

# Contextual Constraints for MT\*

Kurt Eberle & Walter Kasper & Christian Rohrer

University of Stuttgart

Institute for Computational Linguistics

Azenbergstr. 12

W-7000 Stuttgart 1

Germany

E-mail: kurt@adler.philosophie.uni-stuttgart.de

April 10, 1992

## Abstract

We outline an experimental machine translation system currently under development at the Institute for Computational Linguistics which aims at flexibility with respect to the level of transfer. The core of the system is a bidirectional transfer system using LFG-grammars for source and target languages. If translation is not possible on the basis of the linguistic information provided by the core system, i.e. in the case of translation mismatches, the system makes selective use of contextual and background knowledge. To this end we developed the concept of *contextual constraints* which are the interface between the core system and a module that we call the *contextual resolver* which deals among other things with the contextual information provided by the preceding text using *Discourse Representation Theory*.

We concentrate on problems such as the translation of tense and anaphoric expressions between German and French.

## 1 Introduction

In the following we will describe an experimental machine translation system as it is currently under development at the University of Stuttgart with the following goals in mind:

- flexibility regarding the level of transfer

\*The research reported here is supported by the German Science Foundation (SFB 340 - Project B3).

- incorporation and exploitation of contextual and non-linguistic information in the transfer system
- variable depth of semantic and pragmatic analysis required for translation
- modularity

The core of the system is a bidirectional transfer system using LFG-grammars for source and target languages (currently German and French) based on the approach suggested by Kaplan et al.(1989). This core system produces a functional structure of the input sentences together with a shallow semantic analysis (f-s-structures). In order to include contextual intersentential information a *Contextual Resolver*<sup>1</sup> using Discourse Representation Theory (DRT; Kamp(1981)) was added. The system combines a linguistic transfer system with (modules of) a general purpose text understanding system in order to study to what extent translation is possible on the basis of the linguistic information provided by the core system (the f-s-structures) and to develop the control mechanisms for cases where this is not possible. For these cases we developed the concept of what we will call *contextual constraints* which are the interface between the transfer system and the contextual resolver.

The general methodology we followed in developing the system is to keep its modules as independent from the application in MT as possible. This means, for instance, that we use the same grammars in monolingual applications (e.g. in an NL-interface to databases) as in MT. An important consequence is that we do not want to introduce into the grammar of one language distinctions or structures relevant only to the target language.

Our research interest is centered around notoriously difficult problems such as the translation of tense and anaphors. Since German and French differ in their tense systems as well as in their gender system a simple transfer scheme on the level of f-s-structures is not possible. This will be discussed in greater detail after a short description of the overall architecture of the system and the concepts underlying it.

## 2 Background

The approach of Kaplan et al.(1989) assumes that the grammar specifies several levels of analysis such as constituent structure (c-structure), functional structure (f-structure), and semantic structure (s-structure). The different levels of analysis are related to each other via *projections*, that is functions mapping the nodes of one structure to nodes in the other structure(s) thus defining *structural correspondences* among the structures. For MT in addition *transfer (T) projections* mapping

<sup>1</sup>The role of the contextual resolver is similar to that of the *negotiator* in the architecture proposed by Kay et al.(1991).

the structures of the source language to corresponding structures in the target language are used as depicted in figure 1.

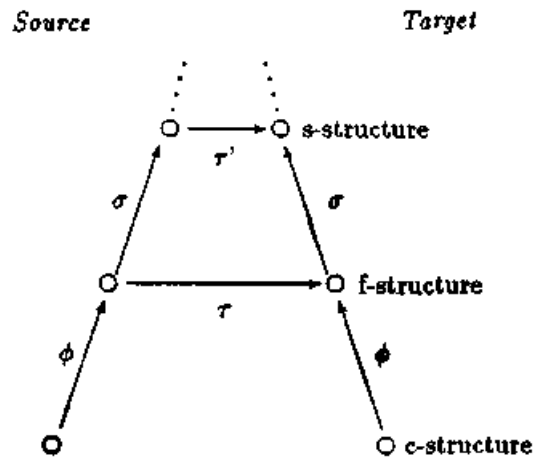


Figure 1: Projections

In this model a transfer is possible on the level of semantic structures as well as a transfer on the level of functional structures. There is no transfer on the level of c-structure. The c-structures of the target language are derived from the semantic or functional structures (Kohl(1991)). Thereby it is possible to derive all sentences which are functionally or semantically equivalent depending on the level of transfer.

