

For the linguist, Comskee is a powerful device especially due to its dynamic data types - dictionary, set, string and sentence - and its dynamic operations.
The system runs on a SIEMENS 7561 under the operating system BS 2000. ASCOF has been conceived as a completely automatic translation system. As yet, we have been less concerned with end-user application than with fundamental research. For this reason, we have focused primarily upon linguistic and computer science problems, rather than upon processing speed and the like.

1.2.2

In ASCOF the "classical" divisions have been adopted: analysis, transfer and synthesis. The result of the sentence analysis is represented as a standardized tree structure, which then serves as input for the transfer and synthesis of the target language (cf. Chapter 1.3).

The ASCOF analysis takes place in three steps based on different grammar and algorithm types. The morphological analysis PHASE I is carried out by an algorithm that realizes a mere pattern matching; in PHASE II a context-free grammar identifies non-complex syntactic phrases and the macro-structure of the sentence. A reduction in the homographies of word classes is

simultaneously achieved for the complete sentence. PHASE III determines the syntactic functions within the sentence, using syntactic and semantic criteria, and carries out the semantic disambiguation of lexemes. This phase of analysis is performed by algorithms similar to ATN, representing an interactive system (interactive in the sense of communication and interaction of different components). Consequently, the ASCOF analysis does not constitute a one-pass parser but a system of parsers, (cf. Fig. 1a) and was chosen for the following reason: the complexity and length of the sentences to be analyzed require - for reasons of efficiency - parsing strategies appropriate to the different problems, that is, context-free grammars for PHASE II, which works exclusively with syntactic information, and formalisms similar to ATN for PHASE III, where syntactic and semantic information is combined.

Beyond the phase of analysis, ASCOF includes a phase of transfer and synthesis, where the words of the source language are exchanged for those of the target language and where structures are simultaneously altered in the tree structure, if necessary. The changes of structure are carried out by a transformational grammar. The grammar operates on trees; grammar and algorithm are separate from each other and the algorithms interprets the externally stored rules of the grammar.

On the leaves of the output tree, produced by the syntactic synthesis, a further algorithm operates, which interprets a set of morphological rules in order to generate the correct word forms of the target language. The transfer and synthesis components of ASCOF are shown in Fig. 1b.

