

Bedeutungsanalyse und Interpretation von Äußerungen
relativ zu einem Dialogmodell im Spracherkennungs- und
Dialogsystem EVAR
Interpretation of Meaning and Interpretation of Utterances
in the Context of a Dialog Model in the Speech
Understanding and Dialog System EVAR

U. Ehrlich*, M. Mast und H. Niemann
Lehrstuhl für Informatik 5 (Mustererkennung),
Friedrich-Alexander-Universität Erlangen-Nürnberg

28. August 1991

Zusammenfassung

Verstehen von gesprochener Sprache wie im Spracherkennungs- und Dialogsystem EVAR (siehe [Nie88]) erfordert die Bedeutungsinterpretation von Worthypothesenkettens in einem vorgegebenen Anwendungsbereich. In EVAR besteht diese Analyse aus zwei Schritten:

- Anwendungsunabhängige semantische Analyse:
Es werden sowohl syntaktische Konstituentenhypothesen bezüglich semantischer Kombinations- und Selektionsrestriktionen verifiziert als auch mehrere syntaktische Konstituentenhypothesen nach syntaktisch-semantischen Kriterien zu größeren Einheiten zusammengefügt und strukturell interpretiert.
- Pragmatische Analyse:
Diese semantischen Strukturen werden im Anwendungskontext „Intercity-Auskunft“ interpretiert, wobei insbesondere die Bestimmung von Zeitbereichen, über die gesprochen wird, eine Rolle spielt.

Für die Interpretation von Zeitangaben und auch für die Auflösung anaphorischer Referenzen wird sowohl während der semantischen Analyse als auch während der pragmatischen Analyse ‚Wissen‘ aus dem situationellen und sprachlichen Kontext integriert.

Um einen Dialog führen zu können, ist eine Interpretation im Dialogkontext erforderlich. Information über den Ablauf eines Informationsdialogs liefert dazu das Dialogmodell.

Understanding spoken language as it is intended in the Speech Understanding and Dialog System EVAR (see [Nie88]) requires that chains of words hypothesized by the word recognition module are interpreted within a special domain of application. In EVAR this kind of analysis is divided into two steps:

- Semantic analysis:
It is independent of a special application. It consists of the verification of syntactic constituent hypotheses using selectional restrictions, the combination of syntactic constituents to larger entities using syntactic and semantic knowledge, and the structural interpretation of these word chains.
- Pragmatic analysis:
Such semantic structures are interpreted in

*jetzt: Triumph-Adler AG, Olivetti Office Research, Fürther Str. 212, D-8500 Nürnberg 80

the domain of application, in our system the information about intercity trains. In this context the determination of temporal ranges referenced during the dialog is an important point for finding an answer for an information inquiry.

For the determination of temporal ranges and for the resolving of anaphoric references knowledge out of the situational and linguistic context is integrated into the semantic and pragmatic analysis.

An interpretation in the dialog context is needed in order to be able to handle a dialog. A dialog model provides information about the structure of an information dialog.

1 Verstehen gesprochener Sprache

Systeme, die gesprochene Sprache ‚verstehen‘ sollen, müssen zusätzlich zu den Aufgaben, die bei der Interpretation geschriebener Eingabe zu lösen sind, weitere Probleme berücksichtigen. Diese liegen an der völlig anderen kommunikativen Situation bei gesprochenen Dialogen als bei geschriebenen Texten. Bei gesprochenen Dialogen sind die außersprachlichen Kommunikationsmittel (z.B. Mimik, Gestik) häufig noch schwieriger zu erfassen und zu interpretieren als die sprachlichen Äußerungen selbst. Daher muß man sich im allgemeinen auf die Sprache konzentrieren (z.B. über den Kanal „Telefon“). Das heißt aber auch, daß alle Informationen nur sprachlich dargestellt werden können. Daher werden die verwendeten grammatikalischen Konstruktionen vielfältiger sein müssen, das heißt die Syntax einer natürlichen Sprache kann (kaum) eingeschränkt werden. Dazu kommt das Problem, daß gesprochene Dialoge meist anderen grammatikalischen Regeln genügen als die geschriebene Sprache (siehe z.B. [Hit86]): Es werden weit weniger komplexe Satzkonstruktionen und Nominalisierungen verwendet, dafür ist aber ein Großteil der Äußerungen syntaktisch unvollständig. Meist bestehen Dialoge nach einer einführenden Äußerung (fast) ausschließlich aus Satzfragmenten, die wie die ebenfalls häufigen anaphorischen Konstruktionen (z.B. Pronomina) nur im Dialogkontext interpretiert werden können. Daraus ergibt sich als Anforderung an ein sprachverstehendes System, daß der sprachliche Kontext und damit

Erwartungen an die Struktur und Bedeutung einer Benutzeräußerung die Analyse steuern oder mindestens unterstützen muß.

