

UNDERSTANDING OF TIME CONSTITUENTS IN SPOKEN LANGUAGE DIALOGUES

Bernd Hildebrandt, Gernot A. Fink, Franz Kummert, Gerhard Sagerer

Universität Bielefeld, Technische Fakultät, AG Angewandte Informatik
Postfach 100 131, 33501 Bielefeld, Federal Republic of Germany
Tel.: +49 521 106 2935, Fax: +49 521 106 2992

ABSTRACT

The analysis and interpretation of time constituents is a rather complex enterprise, since diverse time constituents are distributable in variable positions within an utterance. The first step in order to manage this complexity and variability is the syntactic analysis at *phrase structure level*. Within a utterance each time constituent is analyzed independently and tested for its syntactic coherence. A semantic interpretation of the time constituent has to follow. The second step consists of the analysis and interpretation at *sentence structure level*. The time interpretations need to be tested for consistency and merged into a single representation. Here it is usually possible to resolve ambiguities. As a last step, the interpretations of time constituents have to be merged at *dialogue level*. Although the system asks the user for verification of the time computed, users do seldom just reply 'yes' or 'no'. Mostly they add new information about time, and sometimes users correct the system's interpretation without explicit negation. Thus, the merging of time constituents at dialogue level becomes a rather complex issue.

1. INTRODUCTION

For many applications of speech understanding systems (SUNDIAL, ATIS, ASL, VERBMOBIL) the analysis and interpretation of time constituents is important. Since diverse time constituents are distributable in variable positions within an utterance, problems of modeling can emerge. For example, in German it is possible to say „Ich möchte *am Dienstag um acht Uhr abends* nach Bielefeld fahren" (I want to go to Bielefeld on Tuesday at 8 o'clock in the evening.) or „*Am Dienstag* möchte ich *abends um acht Uhr* nach Bielefeld fahren". Both sentences have the same meaning and in spoken language even the following utterance would be acceptable: „*am Dienstag* möchte ich *abends* nach Bielefeld *um acht Uhr*."

This research was supported by the German Ministry of Research and Technology (BMFT) under grant number 01IV102G/7. Only the authors are responsible for the contents of this publication.

The linguistic analysis is based on a semantic network representation of linguistic knowledge using the ERNEST formalism [6, 4, 5]. ERNEST makes a uniform representation of all knowledge that is needed for a linguistic analysis possible. The aim of this linguistic analysis is the symbolic interpretation of speech data and the instantiation of a concept that represents each dialogue step of the user. In the process of this analysis, separate time constituents are found in several sentences and in subsequent dialogue steps. The semantic interpretations of these time constituents have to be merged step by step to specify the time intended more and more.

2. MODELING OF TIME CONSTITUENTS

Time constituents are one kind of syntactic unit, which organizes semantic concepts. In a corpus of domain specific utterances (train schedule) four types of time constituents can be found [3]. One type is the *question phrase*, which asks for an answer about time. Other types are the *time of day* (e.g. 'um acht Uhr und zwölf Minuten' (at eight hours and twelve minutes)), the *section of day* (e.g. 'am frühen Abend' (early in the evening)), and the *date* (e.g. 'am fünften Mai' (May 5th)). In spoken language only one of the four types or a combination of up to three of the types mentioned can be found within a single utterance. Furthermore, in spoken language there does not seem to be any rule that restricts the position or the combination of time constituents in a German sentence.

2.1 Modeling at phrase structure level

The first step in order to manage this complexity and variability is the syntactic analysis at *phrase structure level*. Each time constituent of an utterance is analyzed independently and tested for its syntactic coherence. A semantic interpretation of every time constituent follows. Initially, the *core time* of a time constituent has to be computed, e.g. *May 5th* is the core time not only of 'am fünften Mai' (May 5th) but also of 'vor dem fünften Mai' (before May 5th) and of 'nach dem fünften Mai' (after May 5th). After that the *time of reference* [7, 1] is

computed depending on the temporal relation *before*, *on* or *after* and depending on the *time of speech* and on the core time mentioned. The result is a time interval within which the intended action (going by train) is meant to take place. A so-called *time table* is used to represent such a time interval.

	from	until
day		
month		
hour	time of speech	18
minute		

Figure 1 Time table to represent the interpretation of the time constituent 'vor achtzehn Uhr' (before 6 p.m.)