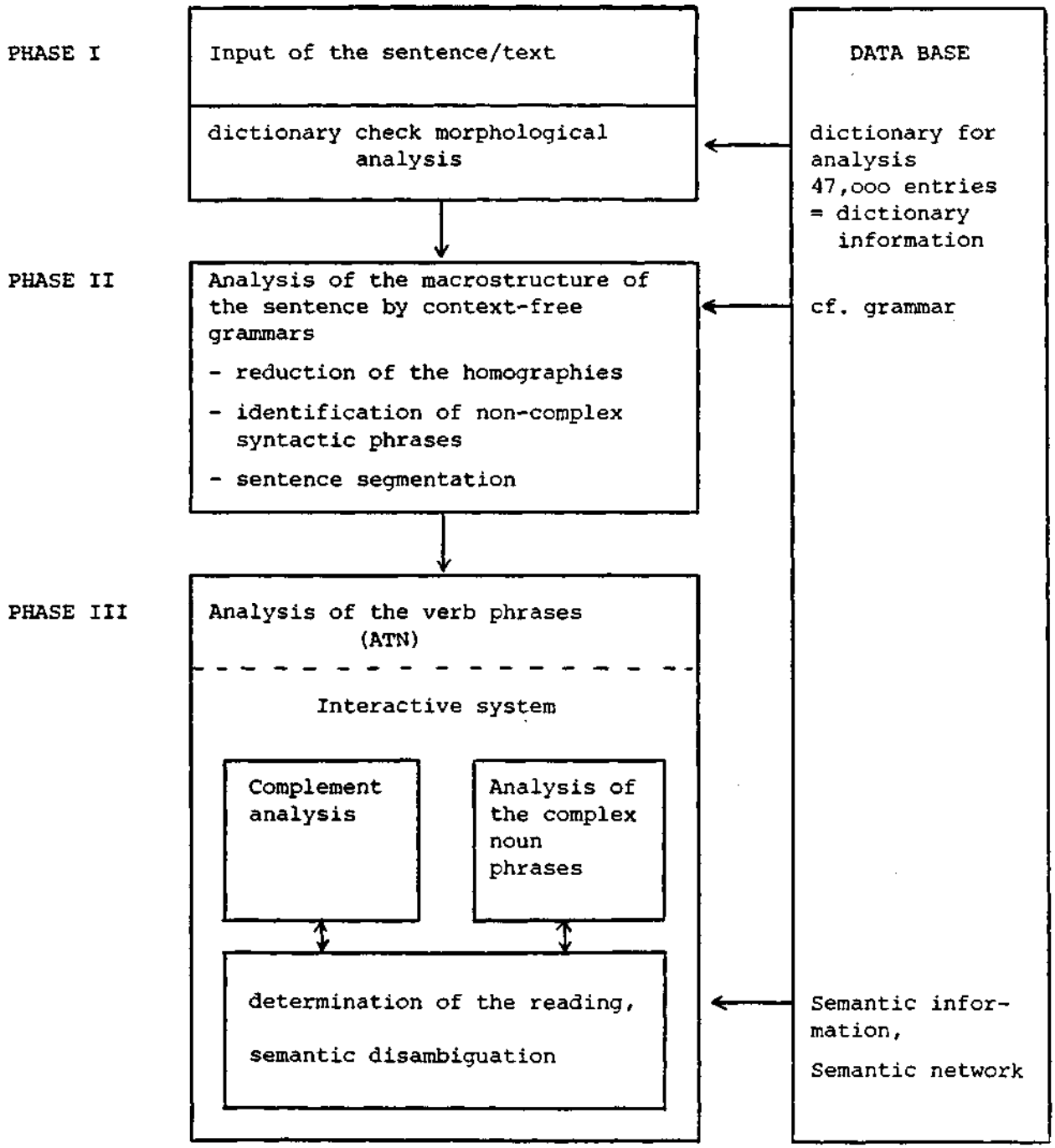


Figure 1a. Analysis components in ASCOF.

