# Ontology and lexical semantics for generating temporal discourse markers

**Brigitte Grote**
Otto-von-Guericke Universität Magdeburg
Inst. für Wissens- und Sprachverarbeitung
P.O. Box 4120
39016 Magdeburg, Germany
grote@iws.cs.uni-magdeburg.de

**Manfred Stede**
Technische Universität Berlin
FB Informatik, Projektgruppe KIT
Sekr. FR 6–10, Franklinstr. 28/29
10587 Berlin, Germany
stede@cs.tu-berlin.de

## Abstract

In text, temporal relations between events can be signalled in several ways; among them are specific lexical items, here called temporal discourse markers. We analyse the semantics of about 20 German subordinating conjunctions and prepositions and transfer these findings to a sentence generation framework that uses a dedicated discourse marker lexicon for producing complex sentences. After discussing the ontological decisions and the lexical representations, we demonstrate how this information can be used to choose an appropriate temporal marker when verbalizing a pair of time-stamped event representations.

## 1 Introduction

In knowledge-based natural language generation (NLG), an abstract meaning representation is successively transformed into a linguistic sentence or text. Specifying this mapping involves both designing the abstract representation and defining its relationship with lexical knowledge, which involves ontological decisions.

In this paper, we investigate the specific problem of signalling a temporal relationship between two events in the German language. Focusing on explicit lexical signals, this is one aspect of the general problem of choosing *discourse markers* in NLG. Our generation framework (Grote and Stede 1998) sees sentence production as a two-step process of first mapping the abstract representation to a sentence-semantic specification, which can then be given to a standard NLG front-end such as KPML (Bateman 1997) for realizing it in natural language.

The first step involves the decisions that affect meaning; most importantly it involves lexical choice. When going further to the production of complex sentences from a representation of propositions and a coherence relation, the first step turns into the more complicated task of *sentence planning*. Then the idea is to employ a dedicated discourse marker lexicon holding the information on how to signal a discourse relation under the contextual conditions; sentence planning now involves interleaving marker choice with the other generation decisions appropriately.

The goals for this paper are twofold: First, we provide a (brief) analysis of the semantics of German temporal markers, in particular of subordinate conjunctions and prepositions (section 3). Then, by extending the ontological system of a previous generator to include temporal information (section 4), we carry our analyses over to NLG, suggest lexical entries for temporal markers grounded in the ontology (section 5), and propose a procedure for generating a range of paraphrases from representations of two time-stamped events, such that the temporal relation between the events is signalled adequately (section 6).

## 2 Related work

Temporal discourse markers have not received much attention in NLG so far. The most complex study in this area is that by Dorr and Gaasterland (1995) for English temporal conjunctions. They examine the application of linguistic theories of tense and aspect to temporal marker choice starting from a pair of time-stamped events. Klenner (1991) discusses the generation of German complex temporal clauses, but only deals with few markers, and moreover does not address questions of lexical and syntactic constraints. In contrast, Grote (1998) is mainly concerned with syntactic and lexical features that characterize German temporal connectives, and provides some ideas for representing them in a lexicon for NLG. Finally, Gagnon and Lapalme (1996) describe the generation of

French temporal adverbs based on a DRT representation of discourse, but only briefly address temporal connectives.

In brief, existing studies focus either on lexical and syntactic constraints of marker selection, or on the relation of marker choice to event types. But the entire process of representing the semantics of temporal marker meaning, relating it to complex event representations, and using this information in knowledge-based NLG for describing the interaction of marker choice and other sentence planning tasks, has not been addressed yet.

## 3 Dimensions of temporal discourse marker description

Descriptive work on German temporal markers such as Bäuerle (1995), Herweg (1991), Sinn (1991), Steube (1980), and grammars such as Helbig and Buscha (1991), suggest that quite diverse factors influence the use of a particular temporal discourse marker. These factors can be divided in two groups: The first group relates to the semantics of markers (mainly the temporal relation holding between two events), and the second relates to the interaction with the lexical and syntactic environment (constraints on Aktionsart, aspect, tense, part of speech). Starting from these classifications, we analysed a wide range of German temporal discourse markers in order to establish a set of features that characterize the different usage conditions of the markers. Table 1 lists the German temporal markers (and their approximate English translations) we examined; they are classified according to the three broad classes of temporal relations that are generally acknowledged in grammars: simultaneity, anteriority and posteriority (e.g., Helbig and Buscha 1991). In the following, we only discuss those dimensions that relate to ontology and verb semantics, namely the temporal relation signalled, and constraints regarding situation type and Aktionsart (see Grote (1998) for a discussion of lexical and syntactic constraints).