Die Charakteristika gesprochener Sprache bedingen auch technische Probleme bei der maschinellen Analyse und Interpretation. Eine besondere Schwierigkeit entsteht bei kontinuierlich gesprochener Sprache dadurch, daß keine Wortgrenzen erkennbar sind. Jedes Wort des Lexikons kann prinzipiell an jeder beliebigen Stelle des Sprachsignals hypothetisiert werden. Die Menge der von einem Spracherkennungssystem erzeugten Worthypothesen enthält im allgemeinen ein Vielfaches der Anzahl der wirklich gesprochenen Wörter, wobei nicht notwendigerweise auch alle diese in der Menge enthalten sind (beim gegenwärtigen Stand der Worterkennung in EVAR werden für einen Satz von 5-10 Wörtern etwa 260 Worthypothesen benötigt, so daß 80% der gesprochenen Wörter in der Hypothesenmenge enthalten sind). Diese Hypothesen können wieder zu sehr vielen Wortketten zusammengefaßt werden, die das ganze Sprachsignal oder einen Teil davon überdecken. Da aber aus den oben erwähnten prinzipiellen Gründen nicht alle wirklich gesprochenen, aber zum Teil „schlecht“ artikulierten Wörter und alle intendierten, aber überhaupt nicht ausgesprochenen Wörter im Sprachsignal gefunden werden können, ist es nicht möglich, alle Ketten zu suchen und diese nach den Methoden der Computerlinguistik für geschriebenen Text zu analysieren:

- In der Computerlinguistik wird vorausgesetzt, daß die zu analysierenden Sätze korrekt nach den vorgegebenen Regeln sind, und die Aufgabe in der Interpretation liegt (siehe z.B. [Gri86]).
- Wegen der kombinatorischen Explosion würde diese zweistufige Art der Analyse nicht durchführbar sein, so daß es nötig wird, Erwartungen aufgrund von schon analysierten Teilen des Sprachsignals und bedingt durch den situationellen und sprachlichen Kontext möglichst zeitig in die Analyse zu integrieren.

Erst durch eine Bedeutungsinterpretation werden spracherkennende Systeme zu sprachverstehenden Systemen. Zum ‚Verstehen‘ einer Äußerung ist es notwendig, die Bedeutung in einer geeigneten Form zu repräsentieren. Diese Analyse erfolgt in den meisten Systemen erst nach der

eigentlichen Erkennungsstufe, indem Hypothesen über syntaktische Strukturen, die zum Teil semantisch verifiziert sind, Bedeutungsinterpretationen zugewiesen werden. Nur in ganz wenigen Systemen wird die Bedeutungsrepräsentation auch zur Steuerung der Analyse eingesetzt (z.B. [Shi86], [Hay86]).

Für ein Dialogsystem genügt die Interpretation einzelner Äußerungen nicht, es muß auch Dialogwissen berücksichtigt werden. Sprache ist auch über die Satz- bzw. Äußerungsebene hinaus strukturiert. So lassen sich praktisch alle Dialoge in eine Einleitung, Haupt- oder Mittelteil und Abschluß einteilen. Für spezielle Dialogtypen, wie z.B. Informationsdialog, ist noch eine feinere Strukturierung erkennbar. Ein Dialogmodell enthält deshalb Information über die Struktur von Informationsdialogen. Allerdings läßt sich bei gesprochener Sprache selten eine scharfe Grenze zwischen akzeptablen und nicht akzeptablen Dialogen ziehen.

2 Das Spracherkennungs- und Dialogsystem EVAR

Im Spracherkennungs- und Dialogsystem EVAR (Erkennen - Verstehen - Antworten - Rückfragen) sollen einem Benutzer telefonisch Auskünfte über den Fahrplan, Reservierungsmöglichkeiten, Fahrpreise und ähnliche Informationen über Intercity-Züge in Deutschland gegeben werden, z.B.