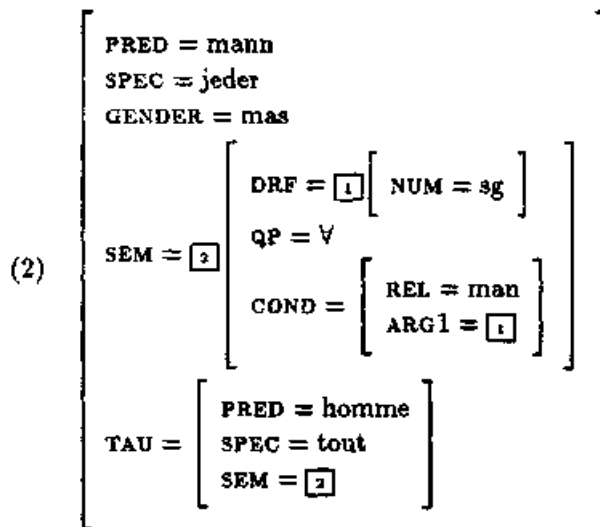
Transfer on the level of f-structures in addition to transfer on the semantic level allows a more fine-grained control of the target language output because it preserves the functional characteristics of the source language expression whenever possible. Let us consider an example. Semantically active and passive sentences should be equivalent. So if we have to translate a sentence like:

- (1) John was admired by Mary.

this sentence would get the same semantic representation as its active counterpart, and so a transfer of only the semantic structure would allow the generation of the active or the passive sentence in the target language. On the other hand, if functional information is added to the semantic information only the passive sentence can be derived because the by-object of the original sentence would be transferred to a par-object in French. It should be obvious that this approach easily allows to combine an interlingual approach (e.g. in terms of a common semantic representation) with a transfer approach.

In our system the  $\sigma$  and the  $T$  projections have been integrated into the functional structure by using special SEM and TAU features. Such an *f-s-structure* e.g. for the NP *jeder Mann* (*every man*) will look as follows<sup>2</sup>:

<sup>2</sup>DRFs here represent *Discourse Referents* in the sense of DRT, standing for the objects talked about. They are complex structures carrying various kinds of sortal information with them which can be used to express e.g. selectional restrictions etc.



The overall structure of our system including this transfer model is shown in figure 2. The source language parser gives a functional-semantic-structure (f-s-structure) of each sentence together with a transfer structure for the target language as illustrated above. Additionally, if necessary, the structure is passed to the contextual resolution component for the evaluation of the contextual constraints discussed below. This component consists at present basically of a temporal and nominal resolver which may call several specialized inference systems such as an event calculus based on ideas from Kowalski/Sergot(1985), a module for computing discourse or rhetorical relations between sentences, a sortal subsumption system and an anaphoric binding module. These have access to knowledge bases containing generic conceptual knowledge such as typical properties of objects, the sort lattice, knowledge about causes and effects, and factual knowledge about what is going on in the world as presented by a text. There is also a system of "text histories" storing computed information about the text as e.g. the available discourse objects for pronouns, the discourse structure etc.

In this paper we will not discuss the problems which the structural correspondence approach gives rise to in general (see e.g. Sadler/Thompson(1991)) but rather illustrate some problems in translating tenses and pronouns involving contextual resolution.

### 3 Problems of translating tenses and pronouns

#### 3.1 Tenses

One well known problem in translating German to French is the fact that the German past tense system has no tense forms which correspond to the distinction of *passé simple* (PS) and *imparfait* (Imp) in French. This distinction has often been associated with certain aspectual distinctions. It is claimed that the PS presents the described eventuality<sup>3</sup> as event-like and "closed", while the

<sup>3</sup>We use this term *as* covering the described events, actions, states and processes in the sense of Vendler's classification. (Vendler(1967)).