**Temporal relations** The three temporal relatons given in table 1 alone cannot reflect the differences in meaning between, for instance, *nachdem* (after) and *sobald* (as soon as) which both signal anteriority, but differ in that the latter requires both events to have a time point in common. Herweg (1991) argues that this explanation alone is not sufficient to account for the meanings of *nachdem* and *sobald*. Consider the following examples which are somewhat odd:

(1) *?Nachdem die Sonne aufgegangen war,*
After the sun had risen,
*ging er baden. Vorher war die*
he went swimming. Earlier had the
*Sonne wieder untergegangen.*
sun again set.
'?After the sun had risen, he went swimming. Earlier the sun had set again.'

(2) *?Sobald die Sonne aufgegangen war,*
As soon as the sun had risen,
*ging er baden. Vorher schlief er*
he went swimming. Earlier slept he
*aber noch.*
but still.
'?As soon as the sun had risen, he went swimming, but first he took a nap.'

In example (1), the subordinate clause event precedes the main clause event, but does not hold alnymore when the main clause event takes place; in example (2), subordinate and main clause event do not have a time point in common, because the nap occurs in between. To capture these phenomena, Herweg introduces additional constraints on temporal relations: *prox, next* and *imm*. For example, *prox* states that the state induced by the event in the subordinate clause still has to hold when the main clause situation takes place; this is not true for (1), and hence *nachdem* cannot be used. *Next* implies temporal adjacency between two events; this does not hold for *sobald* in example (2). In section 4, we provide definitions for Herweg's relations—and an additional *dur* relation which holds if the main clause event extends to speaking time[1]—in terms of our representation of time-stamped events.

**Types of situation** Sinn (1991) notes that it is not sufficient to describe constraints on situation type with general categories such as state or event; these cannot explain the following examples:

(3) *Nachdem der Tank leer gewesen war,*
After the tank empty had been,
'After the tank was empty, ...'

[1] 'Speaking time' is used in the sense of Reichenbach's (1947) Basic Tense Structure, which distinguishes between Event Time (E), Reference Time (R), and Speaking Time (S).

| marker group | temporal markers | |
|---|---|---|
| simultaneity | subc: | *seitdem (since), sobald (as soon as), solange (as long as), sowie (as soon as), während (while)* |
| | prep: | *während (during)* |
| anteriority | subc: | *kaum daß (no sooner), nachdem (after), seit(dem) (since), sobald (as soon as), sowie (as soon as)* |
| | prep: | *nach (after), seit (since)* |
| posteriority | subc: | *bevor (before), bis (until), ehe (before)* |
| | prep: | *bis (until), vor (before)* |

Table 1: *Subordinating conjunctions (subc) and prepositions (prep) examined in this study*

(4) *?Nachdem John Lennon tot*
After John Lennon dead
*gewesen war, ...*
had been, ...
'?After John Lennon had been dead, ... '

Apparently, *nachdem* can only be used if the state expressed in the subordinate clause is right-bounded, i.e., it is temporally followed by a contradictory state, which does not hold for the state of being dead in (4). Most German temporal markers are sensitive to the boundedness of situations, as section 5.1 will show.

**Aktionsart** Regarding the linguistic environment of markers, two kinds of interdependencies are generally acknowledged, see Herweg (1991), Helbig and Buscha (1991) and Bäuerle (1995): Aktionsart and aspect. With Aktionsart we refer to the inherent features that characterize facets of the situation denoted by a verb, for instance, whether it is iterative, durative, or stative (Bussmann 1983), while aspect refers to the non-inherent grammatical features, for German in particular to the perfective/imperfective distinction. Temporal markers often expect a particular Aktionsart of the verbs they connect, for instance, *während* cannot be used with non-durative verbs in the subordinate clause as in (5), and *solange* is odd when used with a resultative in the main clause (6):

(5) *Während das Kabel schmolz / ?riß,*
While the cable melted / ?tore,
*war ich nicht im Raum.*
I was not in the room.
'While the cable melted / ?tore, I wasn't in the room.'

(6) *?Solange ich in Berlin war, fuhr*
As long as I in Berlin was, drove
*Peter mit dem Auto bis nach München.*
Peter with the car to Munich.
'?As long as I was in Berlin, Peter drove with his car to Munich.'

In addition, temporal markers may shift the Aktionsart of a verb, for instance from a semelfactive to an iterative reading, as *solange* does in the following example:

(7) *Solange es still war im Haus,*
As long as it quiet was in the house,
*klopfte Tom.*
knocked Tom.
'As long as it was quiet in the house, Tom kept on knocking.'

In a nutshell, to support the motivated choice of German temporal discourse markers in NLG, we need an input representation that allows us to infer the temporal knowledge required (introduced in section 4) and a complex discourse marker representation that captures all aspects of marker meaning and usage (introduced in section 5).

## 4 Ontological definitions

### 4.1 Ontology and domain model

For defining the input structures for our generation framework, we build upon the ontology of situation specifications ('SitSpecs') used by Stede (1999) in the 'Moose' generator. The hierarchy of SITUATIONS, shown in Figure 1, follows proposals made in research on aspectual categories and resembles the hierarchy given by Bach (1986), but has some important differences.