- U1: Ich möchte morgen nach Hamburg fahren.
- S1: Sie können den Zug um 10.15 h nehmen.
- U2: Hat der in Hannover Aufenthalt?
- S2: Ja, zum Umsteigen in den Zug nach Bremen.
- U3: Fährt abends auch noch einer?

Die Anwendung bedingt Sprecherunabhängigkeit. Der verfügbare Wortschatz sollte mindestens 1000 Wörter enthalten. Die Syntax des gesprochenen Deutschen läßt sich kaum einschränken, da ein menschlicher Benutzer dies kaum akzeptieren und einhalten wird (z.B. nur direkte Fragen zu verwenden, aber keine impliziten wie „ich möchte morgen nach Hamburg fahren“ als Frage nach einer Abfahrtszeit).

Bild 1 zeigt die Systemstruktur: Die gesamte Analyse muß in Abhängigkeit von der aktuellen Dialogsituation erfolgen. Damit können Vorerwartungen, die vom Dialogmodul erzeugt werden, für die strukturelle Interpretation ausgenutzt werden (Syntax-, Semantik- und Pragmatik-Modul). Die Modulen, die „nahe“ am Sprachsignal arbeiten (Akustik/Phonetik-, Worterkennungs- und Prosodie-Modul), können sowohl zu Anfang der Analyse aktiviert werden, als auch auf Anforderung durch andere Modulen: z.B. kann für eine Wortkette, die aus einzelnen Worthypothesen gebildet wurde, eine Betonungsstruktur ermittelt werden (Prosodie-Modul), oder sie kann erneut mit dem Sprachsignal verglichen werden (Worterkennungs-Modul).

Im Gegensatz zu den meisten anderen sprachverstehenden Systemen wird in EVAR anwendungsunabhängige (Semantik-Modul) und anwendungsabhängige Semantik (Pragmatik-Modul) unterschieden: Die Grundlage der Analyse im Semantik-Modul sind Valenz- und Tiefenkasustheorie (siehe [Tes66], [Fil71]). Dabei werden syntaktische Strukturen (mononukleare Konstituentenhypothesen wie ‚mit dem nächsten Zug‘ oder ‚nach Hamburg‘) semantisch verifiziert, aber auch unter Verwendung syntaktischen und semantischen Wissens komplexere Strukturen aus syntaktischen Konstituentenhypothesen zusammengesetzt und strukturell interpretiert. Diese semantischen Strukturen werden während der pragmatischen Analyse im Anwendungskontext „Intercity-Auskunft“ interpretiert (z.B. ‚ich möchte morgen nach Hamburg fahren‘ als Frage nach der Abfahrtszeit eines Zuges mit dem Ziel-Bahnhof Hamburg). Dabei muß für diese Anwendung die Interpretation von Zeitangaben besonders berücksichtigt werden. Die Bestimmung eines Zeitbereichs ist von zwei Faktoren abhängig:

- Nach welchen Regeln können mehrere Zeitangaben zu neuen Zeitangaben verkettet werden? Z.B.: ‚nächstes Jahr | zu Ostern‘
- Von der aktuellen Situation.

3 Wissensrepräsentation in EVAR

Ein menschlicher Kommunikationsteilnehmer verfügt neben Wissen über die Sprache und wie

sie zu verwenden ist auch über außersprachliches Wissen, z.B. bestimmte Tatsachen oder gesellschaftliche Konventionen. Dieses Wissen ist in mehrere Komponenten unterteilt. Das allgemeine Weltwissen („das Gras ist grün“) und das durch den speziellen Anwendungskontext bedingte Wissen sind während eines Dialogs statisch. In EVAR gehören zu diesem anwendungsbedingten Bereich z.B. einige grundlegende Informationen über den Zugverkehr. Auch die Annahme, daß bestimmte Informationen über Züge und Zugverbindungen erfragt werden sollen und die Konventionen, wie dies zu tun ist, bleiben während des Dialogs unverändert. Daneben gibt es dynamisches Wissen, das während eines Dialogs ständig aktualisiert wird. Für einen Informationsdialog betrifft das insbesondere die Informationen, die den Informationszustand des Kommunikationspartners ändern sollen.