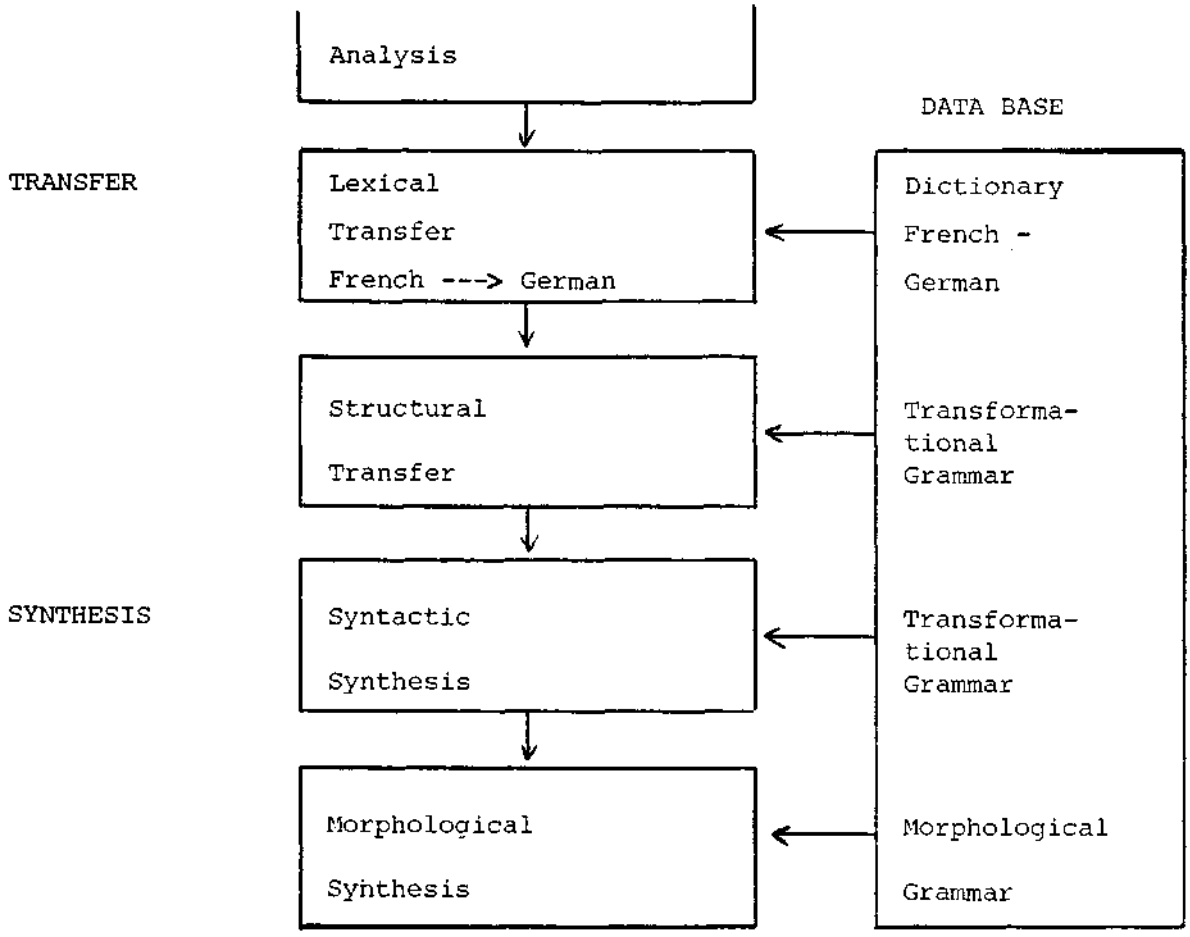


Fig. 1b : Transfer and Synthesis

1.2.3

The most sophisticated phase within ASCOF concerns analysis (French); the synthesizing phase (German) has not yet been developed to such an elaborate extent.

PHASE I of the analysis consists of the sentence/text input and the morphological analysis. Each word form is assigned the set of possible categories as well as the morpho-syntactic information. A full form and a stem dictionary as well as a suffix dictionary (inflectional suffixes) are available. Unknown word forms undergo a derivational analysis.

1.2.4

The second phase of the analysis operates upon a chain of word classes or, given word class ambiguities, upon several word class chains, as they arise from the dictionary check and the flectional analysis.

The range of functions of PHASE II comprises three sections:

a) disambiguation of the word class homographies
b) identification of non-complex syntactic groups
c) segmentation of the entire sentence into parts.

These three steps are not carried out successively, in a particular sequence, but simultaneously, by applying a context-free grammar which operates upon chains of word classes; since the grammar represents a restricted syntactic (macrostructural) sentence analysis, it can be used to resolve word class homographies by the formation of simple syntactic groups and to disambiguate ambiguous syntactic groups when the overall sentence context is taken into consideration.

Each correct result forms the basis upon which the subsequent PHASE III will then operate clause by clause.

1.2.5

The task of analyzing verb sequences, the first step of PHASE III (cf. Fig. 1a), is to group together isolated verb elements (finite verbs, participle I, participle II, infinitive) within a segment to assign a structural description to these phrases (e.g., to determine voice and tense) and - ultimately - to interpret those phrases as nodes of a tree structure.

When the analysis of verb phrases is completed, the sentence is structured in such a way that parts of main and subordinate clauses and their interrelations are identified. Furthermore, non-complex syntactic (one-nuclear) noun and prepositional phrases as well as verb sequences are determined. An interactive component operates on this input performing

(1) the complement analysis,
(2) the analysis of complex (multinuclear) noun and prepositional phrases, and
(3) the disambiguation of lexical items,

all according to syntactic **and** semantic criteria.

In contrast to many other systems of machine translation or linguistic data processing, semantics and syntax are here equally treated, neither having priority over the other.

1.2.6

The phases of the system subsequent to the analysis include a transformational component (transfer and syntactic synthesis) and a morphological synthesis. In the transfer phase, to the source language lemmata are first added their target language equivalents. Here it is not sufficient to mere exchange lexemes since target language differences of a lexical nature arising from syntax and/or semantics need to be borne in mind. Instead the replacement process must take into consideration the information contained in the analysis tree, for example the syntactic (cf. Chapter 1.3) and semantic verb frame actualization. Subsequent to the lexical transfer, the analysis structures are transferred to the structures appropriate to the target language (syntactic synthesis). This transfer component *is* implemented as a tree transformation algorithm that interprets externally stored rules (transformation instructions). The transformation algorithm runs through the analysis tree in preorder, tests for each node wether a package of rules exists for the given node label and - provided that the conditions of a rule are fulfilled - carries out the instructions that refer to some few elementary operations.