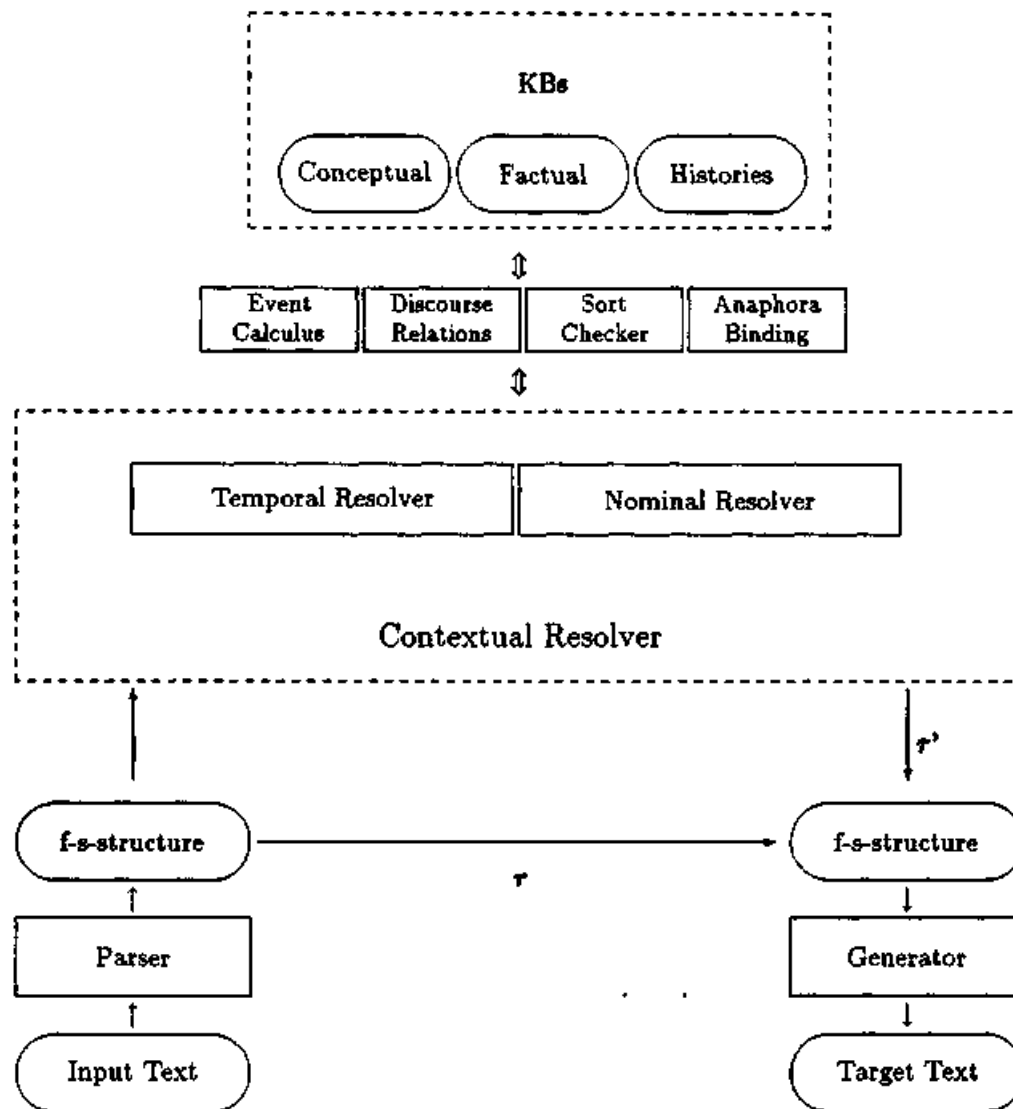


Figure 2: Architecture

Imp presents it as an ongoing process or state-like<sup>4</sup>. The French tense forms force us to make these distinctions explicit while in German they are only implicitly given. Therefore in German one has to get at the required "aspectual" properties by other means<sup>5</sup>. We do not intend to give an explicit specification of the reasoning processes involved here but only point out their complexity<sup>6</sup>.

Let us consider an example:

- (3) a. Der Beamte holte eine Akte aus seinem Schreibtisch.  
(The clerk drew *a file* out of his desk.)
- b. Der Beamte holte Akten aus seinem Schreibtisch.  
(The clerk drew *files* out of his desk.)

<sup>4</sup> A more detailed discussion of the "aspectual" properties of PS and Imp is given in Eberle/Kasper(1991).