As for STATES, Bach distinguished between static and dynamic ones, which for the purposes of (Stede 1999) was not relevant; nor is it in the

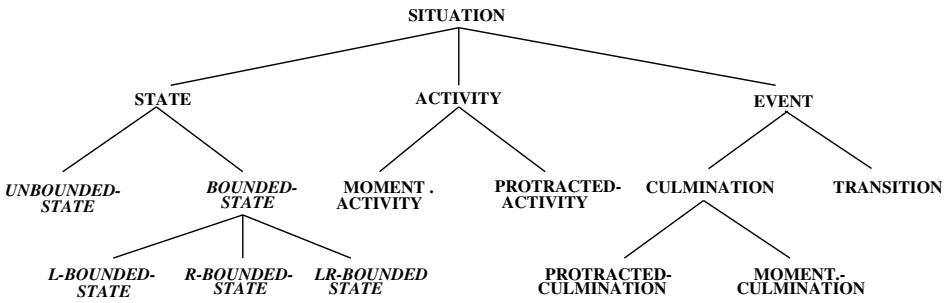


Figure 1: *Extended situation type ontology; additions to Stede (1999) are given in italics*

present work. However, we need to introduce the additional distinctions mentioned in section 3, in response to the observations by Sinn (1991). The BOUNDED feature states whether a situation has 0, 1, or 2 *inherent* temporal limits, which leads to corresponding sub-types of STATE. An example for an unbounded one (where a *from* <time> *to* <time> clause would make no sense) is *The earth is round*. A left-bounded state does not have a right boundary, and hence an *until* <time> clause cannot be added: ?*Jim was dead until Sunday*. The inverse holds for a right-bounded state, and for a bounded state, both left and right boundaries can be present in a sentence.

The remainder of the taxonomy is taken from Stede (1999). ACTIVITIES are quite similar to states, but there is always something "going on", as in *The water was flowing toward the sea*. We distinguish two subtypes here: PROTRACTED ACTIVITIES take place over an extended period of time, whereas MOMENTANEOUS ACTIVITIES occur in an instant. We take all activities to be bounded.

EVENTS are occurrences that have a structure to them; in particular, their result, or their coming to an end is included in them: *to destroy a building, to write a book*. As their central feature we take them to always involve some change of state: The building looses its integrity, the book comes into existence, or gets finished. Basically, we see any EVENT as involving a state change; an activity responsible for the change can *optionally* be present. A plain TRANSITION is necessarily momentaneous (*The room lit up*) (and it can only involve bounded states), whereas a transition-with-activity inherits its protracted/momentaneous feature from the embedded activity. We call these tripartite events CULMINATIONS[2]. They are composed of a pre-state (holding before the event commences), a post-state (holding when the event is over), and an activity that brings the transition about. Like the activities, events are always bounded, i.e. can happen *from* <time> or *until* <time>.

In order to enable a principled selection of temporal markers, we now need to extend the original SitSpecs with time-stamp annotations. For our present purposes, it is sufficient to use rather coarse-grained values and thus we simply take the numbers 1 to 12 as possible time-stamps. Then, STATES can have slots for begin- and end-time, depending on their boundedness. The three momentaneous situations have a single time-stamp; in the case of the TRANSITION, it marks the end of the pre-state and the beginning of the post-state, which are identical. A PROTRACTED-ACTIVITY can have begin- and end-time, and these carry over to a CULMINATION, i.e., the time-stamps of the embedded activity are the time-stamps of the entire culmination (which are identical with the end of the pre-state and the beginning of the post-state).

Subsumed by the general ontological system, Stede (1999) defines a domain model that holds the concepts relevant for representing situations and that specifies the exact conditions for their well-formedness. A network of instances of domain model concepts forms the input to the generator. To reflect our extensions with temporal information, we call these structures **TSitSpecs**. In analogy to SitSpecs, the root node of any TSitSpec is of type SITUATION. As an

[2]Moens and Steedman (1988) also use this term, but they restrict it to momentaneous events. Unfortunately, the terminology used in the literature for these kinds of categories varies so much that a 'standardization' seems out of reach.