For example, Figure (1) shows the representation of the semantic interpretation of the time constituent 'vor achtzehn Uhr' (before 6 p.m.). The core time is six o'clock p.m. But since the temporal relation is *before*, the intended action is meant to take place within an interval between time of speech and six o'clock p.m..

2.2 Modeling at sentence structure level

The second step consists of the analysis and interpretation at *sentence structure level*. After the treatment of separated time constituents the time interpretations need to be tested for consistency and merged into one single representation; i.e. if the system has two or more items of information, let us say the date and the time of day, the system can merge them. On top of that, at sentence structure level, it is often possible to resolve ambiguities. In German isolated time expressions are often ambiguous; e.g. 'um acht Uhr' can refer to 8 a.m. or 8 p.m.. If sufficient information is given, like a time constituent expressing section of day, such an ambiguity can be resolved (see Example 1).

<p>am neunten Mai möchte ich um acht Uhr abends nach Bielefeld fahren (on May 9th I want to go at 8 o'clock evening to Bielefeld)</p>

Example 1

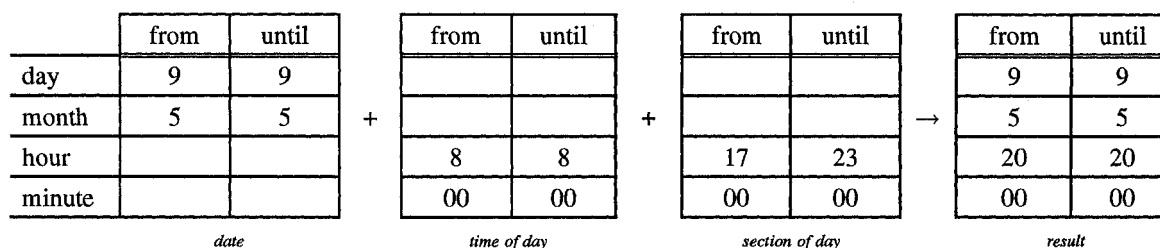


Figure 2 Merging of time constituents

Figure (2) shows the steps of interpretation. First, the date, the time of day and the section of day are interpreted independently: 'abends' (in the evening) is currently defined as 5–11 p.m., and taking 'um acht Uhr' (at 8 o'clock) literally in German means 08.00 a.m.. Secondly, date, time of day and section of day are merged, resolving the ambiguity of 'um acht Uhr' into 08.00 p.m..

2.3 Modeling at dialogue level

As a last step, interpretations of time constituents must be merged at *dialogue level*. If the system has found an utterance with time constituents, it asks the user for verification of the time computed. Users seldom reply just 'yes' or 'no'. Mostly they add new information about time. In such a case, the system begins to construct a time representation at sentence structure level as shown above. Then, it attempts to merge the new time representation with the former one.

<p>User: ich möchte am neunten Mai um acht Uhr nach Hamburg fahren (I want to go May 9th at eight o'clock to Bielefeld)</p> <p>System: Sie möchten von Bielefeld nach Hamburg fahren und am neunten Mai um acht Uhr von Bielefeld abfahren? (do you want to go from Bielefeld to Hamburg; and do you want to leave Bielefeld May 9th at 8 o'clock a.m.?)</p> <p>User: nein, um acht Uhr abends (no, at eight o'clock in the evening)</p> <p>System: Sie möchten von Bielefeld nach Hamburg fahren und am neunten Mai um zwanzig Uhr von Bielefeld abfahren? (do you want to go from Bielefeld to Hamburg; and do you want to leave Bielefeld May 9th at 8 o'clock p.m.?)</p> <p>User: ja genau (yes exactly)</p>
--

Example 2

If the user confirms explicitly the system's interpretation, the merging algorithm is similar to the one at sentence structure level. On the other side, explicit negations by users do not always mean that they negate the

whole interpretation, which would compel the system to start anew. Quite often, the negation only refers to one single time constituent type like the date or to the time of day (see Example 2). In such a case, the first thing the system has to do is detect that time constituent types which remain consistent; and second, the system has to substitute an inconsistent former time constituent by a new time constituent. In Figures (3—5) the main steps of merging and correcting time constituents at dialog level can be seen.