The conditions and the transformation instructions can refer to both subtrees and the attribute-value pairs associated with the node.

1.2.7

The input for the morphological synthesis is the labeled tree taken from the syntactic synthesis. In contrast to the syntactic synthesis, the morphological synthesis operates only locally, in other words, the pre-terminal nodes are examined and processed isolated from each other. Tree transformations are thus no longer carried out. The basic forms of the lexemes and the morpho-syntactic information from the pre-terminal nodes serve as keys that call the appropriate rule if the morphological generative grammar.

The grammar for the morphological synthesis of German is able to generate German word forms, provided that the necessary information supplied from all of the preceding phases is complete and accurate.

1.3

The output tree generated by the analysis has a set structure containing certain syntactic groups located at a particular place within the tree structure:

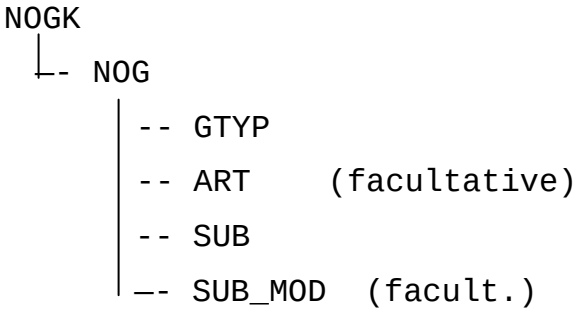
Each sentence contains the node SATZ (= sentence) as the uppermost node. This node governs a sentence-modifier node (S_MOD) and a proposition node (PROP) which are located on the same level.

The node S_MOD represents all the adverbial groups with reference to the sentence. The node PROP governs those nodes which are analyzed as belonging to the verbal framework.

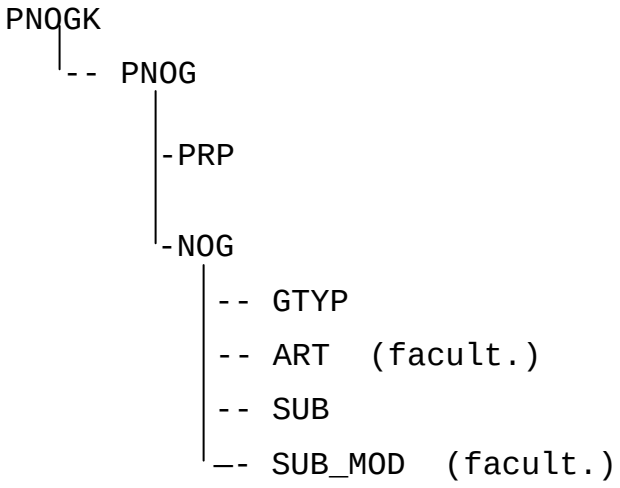
The internal structure of the proposition is represented in the following way: SUBJ (= subject) - REK (= predicate) - D-OBJ (= direct object) - PP-OBJ1 - PP-OBJ2 (= prepositional objects) (cf. Fig. 2).

Each nominal group, whether it falls within the S_MOD or the PROP domain, has the following structure:

1) for nominal groups:



2) for prepositional groups:



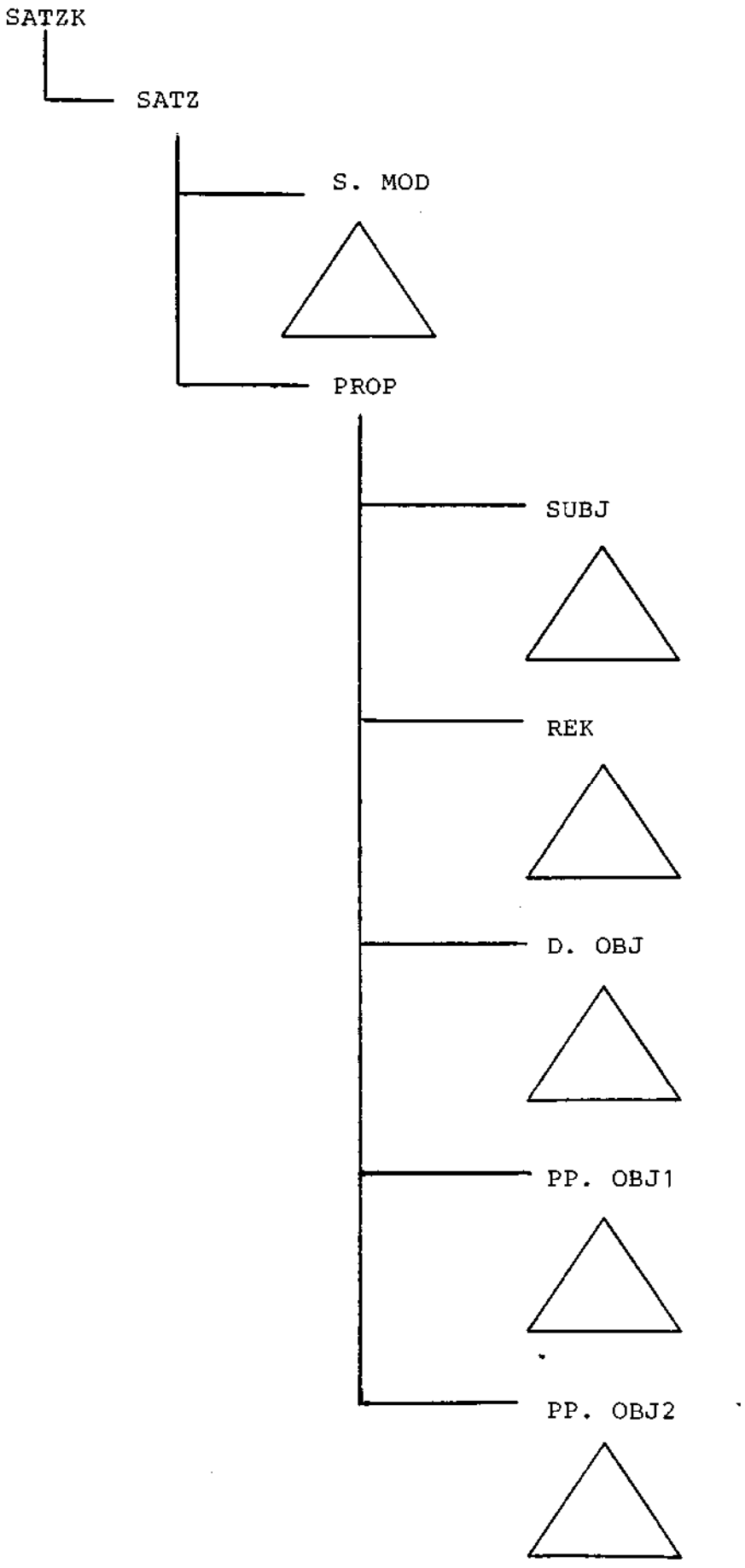


Fig. 2

Modifiers of the nominal groups are placed under the particular noun as a SUB_MOD complex.

The nominal kernel of the nominal/prepositional groups have a special node GTYP (= group type), which holds information applying to the entire group, such as case, gender, number.

An illustration of a complete analysis tree is presented in Fig. 3. The analyzed sentence reads as follows:

(1) A 1'heure de la catastrophe nucléaire, le monde pense à ses enfants.

(In the tree structure "à" is represented as "A2" and "é" as "E1").

The information placed under the individual nodes as labels in the figure stem partly from the analysis and partly from the transfer French-German. The information which was added during the transfer contain the symbol "=", in contrast to the information stemming from the analysis phase.

The prepositions are given a standard translation, e.g., a --> in with the governing dative (cf. for example node 7:1 and 8:8). For the adverbial group "a l'heure de la catastrophe nucleaire", the standard translation is adopted (as default value) and results in: "in der Stunde der nuklearen Katastrophe".

For the same preposition "à" in the prepositional object (node under 4:4), another translation is chosen: in the verbal node (5:3) the source language preposition of the prepositional object ("à l'heure ...") is given as QPRP=A2. As ZPRP=AN, it is noted that the target language equivalence translation is "an". The lemma of this preposition is copied in the synthesis in the prepositional node (8:8). At the same time, the information ZKAS=AKK in the verbal node indicates that in this case the German preposition "an" governs the accusative (in contrast to "an" in the local sense, such as in "an dem Fluss"). This is also copied in the nodes of the preposition and afterwards in the node GTYP (8:9) in order that the correct form of the prepositional group can be generated in the morphological synthesis ("...an seine/ihre Kinder" vs. "... an seinen/ihren Kindern").