| Point relations | Interval relation |
|---|---|
| $b_1 = b_2 \wedge e_1 = e_2$ | *equal*($Sit_1, Sit_2$) |
| $e_1 < b_2$ | *before*($Sit_1, Sit_2$) |
| $e_2 < b_1$ | *after*($Sit_1, Sit_2$) |
| $e_1 = b_2$ | *meets*($Sit_1, Sit_2$) |
| $e_2 = b_1$ | *meets-i*($Sit_1, Sit_2$) |
| $b_1 < b_2 \wedge e_1 < e_2 \wedge b_2 < e_1$ | *overlaps*($Sit_1, Sit_2$) |
| $b_2 < b_1 \wedge e_2 < e_1 \wedge b_1 < e_2$ | *overlaps-i*($Sit_1, Sit_2$) |
| $b_2 < b_1 \wedge e_1 < e_2$ | *during*($Sit_1, Sit_2$) |
| $b_1 < b_2 \wedge e_2 < e_1$ | *during-i*($Sit_1, Sit_2$) |
| $b_1 = b_2 \wedge e_1 < e_2$ | *starts*($Sit_1, Sit_2$) |
| $b_1 = b_2 \wedge e_2 < e_1$ | *starts-i*($Sit_1, Sit_2$) |
| $b_2 < b_1 \wedge e_1 = e_2$ | *finishes*($Sit_1, Sit_2$) |
| $b_1 < b_2 \wedge e_1 = e_2$ | *finishes-i*($Sit_1, Sit_2$) |

Table 2: *Mapping point representations to Allen's interval relation*

example, the EVENT of a person named Jill filling a tank with water is shown in Figure 2 in a graphical description logic notation, with relation names appearing in boxes. The event combines the activity of Jill pouring water into the tank with the fill-state of the tank changing to full. A verbalization of this event can emphasize either of these aspects. The slots `t-begin` and `t-end` give the temporal extension of the situation parts.

When the sentence generator maps portions of a TSitSpec to words, the denotations of verbs correspond to the structure of the event, which thus identifies the Aktionsart of the verb (in straightforward cases where no aspectual composition is involved): For instance, an ACTIVITY is expressed by a durative verb, a TRANSITION by a semelfactive one, a PROTRACTED CULMINATION by a resultative one.

### 4.2 Determining temporal relations

When the task is to produce a complex sentence with a temporal marker from two TSitSpecs, a prerequisite is to infer the temporal relation from the time-stamps. Following, for instance, Dorr and Gaasterland (1995), Klenner (1991) and Herweg (1991), we adopt Allen's framework of temporal interval relations (Allen 1984) to describe the temporal relation indicated by temporal connectives. Allen defines seven basic temporal relations, namely *equals(=), after(>), during(d), overlaps(o), meets(m), starts(s), finishes(f)*, and their inverses, which may hold between two situations.

Our input structure, as described above, consists of a speaking time S and two TSitSpecs ($Sit_1$ and $Sit_2$) with time points `t-begin` and `t-end`, i.e. $(b_1, b_2, e_1, e_2)$, such that $(b_1 \leq e_1)$ and $(b_2 \leq e_2)$. Assuming the three point relations $(=, >, <)$, each of Allen's interval relations can be described by the point relations holding between the pairs $(b_1, b_2), (b_1, e_2), (e_1, b_2)$ and $(e_1, e_2)$. A subset of these relations is sufficient to infer the interval relation; the correspondences are listed in table 2.

In section 3 we argued that, following Herweg (1991), additional temporal constraints are required to capture the exact meaning of a temporal marker. In our framework, they are defined as follows:

- $prox(Sit_1, Sit_2)$ holds if $Sit_2$ has a time point in common with the post-state ($Sit_3$) of $Sit_1$: $(b_3 \leq b_2) \vee (b_2 \leq e_3)$.
- $next(Sit_1, Sit_2)$ holds if there is no other situation ($Sit_3$) in the input structure that is located between $Sit_1$ and $Sit_2$: $\neg\exists Sit_3[(e_1 < b_3) \wedge (b_3 < b_2)]$.
- $imm(Sit_1, Sit_2)$ realizes the idea of 'temporal adjacency' of two situations. We posit that one time point be the maximal time span allowed between $Sit_1$ and $Sit_2$: $(e_1 + 1 = b_2) \vee (b_1 + 1 = b_2)$.
- $dur(Sit_i)$ holds if the end point of $Sit_i$ equals or extends beyond speaking time S: $e_i \geq S$.

## 5 Representing temporal markers

### 5.1 Discourse marker lexicon

In order to perform an informed choice of discourse markers in generation, we assume an independent lexical resource holding information about such markers, as outlined in (Grote, Stede 1998). This discourse marker lexicon stores three types of information for each entry:

- **Applicability conditions:** The necessary conditions that need to be present in the input representation for the marker to be a verbalization candidate.
- **Combinability constraints**: The constraints that the marker imposes on its neighbouring linguistic constituents.
- **Distinguishing features**: for preferential choice.