User: morgen abend möchte ich um acht Uhr nach Hamburg fahren.
 (tomorrow | evening | I want to go | at eight o'clock | to Hamburg)
 System: *Sie möchten von Bielefeld nach Hamburg fahren und morgen abend um neun Uhr von Bielefeld abfahren?*
 (do you want to go from Bielefeld to Hamburg; and do you want to leave Bielefeld tomorrow evening at nine o'clock?)
 User: nein, um acht Uhr
 (no, at eight o'clock)

Example 3

In spontaneous dialogues users sometimes correct the system's interpretation without explicit negation. Thus, the system only detects an inconsistency between the former and the new time interpretation. In both, explicit and

implicit cases of correction, the system has to find out, what kind of negation is meant and has to react accordingly in the way shown above.

In a speech understanding system it can happen, that the system does not understand the user's utterance correctly, e.g. in Example (3) '9 o'clock' instead of '8 o'clock'. The user will probably correct it making the following utterance: "Nein, um acht Uhr" (no, at 8 o'clock). Now the system has one partially false interpretation and produces another correct but incomplete interpretation. The result of merging both interpretations is also incorrect (see Figure 6), because the information of the section of day constituent is dropped. Therefore, the merging process has to take only those unmerged interpretations of previous dialogue steps made at phrase structure level which are consistent with the most recently merged interpretation (see Figure 7). This leads us to a correct interpretation.

Furthermore, it seems reasonable that the system is able to ask for uncertain or missing bits of information immediately. An uncertain item of information is, for example, an ambiguous time constituent like 'um zwei Uhr' (at two o'clock). Users seldom want to go by train at 2 o'clock at night, although it might happen. In those cases in which clues for disambiguation are missing the system has to be able to ask for exact information. The system ought to react in the same way, if it has no information about the day. The drawing of inference using the current day as default is not a reliable strategy.