<sup>5</sup>With respect to English similar problems can arise for choosing between the *progressive* and *non-progressive* form.

<sup>6</sup>For a formal description cf. Eberle(1991).

Possible translations are:

- (4) a. Le fonctionnaire tira un dossier de son bureau.
- b. Le fonctionnaire tirait des dossiers de son bureau.

The essential difference lies in the fact that the first sentence contains a singular count NP *a file* where the second has a bare plural NP *files*. In cases like (3) the singular count NP favors an event-reading - leading to a PS in French - while the bare plural strongly suggests a process-reading, requiring Imp in French<sup>7</sup>.

Another important difference between PS and Imp related to the aspectual properties is that event describing sentences with the PS tend to move narrative time forward. Statives on the other hand normally provide *background* information for the events talked about and so temporally surround the events. Therefore the temporal order of the eventualities involved can give important clues for the translation.

- (5) a. Hans fuhr nach Frankreich<sub>e1</sub>.  
(Hans went to France.)
- b. Hans fuhr nach Frankreich<sub>e1</sub>. An der Grenze wurde er angehalten<sub>e2</sub>. Er mußte seinen Kofferraum öffnen<sub>e3</sub>.  
(Hans went to France. At the border he was stopped. He had to open his boot.)

For (5.a), being also part of (5.b), the information contained in the sentence independent of the context would result in the aspectual value *non-stative*, and so a translation by PS should be appropriate. However, in (5.b), an Imp for *e1* will be preferred, since the information that *e2* happens at the border (to France) takes the reader to locate *e2* (and *e3*) within *e1* thus enforcing a stative reading for *e1*. This shows that the aspectual value predicted from sentential information can be changed by the context. Notice also, that often the relevant information must be inferred from the *subsequent* context as in (5.b). This is an additional processing problem indicating that sentence by sentence translation is in general not possible.

The temporal anchoring of an eventuality in a new sentence interacts also with other kinds of resolution, especially of local anaphors, complicating matters further. Consider the following alternative continuations (a), (b) and (c) of the first sentence of (6):

- (6) Hans fuhr über die Schweiz nach Frankreich<sub>e1</sub>.  
John went to France via Switzerland.
- a. Er verbrachte das Wochenende *dort* mit einer Klettertour<sub>e2</sub>.  
(He spent the weekend *there* making a climbing tour.)
- b. Er flog *von dort aus* nach Spanien<sub>e2</sub>.  
He flew to Spain *from there*.)
- c. Er besuchte *dort* seinen Freund Guido<sub>e3</sub>, der in Basel wohnte<sub>e3</sub>.  
(He visited *there* his friend Guido, who lived in Basel.)

<sup>7</sup>Such modifications of sentence aspect by e.g. the type of complements and adverbials have been described e.g. in Moens/Steedman(1988), Krifka(1987) and Eberle(1991).

In (6.a) the demonstrative *dort* can refer to Switzerland or to France. In the first case we get a situation as in (5.b) and the event  $e_1$  should be described using Imp. In the second case  $e_1$  precedes  $e_2$  and, therefore, the translation must use the PS. So the temporal anchoring depends on how *there* is resolved.

(6.b) and (6.c) are different. Here it is knowledge about the incompatibility of event types which decides the temporal anchoring and, as a consequence, the local resolution, too: *Flying to Spain-events* and *Going (by car, by train, ...) to France-events* caused by the same agent cannot overlap. This rules out that the  $e_2$  of (6.b) can be included in  $e_1$  and that *dort* can refer to Switzerland. To eliminate the interpretation that the  $e_2$  of (6.c) follows  $e_1$  requires a sophisticated default reasoning which uses the rule that normally visiting events take place at the domicile of the visited person, and being in France therefore is incompatible with visiting somebody in Switzerland. This is compatible only with the interpretation that  $e_2$  is included in  $e_1$ . Otherwise one would have to assume an unmentioned event of changing the location between  $e_1$  and  $e_2$ . But this would violate principles of minimizing event assumptions as stated e.g. in Shoham(1988) or Sandewall(1988).