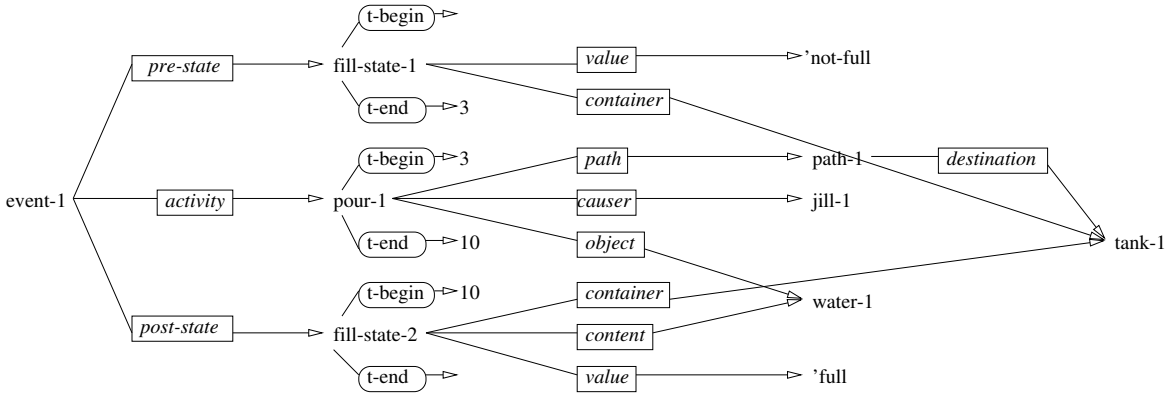


Figure 2: *TSitSpec representing a fill-event lasting from 3 to 10*

| Feature | Possible values |
|---|---|
| name | [name of lexicon entry] |
| spelling | {markers given in table 1} |
| temprel | {Allen's relations}, prox, next, imm, dur |
| sittype mc | {situation types} |
| sittype sc | {situation types} |
| Aktart mc | {Aktionsarten} |
| Aktart sc | {Aktionsarten} |
| aspect mc | perfective, imperfective |
| aspect mc | perfective, imperfective |
| tense | relation of E(mc), E(sc) and S |
| syncat | subconj, prep |
| quantify | quant, nonquant |
| style | neutral, brief, formal, judging |

Table 3: *Features and values of lexicon entries*

For our temporal markers and the features identified in section 3, this means: Temporal relations and situation types are applicability conditions (given by the TSitSpecs and an additional inference step). Aktionsart, aspect and tense are additional means to signal temporality in text; these together with syntactic structure and quantification form the combinability constraints of temporal markers. Stylistic variation is a distinguishing feature, but plays only a limited role for temporal markers.

### 5.2 Lexicon entries

The feature values of our lexicon entries result mainly from an analysis of marker occurrences in German texts, employing a substitution test as described by Knott and Mellish (1996); these findings are supplemented by results from the research literature, as mentioned in section 3. Table 3 lists the features and their possible fillers that we use for representing temporal markers, divided in the zones given above. The values for the applicability conditions (temporal relations (`temprels`), situation type (`sittype`)) have been discussed in section 4; likewise the Aktionsart values (`Aktart`). Tense constraints are expressed using Reichenbach's (1947) BTS notation, and define the legal ordering of event time E and speaking time S. `Syncat` gives the syntactic category of the marker (which entails constraints on the syntactic type of conjoined material), `quantify` says whether a quantifier can be added to the marker, and finally, `style` accounts for stylistic variation. Several features can apply to main and subordinate clause separately; these have an `mc` or `sc` suffix to mark their scope.

We have developed lexical representations for the markers given in table 1, but for brevity show only six examples here. Table 4 gives the lexicon entries for four 'anterior' and two 'simultaneity' markers. If a marker involves no constraint for a particular feature, the slot in the table remains empty. Within value sets, hyponomy relations hold, for instance, *next* is more specific than *prox*, and *after* more general than *after*$\wedge$*imm*. Notice that the major differences within the anterior group concern the kind of temporal relation they signal, while the simultaneity markers show variation in combinability conditions (mainly Aktionsart).

| **name** | nachdem-ant | sobald-ant | kaum-dass-ant | seitdem-ant | waehrend | solange |
|---|---|---|---|---|---|---|
| **spelling** | nachdem | sobald | kaum daß | seitdem | während | solange |
| **temprel** | after∨ meets-i prox, dur | (after∧imm)∨ meets-i, next | (after∧imm)∨ meets-i, next | (after∧imm)∨ meets-i, dur, next | equal∨overlaps∨ during∨starts∨ finishes∨*inverses* | finishes |
| **sittype mc** | l-bounded | l-bounded | l-bounded | l-bounded | | |
| **sittype sc** | r-bounded | r-bounded | r-bounded | r-bounded | lr-bounded | r-bounded |
| **Aktart mc** | | | | | | not-result. |
| **Aktart sc** | | | | durative ∨ stative | durative ∨ iterative | durative ∨ iterative |
| **aspect mc** | imperfective | | | | | |
| **aspect sc** | perfective | perfective | perfective | perfective | | |
| **tense** | $E(sc) < S \wedge E(mc) \leq S$ | $E(sc) < S \wedge E(mc) \leq S$ | $E(sc) < S \wedge E(mc) < S$ | $E(sc) < S \wedge E(mc) \leq S$ | $E(sc) = E(mc)$ | $E(sc) = E(mc)$ |
| **syncat** | subconj | subconj | subconj | subconj | subconj | subconj |
| **quantify** | quant | nonquant | nonquant | nonquant | nonquant | nonquant |
| **style** | neutral | neutral | judging, formal | neutral | neutral | neutral |