The possessive adjective "ses", which forms the SUB-MOD node under the nominal kernel "enfants", is transformed in

```
1:1 SATZK
      2:1 SATZ
            3:1 S_MOD
                  4:1 PNOGK
                        5:1 PNOG
                              6:1 PRP
                                    7:1 A2
                                        A2
                                        TRAD=IN
                                        ZPRP=DAT
                              6:2 NOG
                                    7:2 GTYP
                                          8:1 DEF
                                              NUMER:SING
                                              GENUS:FEM
                                              PERS:3
                                    7:3 ART
                                          8:2 L
                                              L
                                              TRAD=D
                                    7:4 SUB
                                          8:3 HEURE
                                              HEURE
                                              TRAD=STUNDE
                                              GEN=FEM
                                              TYP=S0P6
                                    7:5 SUB_MOD
                                          8:4 PNOGK
                                                9:1 PNOG
                                                      10:1 PRP
                                                            11:1 DE
                                                                 DE
                                                                 TRAD=VON
                                                                 ZPRP=DAT
                                                      10:2 NOG
                                                            11:2 GTYP
                                                                  12:1 DEF
                                                                       NUMER:SING
                                                                       GENUS:FEM
                                                                       PERS:3
                                                            11:3 ART
                                                                  12:2 LA
                                                                       L
                                                                       TRAD=D
                                                            11:4 SUB
                                                                  12:3 CATASTROPHE
                                                                       CATASTROPHE
                                                                       TRAD=KATASTROPHE
                                                                       GEN=FEM
                                                                       TYP=S0P6
                                                            11:5 SUB_MOD
                                                                  12:4 ADJ
                                                                        13:1 NUCLE1AIRE
                                                                             NUCLE1AIRE
                                                                             TRAD=NUKLEAR
            3:2 PROP
                  4:2 SUBJ
                        5:2 NOGK
                              6:3 NOG
                                    7:6 GTYP
                                          8:5 DEF
                                              NUMER:SING
                                              GENUS:MAS
                                              PERS:3
                                    7:7 ART
                                          8:6 LE
                                              L
                                              TRAD=D
                                    7:8 SUB
                                          8:7 MONDE
                                              MONDE
                                              TRAD=WELT
                                              GEN=FEM
                                              TYP=S0P5
```

```
4:3 REK
     5:3 PENSE
         PENSER(,)
         TEMPUS:PRESENT
         GENV:AKTIV
         NUMER:SING
         PERS:3
         MODUS:IND
         TRAD=DENKEN
         REFL=0
         QPRP=A2
         ZPRP=AN
         ZKAS=AKK
4:4 PP OBJ1
     5:4 PNQGK
          6:4 PNQG
               7:9 PRP
                    8:8 A2
                        A2
                        TRAD=IN
                        ZPRP=DAT
               7:10 NQG
                    8:9 QTYP
                         9:2 DEF,NUMER:PLUR,GENUS:MAS
                             FEM
                             PERS:3
                    8:10 SUB
                         9:3 ENFANTS
                             ENFANT
                             TRAD=KIND
                             GEN=NEU
                             TYP=S&P3
                    8:11 SUB_MOD
                         9:4 ADJ POSS
                              10:3 SES
                                   SES
                                   TRAD=SEIN
```

Fig. 3

the synthesis - in keeping with the German word order - before the nominal kernel.

In French the form of the possessive adjective is influenced by the gender and number of the object - with the number of "possessors" also determining the lexeme ("son/sa/ses" vs. "leur/leurs"), whereas in German the gender of the "possessor" determines the lexeme of the possessive: "sein" vs. "ihr".

If automatic translation takes into consideration only the sentence context, this translation problem can obviously not be solved.

For purposes of demonstration a number of transformation rules were included in the synthesis phase of ASCOF, which operates only in reference to the sentence. These rules translate possessive adjectives in object groups according to the gender of the sentence subject (as a possible "possessor" of the object). For the example sentence, this would result in the translation "an ihre Kinder".