Für ein sprachverstehendes System kommt noch eine weitere, mehr ‚technische‘ Komponente des Gedächtnisses hinzu: das ‚Analyse-Gedächtnis‘, das heißt alle Informationen und Hypothesen, die während der Analyse einer Äußerung generiert werden, müssen während der Analyse dieser Äußerung verfügbar bleiben.

Das bedeutet aber, man muß das Wissen eines sprachverstehenden Systems danach unterscheiden können, wie ‚langlebig‘ es ist. In EVAR sollen drei Ebenen unterschieden werden:

- *Langzeitgedächtnis*: enthält das während eines Dialogs unveränderbare Wissen, das ‚Modell‘ des Systems. Dazu gehört das gesamte statische Wissen über Semantik, Pragmatik, Dialog und die Vorannahmen über den Benutzer.
- *Dialoggedächtnis*: enthält das während eines Dialogs aktualisierte Wissen. Dazu gehören situationsabhängige Komponenten, die bei Beginn des Dialogs initialisiert werden (z.B. die Sprechzeit) und die Änderungen, die durch die Kommunikationsbeiträge bewirkt wurden. Dieses Gedächtnis ist gültig während eines Dialogs.
- *Analysegedächtnis*: enthält alle während der Analyse einer Äußerung generierten Hypothesen. Es ist nur während dieser Analyse gültig und für die folgenden Dialogschritte nicht mehr verfügbar.

Das statische und dynamische Wissen wird in EVAR durch ein semantisches Netz re-

präsentiert. Der verwendete Netzwerkformalismus geht auf die Repräsentationssprache KL-ONE zurück (siehe [Bri84], [Bra77]). Personen, Objekte, Sachverhalte oder Ereignisse werden durch **Konzepte** („Modell“) oder **Instanzen** repräsentiert (Individuen einer durch ein Konzept definierten Klasse). Jedes Konzept kann eine Spezialisierung eines allgemeingültigeren Konzepts sein und ‚erbt‘ damit alle dessen Eigenschaften. Für Instanzen eines spezielleren Konzepts gilt dann, daß sie auch Instanzen aller Generalisierungen des entsprechenden Konzepts sind (z.B. wenn ‚Chicco‘ ein Hund ist, also eine Instanz des Konzepts **HUND**, dann wäre dieses Objekt auch eine Instanz der Konzepte **TIER** und **LEBEWESEN**). Jedes Konzept kann durch eine Menge von Attributen näher spezifiziert werden. Diese Attribute können einerseits bestimmte Eigenschaften beschreiben, z.B. Farbe, Form oder Größe, oder aber Beziehungen zu anderen Individuen im semantischen Netz darstellen, z.B. ein Auto **HAT 4 RÄDER**.

Zur Analyse von Äußerungen müssen Instanzen auch über Informationen verfügen, die mit der eigentlichen Repräsentation der Äußerung als linguistische Einheit oder ihrer Interpretation in einer bestimmten ‚Welt‘ nichts zu tun haben (z.B. Bezug zum Sprachsignal, woraus ist die Hypothese entstanden?). Diese Art von Informationen wird in EVAR als ‚Instanzinformation‘ repräsentiert, das heißt jeder Instanz werden zusätzlich zu den Instanzattributen eine Reihe weiterer Merkmale zugeordnet, die der Kontrolle der Analyse dienen.

3.1 Semantik-Netz

Die semantische Analyse in EVAR basiert auf der Valenz- und Kasustheorie: In der Valenztheorie, die in ihrem Ursprung auf [Tes66] zurückgeht, wird das finite Verb als das strukturelle Zentrum eines Satzes betrachtet, das seine „Mitspieler“ regiert. Das Verb besitzt die Fähigkeit, als Nukleus weitere Stellen im Satz zu fordern. Diese Leerstellen werden durch valenzgebundene Glieder, die sogenannten *Aktanten* besetzt. Zusätzlich können dem Verb *freie Angaben* untergeordnet sein, die (fast) unabhängig von der Bedeutung des jeweiligen Verbs hinzugefügt oder weggelassen werden können. In der Kasustheorie (vgl. insbesondere [Fil71]) werden semantische Relationen (*Tiefenkasus*) definiert, die die valenzgebundenen Glieder zum

Verb haben. Fillmore unterscheidet dabei strikt zwischen den semantischen Relationen des Nukleus zu seinen Aktanten und deren oberflächenstrukturellen Repräsentation.