Table 4: *Lexicon entries for some German temporal markers ('inverses' is a shorthand for the inverses of the temporal relations given for 'während').*

## 6 Generating complex temporal clauses

Given the time-stamped input representations and the discourse marker lexicon, we now turn to the procedure of producing a (set of) complex temporal clause(s) that adequately signals the temporal relation holding between the given situations. At present, our work is restricted to sentence-pairs in a temporal relation only; we do not consider the broader discourse context, and we thus ignore possible additional relations (such as causal links), which in general is a simplification (cf. Moens and Steedman 1988). We do not address the selection of tense, either; however, Dorr and Gaasterland (1995) have shown how complex tenses can be derived from time-stamped event representations.

### 6.1 Procedure

We assume that decisions about focusing on propositions have already been made, so that one of the two input TSitSpecs is marked as temporal locator (*TempLoc*, the situation that serves as temporal 'anchor' for the other situation; this serves to distinguish, e.g., *after A, B* from *before B, A*). Also, we take the speaking time S as additional input. Figure 3 shows a procedure for generating temporal clauses from this information, on the basis of ontological knowledge and the temporal marker lexicon.

**Input:** two TSitSpecs, one with *TempLoc* label, speaking time $S$

1. infer the temporal relations holding between $Sit_1$ and $Sit_2$, assuming $Sit_2$ = *TempLoc*
2. test $Sit_1$ and $Sit_2$ using $S$ for *prox, next, imm, dur* and obtain situation types
3. determine set of possible markers by matching results from Step1 and Step2 with lexicon entries
4. choose verbalizations for propositions (yields Aktionsart)
5. choose marker that satisfies the constraints posited by verb (and other sentence planning modules)
6. produce SemSpec using MOOSE (Stede 1999)
7. send to surface generator (KPML, Bateman 1997)

**Output:** a (set of) complex temporal clause(s)

Figure 3: *Procedure for clause generation*

In this paper, we focus on the role of ontological knowledge and lexical semantics in marker choice, and hence on steps 1 to 5 of our procedure. Having completed step 5, we know the set of temporal markers that meet the applicability conditions imposed by the two TSitSpecs, and that satisfy the Aktionsart constraints for each proposition partaking in the temporal relation. For verbalizing propositions via the intermedi-

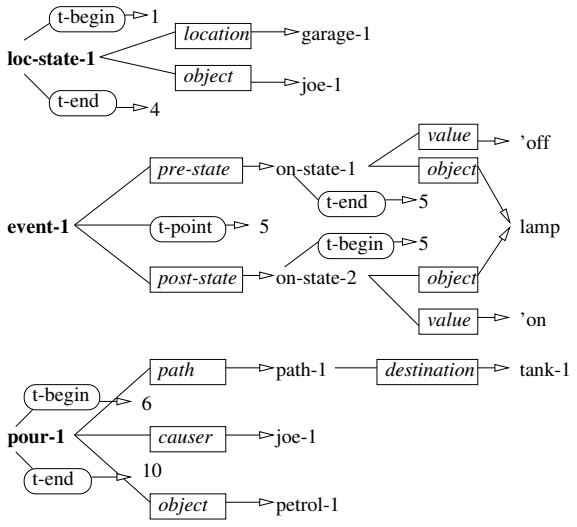


Figure 4: *TSitSpecs in temporal succession*

ate level of 'SemSpecs', we use the approach described in (Stede 1999); our extension of building SemSpecs for complex sentences, using lexical entries for discourse markers, were explained in (Grote, Stede 1998).

### 6.2 Examples

**Joe's garage** Figure 4 shows three situations in temporal succession; each pair of TSitSpecs can serve as input to our procedure. Consider the first two TSitSpecs (example $joe_1$): The procedure is called with loc-state-1 (= *TempLoc*), event-1, $S$=10.

Using the mapping in table 2, step1 infers the interval relation: *after(event-1,loc-state-1)*. With the definitions from section 4.2, step2 determines that the relations *prox(loc-state-1,event-1)*, *imm(loc-state-1,event-1)* and *next(loc-state-1,event-1)* hold. Given these relations and the situation types (TRANSITION and BOUNDED-STATE, which can be read off the TSitSpecs), the temporal marker lexicon given in figure 4 is accessed in step3: The results from step1 and step2 are matched against the applicability conditions, resulting in a set of marker candidates with the correct denotation: *sobald, kaum daß*. *Seitdem* fails because *dur(loc-state-1)* does not hold (event-1 does not extend to speaking time $S$).