1.4

References concerning ASCOF:

Biewer, Axel 1985 Ein ATN zur Verbalgruppenanalyse des Franzoesischen. In Figge, Udo L., 1985: 13-26.

Biewer, Axel;Féneyrol, Christian; Ritzke, Johannes; and Stegentritt,Erwin 1985 ASCOF - A Modular Multilevel System forFrench-German Translation. Computational Linguistics, Vol. 11, Number 2-3, April-September 1985: 137-154.

Féneyrol, Christian 1982 La forme en (-ant): un problème de l'analyse automatique du français. In Fix et al. 1985: 237-262.

Féneyrol, Christian 1983 La segmentation automatique de la phrase dans le cadre de l'analyse du français. In Actes du Colloque International Informatique et Sciences Humaines. 1981. LASLA. Université de Liège. Liège, Belgium: 353-368.

Féneyrol, Christian; Ritzke, Johannes; and Stegentritt, Erwin 1984 Esquisse d'une analyse fonctionnelle du français (ASCOF). In Actes de la XIe Conférence Internationale de l'ALLC. 1984 Louvain-la-Neuve. Champion-Slatkine, Paris-Genève: 153-172.

Féneyrol, Christian and Stegentritt, Erwin 1982 Komplementanalyse in komplexen Saetzen des Franzoesischen. In Stegentritt, Erwin, Ed., Maschinelle Sprachverarbeitung 1981. Vortraege auf der 12. Jahrestagung der GAL, Mainz 1981. Sektion Maschinelle Sprachverarbeitung. Sprachwissenschaft-Computerlinguistik 8. AQ, Dudweiler, W. Germany: 55-68.

Figge, Udo L., Ed. 1985 Romanistik und Datenverarbeitung

1983. Akten des Deutschen Romanistentages, Berlin 1983. Sektion Romanistik und Datenverarbeitung. Sprachwissenschaft-Computerlinguistik 11. AQ, Dudweiler. W. Germany.

Fix, Hans; Rothkegel, Annely; and Stegentritt, Erwin. Eds. 1985 Sprachen und Computer. Festschrift zum 75. Geburtstag von Hans Eggers. Sprachwissenschaft-Computerlinguistik 9. AQ, Dudweiler. W. Germany.

Reding, Helena 1985 Transformationskomponente fuer Baumstrukturen. Linguistische Arbeiten, Neue Folge 11. SFB 100, Saarbruecken, W. Germany.

Ritzke, Johannes 1981 Probleme de l'analyse automatique des prépositions du français. In Schwarze, Christoph, Ed. Analyse des prépositions. IIIe Colloque franco-allemand de linguistique théorique du 2 au 4 fevrier 1981 à Constance. Niemeyer, Tuebingen, W. Germany: 139-157.

Ritzke, Johannes 1982 Zur automatischen Analyse adverbialer Relationen. In Fix et al. 1985: 287-307.

Ritzke, Johannes 1985 Automatische Analyse des Franzoesischen: Komplexe Nominalstrukturen. In Figge, Udo L., 1985: 37-54.

Stegentritt, Erwin 1982 LAISSER + Infinitiv - Konstruktionen in der automatischen Analyse des Franzoesischen. In Fix et al. 1985: 309-327.

Stegentritt, Erwin 1983 Aperçu d'une analyse automatique du français. In Cignoni, Laura and Peters, Carol, Eds. Linguistica Computazionale. Vol. III. Supplement Actes du VIIe Congrès de l'ALLC, Pisa 1982: 243:250.

Stegentritt, Erwin 1984 Automatische Subjektidentifizierung in franzoesischen Infinitivsaetzen des Typs FAIRE + Infinitiv. In Figge, Udo L., Ed. Romanistik und Datenverarbeitung 1981. Akten der 10. Sektion des Deutschen Romanistentages, Regensburg 1981. Sprachwissenschaft-Computerlinguistik 7. AQ, Dudweiler, W. Germany: 22-38.

Other references :

Messerschmidt, Jan. 1984 Linguistische Datenverarbeitung mit Comskee. Teubner, Stuttgart, W. Germany.

Mueller-v. Brochowski et al. 1981 The Programming Language Comskee. 2nd Revised Report. Linguistische Arbeiten, Neue Folge, 4. SFB 100, Saarbruecken, W. Germany.