### 3.2 Pronouns

Similar problems with respect to incomplete information arise in the case of anaphora. We will discuss this problem here mainly with respect to gender. German and French have different gender systems. So the German noun *Tisch*, for instance, is masculine, the French *table* is feminine (and the English *table* is neuter). Therefore a pronoun can only be translated if its referent is known. But the notion of *reference* as required for MT differs a bit from standard accounts. In a theory like DRT the resolution of pronouns would usually consist in establishing the co-reference of two discourse entities, the one representing the referent of the pronoun, the other being the antecedent. But for MT this would not be sufficient. In addition one needs information about how the antecedent entity had been introduced, e.g. by what noun. The following example might illustrate this:

- (7) a. Das Mädchen verjagte die Katze, weil *sie* seine Wurst fressen wollte.  
(The girl chased the cat because *it* wanted to eat her sausage.)
- b. La fille a chasse le chat, parce qu'il voulait manger son saucisson.

The translation requires the resolution of the pronoun *sie* because *cat* is feminine in German, masculine in French (and neuter in English). But in order to get the transfer relation right it is not sufficient to know that *die Katze* and *sie* refer to the same object, but one has also to know that this object was introduced by the noun *Katze* because only on the basis of the transfer relation between *Katze* and *chat* the correct gender for the French pronoun can be established. Another interesting point concerns the possessive pronouns: in the direction German-French a resolution is not necessary because in French the choice of the possessive pronoun (*son* or *sa*) depends only on the gender of the head noun. The direction French-German, however, requires resolution as the

German possessive pronoun *sein/ihr* depends on the gender of its antecedent.

## 4 Towards a Solution: Contextual Constraints

How does the system know when a more or less direct transfer is sufficient and when additional information from the context or a deeper semantic analysis is required? The points in questions we illustrated concerned differences in the structure of source and target language. Evidently, it is the *translator*, not the speaker of one of these languages who has to know about such differences. So the information about when and what kind of additional information is required to resolve such "translation mismatches" (Kameyama et al.(1991)) should be part of the transfer system - representing the translator - not part of the individual grammars. This kind of architecture should also help to solve the old problem - especially for interlingual approaches - in translation theory of how one can get around the necessity to introduce into a (possible source) language distinctions which are only relevant with respect to another (possible target) language. Furthermore, the architecture should contribute to solving the problem of how one can decide when a refined semantic analysis with respect to MT might be necessary and when it can be avoided, without requiring specific adaptations in the grammars for the languages involved. In our system the transfer relation is basically defined on the level of sentential f-s-structures. The question then arises how transfer at the level of these grammatical structures can be combined with mechanisms which resolve such mismatches.

Another consideration which has to be taken into account is that the information obtained from the context must flow back to the structure forming the input for the generator. Since we want to use the same grammar for parsing and generation (in the target language), the contextual reasoning component should not provide more information or other types of information than is used in the target language grammar. Otherwise the generator would get overspecified structures not derivable from the grammar. This makes it necessary to specify exactly what *grammatical* information should be provided by the reasoning processes.

In our approach we assume that the use of such additional mechanisms is controlled directly by a special type of constraints in the transfer component in the grammars for a language pair. We call such constraints "contextual constraints" and mark them by " $=_c$ " equations. These constraints need not to be satisfied by information the grammars could provide, since they are not grammatical well-formedness constraints in the usual sense but rather must be evaluated with respect to the information about how the current sentence is embedded in the context and so has to be provided by additional mechanisms. These constraints can be seen as specific queries to the contextual resolver (CR) about additional information the translator would like to have, but the translator does not have to care about how the CR gets this information. It is up to the CR to know what kind of analysis is required to answer a specific query. This involves a clear separation of specifying what



kind of information is required with respect to the target language grammar and the processing issues for finding this information.

Any NL-understanding system requires such additional mechanisms for a full semantic analysis of sentences or discourses. We make a specific use of them for MT. They are invoked only in case the transfer structures contain constraints which are marked for contextual evaluation. If the structure does not contain any such constraints the contextual resolver does not make any contribution to the translation process and no deeper analysis is looked for, since it is assumed that it is not required for the translation.

In the following we will exemplify our approach by showing how it is applied to the problems of tense and anaphora discussed in the previous sections.