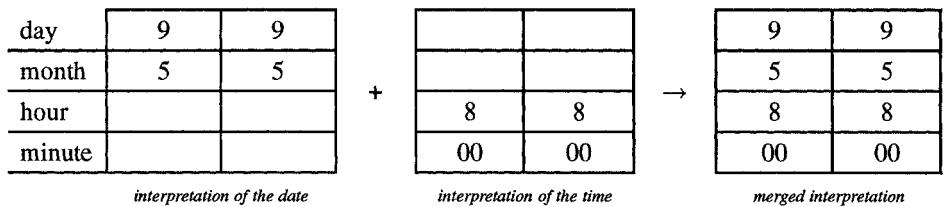


Figure 3 The interpretation of the user's first utterance

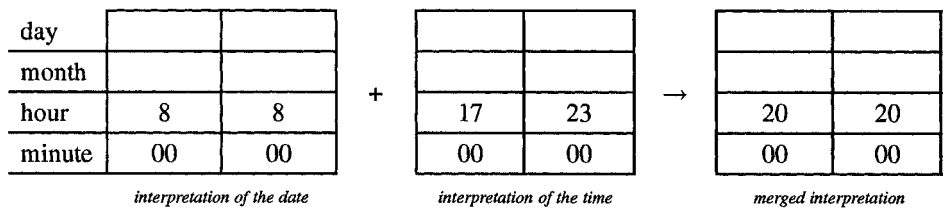


Figure 4 The interpretation of the user's second utterance

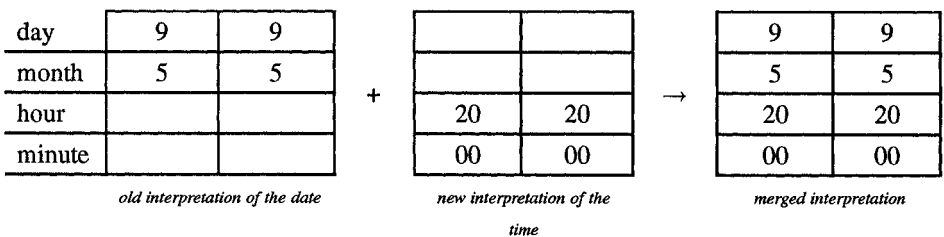


Figure 5 The merging of the user's first and second utterances

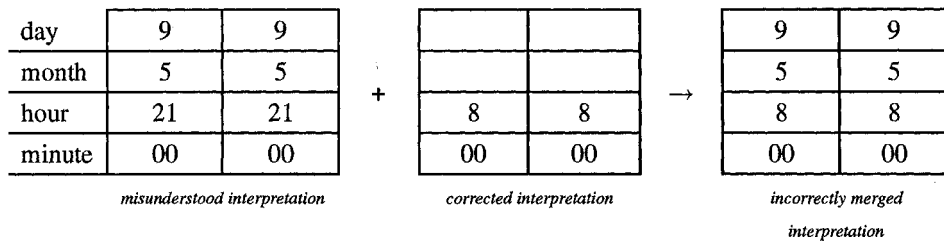


Figure 6

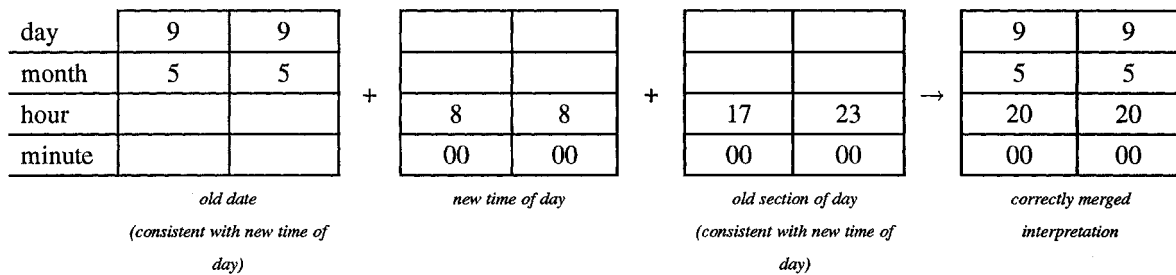


Figure 7

3. FIRST RESULTS

To assess the reliability of the modeling presented, 36 sentence types with time constituents were analyzed. These sentence types were taken from a small corpus of 51 dialogues of spontaneous utterances of 43 naive speakers collected at Hannover Industrial Fair 1993. 29 sentence types were successfully analyzed and interpreted. Some of the remaining seven sentence types could only be analyzed partially (see Figure 8). Nevertheless, because of the robustness of the system [2] an acceptable train schedule was produced for every request. Other sentence types were correctly analyzed, but under pragmatic considerations the interpretations were questionable. For example, the ambiguity of time of day constituents mentioned leads to an insufficient interpretation if the phrase is 'am Freitag um zwei Uhr' (on Friday at 2 o'clock), since the output will be 2 o'clock a.m., which is unlikely to be correct. As shown above, in a complete dialogue the user is able to correct the system's interpretation if necessary by adding bits of information. All in all, 80,5% of time structures found in the corpus were analyzed completely. A similar test for dialogues is in preparation.

total	completely analysed	satisfactorily analysed
36	29	7
100%	80,5%	19,5%

Figure 8 Results

REFERENCES

- [1] R. Bäuerle. *Temporale Deixis, temporale Frage. Zum temporalen Gehalt deklarativer und interrogativer Sätze*. Narr, Tübingen, 1979.
- [2] G. A. Fink, F. Kummert, G. Sagerer, and B. Seestaedt. Robust interpretation of speech. In *Proc. European Conf. on Speech Communication and Technology*, Berlin, 1993.
- [3] B. Hildebrandt, G. A. Fink, F. Kummert, and G. Sagerer. Modelling of time constituents for speech understanding. In *Proc. European Conf. on Speech Communication and Technology*, Berlin, 1993.
- [4] F. Kummert, H. Niemann, R. Prechtel, and G. Sagerer. Control and Explanation in a Signal Understanding Environment. *Signal Processing, special issue on 'Intelligent Systems for Signal and Image Understanding'*, 32:111-145, 1993.
- [5] M. Mast, F. Kummert, U. Ehrlich, G. A. Fink, T. Kuhn, H. Niemann, and G. Sagerer. A Speech Understanding and Dialog System with a Homogeneous Linguistic Knowledge Base. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 16:179-193, 1994.
- [6] H. Niemann, G. Sagerer, S. Schröder, and F. Kummert. ERNEST: A Semantic Network System for Pattern Understanding. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 12:883-905, 1990.
- [7] H. Reichenbach. *Elements of symbolic logic*. Free Press, New York, 1947.