Step4 now chooses the verbs, here *angehen* (turn on) for event-1 and *sein* (be) for loc-state-1, thus giving the Aktionsarten semelfactive and stative. These do not restrict the set of markers any further (cf. table 4). When combining all the lexical information, respecting the constraints on syntactic realization given in the lexicon entries for *sobald* and *kaum daß*, into SemSpecs, we get the results:

(8) *Sobald Joe in der Garage*
As soon as Joe in the garage
*gewesen war/ist, ging/geht die*
had/has been, turned/turns the
*Tankleuchte an.*
lamp on.
'As soon as Joe had/has been in the garage, the lamp turned/turns on.'

(9) *Kaum daß Joe in der Garage*
No sooner than Joe in the garage
*gewesen war, ging die Tankleuchte an.*
had been, turned the lamp on.
'No sooner than Joe had been in the garage, the lamp turned on.'

Choosing among these options is the task of the sentence planner in step5; variation between *sobald* and *kaum daß* is merely stylistic in that *kaum daß* expresses the speaker judgement that the situation in the main clause occurred too quickly.

Generation of the remaining pairs (examples $joe_2$ and $joe_3$) works in a similar manner, therefore we compactly present the results of the different steps in figure 5. Notice that despite the fact that pour-1 in example 3 extends to speaking time (i.e. *dur(pour-1)* holds), *seitdem* is ruled out, because *imm(loc-state-1,pour-1)* and *next(loc-state-1,pour-1)* do not apply. Instead, *nachdem* is chosen. These examples illustrate how ontological knowledge and complex event representations enable the selection of an adequate temporal discourse marker.

**Lucy's diamonds** A central feature of Stede's (1999) generation approach is its ability to express complex events in different ways, placing emphasis either on the fact that the post-state has been reached, or on the activity leading toward that post-state. For the TSitSpec given in figure 2, this means, for example, a choice between *Jill filled the tank with water* and *Jill poured water into the tank*. Using this example, we demonstrate the role of Aktionsart in discourse marker generation.

Suppose that the event in figure 2 is combined with one of Lucy polishing dimanonds, which is an ACTIVITY and would be represented similar

| **Example** $joe_2$: | |
|---|---|
| **Input:** | event-1 (= *TempLoc*), pour-1, *S*=10 |
| **Step1:** | *after(pour-1,event-1)* |
| **Step2:** | *prox(event-1,pour-1), next(event-1,pour-1), imm(event-1,pour-1), dur(pour-1)*<br>event-1 = TRANSITION, pour-1 = PROTRACTED-ACTIVITY |
| **Step3:** | *seitdem* |
| **Step4:** | transformative, durative |
| **Step5:** | *seitdem* |
| ... | |
| **Output:** | *Seitdem die Tankleuchte angegangen ist, gießt Joe Benzin in den Tank.*<br>Since the tank lamp has turned on, pours Joe petrol into the tank.<br>'Since the tank lamp has turned on, John has been pouring petrol into the tank.' |

| **Example** $joe_3$: | |
|---|---|
| **Input:** | loc-state-1 (= *TempLoc*), pour-1, *S*=10 |
| **Step1:** | *after(pour-1,loc-state-1)* |
| **Step2:** | *prox(loc-state-1,pour-1), dur(pour-1)*<br>loc-state-1 = LR-BOUNDED-STATE, pour-1 = PROTRACTED-ACTIVITY |
| **Step3:** | *nachdem* |
| **Step4:** | stative, durative |
| **Step5:** | *nachdem* |
| ... | |
| **Output:** | *Nachdem Joe in der Garage gewesen ist, gießt Joe Benzin in den Tank.*<br>After Joe in the garage has been, pours Joe petrol into the tank.<br>'After Joe has been in the garage, Joe has been pouring petrol into the tank.' |

Figure 5: *Results of Step1 to Step5 for situation pairs (event-1,pour-1) and (loc-state-1,pour-1)*

to the third TSitSpec in figure 4. We assume that `t-begin=2` and `t-end=10`, in other words, that the two situations are co-extensive, and end at the same time point. This gives the temporal relation *finishes*. When expressing the ACTIVITY of Jill pouring water, the Aktionsart of the verb *gießen* (pour) is durative, which combines with *solange* (cf. table 4):

(10) *Solange Lucy die Diamanten polierte,*
As long as Lucy the diamonds polished,
*füllte Jill Wasser in den Tank.*
poured Jill water into the tank.
'As long as Lucy was polishing the diamonds, Jill was pouring water into the tank.'

If we prefer to emphasize the resulting fill-state, however, *solange* does not work, because now we have to use *füllen* (fill) in its resultative Aktionsart:

(11) *?Solange Lucy die Diamanten polierte,*
As long as Lucy the diamonds polished,
*füllte Jill den Tank mit Wasser.*
filled Jill the tank with water.
'As long as Lucy was polishing the diamonds, Jill was filling the tank with water.'