We will illustrate our treatment of the translation of tenses by the following examples, the first requiring a translation of the German simple past to a *Passe Compose* (or *Passe Simple*) with respect to  $e_1$ , the other one an *Imparfait* translation, depending on the continuation ( $e_2$  or 63):

- (8) a. Hans ging ins Hotel zurück ( Dort setzte er sich an die Bar ( $e_2$ ).  
 (John went back to the hotel. There he sat down at the bar.)  
 b. Hans ging ins Hotel zurück ( $e_1$ ). Unterwegs traf er einen Freund. ( $e_3$ )  
 (John went back to the Hotel. On the way he met a friend.)
- (9) a. Jean est rentré à l'hôtel. Là, il s'est assis au bar.  
 b. Jean rentrait à l'hôtel. En route il a rencontré un ami.

For the correct translation of the German (or English) past tense sentences information about the "completeness" or "incompleteness" of the described events with respect to the context is required, as discussed above. The grammar specifies the following for the translation of the German past tense to French:

(10) pret: (↑TENSE) = past  
 {/ (↑TAU TENSE = PC  
 (↑TAU SEM DRF COMPLETED =<sub>tc</sub> +  
 / (↑TAU TENSE = Imp  
 (↑TAU SEM DRF COMPLETED =<sub>tc</sub> - /}

These equations state that the past tense should be translated as a *Passe Compose* (PC) if the contextual constraint (marked by "=<sub>tc</sub>") that the described eventuality is "completed" (+) can be proven by the CR (we use the feature COMPLETED as a name for the aspectual property by which the correct translation can be determined; the feature is also used in the description of the tenses in the French grammar). In the opposite case (-) the *Imparfait* should be used. The constraint is not a well-formedness constraint in the usual sense since both translations would result in grammatically well-formed sentences. But only one of them will be contextually appropriate. Therefore if these constraints cannot be satisfied on the sentential level the generator - instead of simply failing as it would do in the case of unsatisfied grammatical constraints - can pass these constraints to

the contextual resolution component which will try to deduce the necessary information from the context. This component decides the constraint on the basis of the temporal relation holding between the *return-to-the-hotel* eventuality of the first sentence ( $e_1$ ) and that described in the following sentence ( $e_2$  or  $e_2$ ), the *sitting-at-the-bar* or the *meeting*, respectively. If the new eventuality follows  $e_1$  temporally the feature COMPLETED has the value +; if the eventualities overlap the value is —. This result is passed to the generator by merging the result into the transfer structure and this excludes one of the translations. The reasoning process leading to the conclusion that in the first example the return-to-the-hotel precedes the sitting-at-the-bar and therefore is regarded as closed, involves roughly the following steps: from the sentences we know that the first eventuality involves a movement towards the place where the second is situated (after resolving *there*). From events of the type *move-towards-X* (sortal subsumption) we can infer that while the event lasts, the agent cannot be at X (event calculus). Because doing something at X is incompatible with not being at X (event calculus), the existence of an arrival at X can be inferred which finishes the return-to-the-hotel event (event calculus). Thereby we have established that the return-to-the-hotel event is finished when the second event takes place. It also establishes *continuation* as the discourse relation between the two sentences. In the other case the adverbial *on the way* makes it clear that the meeting took place during the return to the hotel and so it provides the *background* for the meeting.

Pronouns can be dealt with in a similar way. We assume that the resolution component will try to relate or identify the pronominal DRF with some otherwise contextually available DRF, and that the component can access information about the DRFs: the information relevant here is by what description (e.g. head noun) they have been introduced together with the corresponding transfer information<sup>8</sup>. Both kinds of information are available when a DRF is first introduced. A pronoun now appears together with a substructure for antecedent information as illustrated in the following example for the German pronoun *er* (*he*):

(11) **er:** { ↑SEM ANTEC GENDER) = *mas*  
           { ↑SEM ANTEC NUM) = *sg*  
           { ↑TAU SEM ANTEC NUM = *sg*  
           { ↑TAU SEM ANTEC GENDER) =<sub>tc</sub> *X*

The transfer rule here says that the GENDER-value of the transferred structure has to be defined contextually (we use *X* as a variable which will be instantiated by the resolution process). It can be retrieved from the transfer information associated with the DRF which is the antecedent.

The flexibility of this approach can be illustrated by the circumstance that e.g. in a German-English translation system one could dispense with pronoun resolution in the case of plural personal pronouns because neither German nor English makes a gender distinction then. This just means that it is not necessary to specify a corresponding contextual constraint as for French.

<sup>8</sup>In case, the DRF has not been introduced explicitly in the text but has to be constructed - as might be necessary e.g. for plural anaphora - some more steps involving e.g. sort generalization are required.

## 5 Conclusions

It is well known that high quality translation is not possible without recourse to contextual and world knowledge in general. Here we made a proposal for setting up a system for integrating linguistic and extra-linguistic knowledge and suggested the mechanism of contextual constraints as a suitable and flexible interface for specifying in a declarative way the information required for the translation process. At the same time the concept of contextual constraints ensures that only grammatically relevant information enters the structures used for generating the target language text.

Naturally, not all components of the system have been fully implemented yet. However, the grammars, the parser and the generator work (even generation from semantic structures). Of the temporal and nominal resolver prototypical versions exist. Their integration into the system is now being examined.

An important issue we will turn to in future work is the investigation of how other types of translation mismatches, especially lexical differences, can be treated using the concept of contextual constraints for efficiently controlling the inference procedures needed therefore. This includes relating our approach to contextual reasoning to the use of abduction in Hobbs/Kameyama(1990) and information flow graphs of situation semantics (Kameyama et al.(1991)).

## References

- Eberle, K. (1991): *Ereignisse: Ihre Logik und Ontologie aus textsemantischer Sicht*. PhD thesis, Universität Stuttgart, 1991
- Eberle, K./Kasper, W. (1991): Tense, Aspect, and Temporal Structure in French. In: Kamp, H. (ed.), *Tense and Aspect in English and French*, DYANA-Deliverable R2.3.B, pp. 4-40. Center for Cognitive Science, Edinburgh, 1991
- Hobbs, J./Kameyama, M.** (1990): Translation by Abduction. In: *Papers presented to the 13th International Conference on Computational Linguistics*, pp. 155-161, Helsinki, 1990
- Kameyama, M./Ochitani, R./Peters, S.** (1991): Resolving Translation Mismatches With Information Flow. In: *Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics*, Berkeley, 1991
- Kamp, H.** (1981): A Theory of Truth and Semantic Representation. In: Groenendijk, J./Janssen, T./Stokhof, M. (eds.), *Formal Methods in the Study of Language*, pp. 277-322. Mathematisch Centrum, Amsterdam, 1981

- Kaplan, R./Netter, K./Wedekind, J./Zaenen, A. (1989): Translation by structural correspondences. In: *Proceedings of the Fourth Conference of the European Chapter of the Association for Computational Linguistics*, pp. 272-281, Manchester, 1989
- Kay, M./Gawron, J. M./Norvig, P. (1991): VERBMOBIL: A Translation System for Face-to-Face Dialog, 1991
- Kohl, D. (1991): Generierung aus unter- und überspezifizierten Merkmalsstrukturen in LFG. Arbeitsberichte des SFB 340 *Sprachtheoretische Grundlagen für die Computerlinguistik* 9, IMS, Stuttgart, 1991
- Kowalski, R./Sergot, M. (1985): A Logic-Based Calculus of Events. *New Generation Computing*, 4(1), 1985
- Krifka, M. (1987): *Nominalreferenz und Zeitkonstitution. Zur Semantik von Massentermen, Pluraltermen und Aspektklassen*. PhD thesis, Universität München, 1987
- Moens, M./Steedman, M. (1988): Temporal Ontology and Temporal Reference. *Computational Linguistics*, 14.2 (Special Issue on Tense and Aspect):15-28, 1988
- Sadler, L./Thompson, H. S. (1991): Structural Non-Correspondence in Translation. In: *Proceedings of the Fifth Conference of the European Chapter of the Association for Computational Linguistics*, pp. 293-298, Berlin, 1991
- Sandewall, E. (1988): Non-Monotonic Entailment for Reasoning about Time and Action. LiTH-IDA R-88-27 - R-88-29, Linköping University, Department of Computer and Information Science, Linköping, Sweden, 1988
- Shoham, Y. (1988): *Reasoning about Change. Time and Causation from the Standpoint of Artificial Intelligence*. Cambridge (Mass.): MIT Press, 1988
- Vendler, Z. (1967): *Linguistics in Philosophy*. Cornell University Press, Ithaca, New York, 1967