In this case, as the dictionary entries show, only the highly ambiguous *während* is appropriate (signalling 9 different temporal relations), thereby losing the information of the simultaneous ending on both situations:

(12) *Während Lucy die Diamanten polierte,*
While Lucy the diamonds polished,
*füllte Jill den Tank mit Wasser.*
filled Jill the tank with water.
'While Lucy was polishing the diamonds, Jill was filling the tank with water.'

Interdependencies of Aktionsart and marker choice concerning the +/-durative variation are far more frequent than the +/-telic variation just illustrated. Most temporal markers have a durative counterpart; pairs are for instance, *nachdem/seitdem* and *bevor/bis*. Using the complex event representation, the knowledge on how verbal Aktionsart relates to these, and our marker lexicon, we can produce pairs of temporal clauses reflecting the +/-durative opposition, as in *Nachdem die Ampel grün geworden war, ...* (After the lights turned green ...) and *Seitdem die Ampel grün war, ...* (After the lights were green, ...). Here, the TRANSITION in the sub-

ordinate clause can be realized by a transformative verb signalling the beginning of a state and *nachdem*, or a durative verb plus the durative connective *seitdem*.

## 7 Summary

Language offers many ways to express temporal relationships between events, and exploiting them in NLG presupposes a detailed analysis of the semantics of temporal discourse markers. We have undertaken such a study for about 20 German markers, and then fused the results with an (already existing) sentence generation framework that can produce a range of paraphrases from the same input. We augmented the input representations used for that generator with time stamps, introduced additional categories to the ontology, and extended the approach to the production of complex sentences that include an appropriate temporal marker.

The generation procedure has been sketched in section 6, and its implementation on the basis of pre-existing modules is currently under way. We envisage to integrate this work, which focused on the production of temporal markers, into a general framework for paragraph generation, which employs a dedicated discourse marker lexicon to enable motivated choices of relation signals during the sentence planning phase.

**Acknowledgement** For helpful comments on improving earlier versions of this paper we thank John Bateman and two anonymous reviewers.

## References

J. Allen. Towards a general theory of action and time. *Artificial Intelligence*, 23(2), 1984.

E. Bach. The Algebra of Events. In: *Linguistics and Philosophy* 9, 1986.

J. Bateman. Enabling Technology for Multilingual Natural Language Generation: The KPML Development Environment. In *Journal of Natural Language Engineering*, 3(1), 15-55, 1997.

R. Bäuerle. Temporalsätze und Bezugspunktsetzung im Deutschen. In B. Handwerker (ed.) *Fremde Sprache Deutsch.* Tübingen: Gunter Narr, 1995.

H. Bussmann. *Lexikon der Sprachwissenschaft.* Stuttgart: Kröner, 1983.

B. Dorr, T. Gaasterland. Selecting tense, aspect and connecting words in language generation. In *Proc. of the 14th International Joint Conference on Artificial Intelligence*, Montreal, 1995.

M. Gagnon, G. Lapalme. From conceptual time to linguistic time. In: *Computational Linguistics* 22(1), 1996.

B. Grote. Representing temporal discourse markers for generation purposes. In: *Proc. of the COLING-ACL workshop on Discourse Markers and Discourse Relations.* Montreal, 1998.

B. Grote, M. Stede. Discourse marker choice in sentence planning. In *Proc. of the 9th International Natural Language Generation Workshop.* Niagara-on-the-Lake, Canada, 1998.

G. Helbig, J. Buscha. *Deutsche Grammatik: Ein Handbuch für den Ausländerunterricht.* Berlin, Leipzig: Langenscheidt, Verlag Enzyklopädie, 1990.

M. Herweg. Temporale Konjunktionen und Aspekt: Der sprachliche Ausdruck von Zeitrelationen zwischen Situationen. In *Kognitionswissenschaft* 2, 1991.

M. Klenner. TELOS: Ein System zur Generierung temporaler Konjunktionalsätze. Research Report 75, Psychological Institute, University of Freiburg, Germany, 1991.

A. Knott, C. Mellish. 1996. A Feature-based Account of the Relations Signalled by Sentence and Clause Connectives. In *Language and Speech* 39(2-3).

M. Moens, M. Steedman. Temporal Ontology and Temporal Reference. In: *Computational Linguistics* 14(2), 1988.

H. Reichenbach. *Elements of Symbolic Logic.* London: Macmillan, 1947.

H. Sinn. Situationaler Fokus: Prinzipien der Generierung temporaler Konjunktionen. In *Kognitionswissenschaft* 2:136-155, 1991.

A. Steube. *Temporale Bedeutung im Deutschen.* studia grammatica XX. Berlin: Akademie-Verlag, 1980.

M. Stede. *Lexical semantics and knowledge representation in multilingual text generation.* Dordrecht: Kluwer, 1999.