Die Kasusrahmen des Lexikons werden in einem semantischen Netz repräsentiert (*Semantik-Netz*), wobei jeder Kasusrahmen seine obligatorischen und fakultativen Aktanten als eigene Attribute hat (z.B. die Tiefenkasus *OBJEKT* und *MASSANGABE* des Verbframes *kosten.1.2* in Bild 2, also ‚die Fahrkarte‘ bzw. ‚10 Mark‘ in ‚die Fahrkarte kostet 10 Mark‘). Die freien Angaben werden vom übergeordneten Konzept geerbt (z.B. *ZEITANGABE* in Bild 2, also in ‚am Sonntag‘ in ‚am Sonntag kostet die Fahrkarte mehr‘).

3.2 Pragmatik-Netz

Im *Pragmatik-Netz* sind verschiedene Typen von Auskunftswünschen durch *Auskunfts-Konzepte* modelliert. Deren Attribute geben an, welche Informationen benötigt werden, um die Informationsfrage des Benutzers beantworten zu können. Z.B. muß das System für eine Fahrplanauskunft wissen, wohin der Benutzer fahren möchte, es muß also das Attribut *Ankunftsort* realisiert sein. In EVAR werden 7 verschiedene Auskunftstypen unterschieden (siehe dazu Bild 3):

- **Zuginformation:** Dieses Konzept repräsentiert den gesamten Anwendungsbereich. Es ist eine Spezialisierung des Konzepts ‚Thema‘ (für eine Änderung des Anwendungsbereichs müßte also eine weitere Spezialisierung zu ‚Thema‘ parallel zu ‚Zuginformation‘ eingehängt werden). Alle weiteren anwendungsspezifischen Auskunftskonzepte sind damit Spezialisierungen von ‚Zuginformation‘.
- **Objektauskunft:** Hier handelt es sich um allgemeines Wissen über Züge und Bahnhöfe, z.B. „Wieviel Zuschlag muß man für einen Intercity-Zug zahlen?“.
- **Verbindungsauskunft:** Enthält alle Informationen über die Verbindungsmöglichkeiten zwischen zwei Städten (z.B. „Gibt es eine direkte IC-Verbindung von Kassel nach Konstanz?“). Obligatorische Attribute sind hier ‚Source‘, ‚Goal‘ und ‚Instrument‘.

- **Fahrplanauskunft:** Auf einer Fahrtroute (bestimmt durch eine Instanz von ‚Verbindungsauskunft‘) gibt es mehrere Zugverbindungen oder Züge. Um davon eine oder mehrere zu spezifizieren, wird entweder ein Zeitraum für die *Abfahrtszeit* oder die *Ankunftszeit* als notwendig für eine Auskunft gefordert (z.B. „Wann fährt der nächste Zug nach Hamburg?“).
- **Zugausstattung:** Jeder Zug auf einer bestimmten Strecke hat eine bestimmte Ausstattung (z.B. „Hat dieser Zug auch einen Speisewagen?“).
- **Reservierungsauskunft:** Fragen wie „Können Sie mir einen Fensterplatz reservieren?“ sind von einer ganz bestimmten Zugverbindung zu einer festen Zeit abhängig. Daher werden für eine Reservierungsauskunft alle Merkmale einer Fahrplanauskunft benötigt.
- **Fahrpreisauskunft:** Da der Preis für eine Zugreise nur von der Streckenlänge abhängig ist, ist dieses Konzept eine Spezialisierung von ‚Verbindungsauskunft‘.

Um beschreiben zu können, nach welchen Informationen ein Benutzer fragen kann, wurde für jeden Auskunftstyp ein Attribut „*Information*“ eingeführt, das die Informationen spezifiziert, die das System geben kann. Diese werden jeweils bei Bedarf nach einer Datenbankabfrage instantiiert (z.B. die Streckenführung von einer Stadt zu einer anderen wie München - Nürnberg - Würzburg - Hannover, die Länge der Strecke, der Fahrpreis, Ankunfts- oder Abfahrtszeit, siehe Bild 4).

Semantik-Netz und Pragmatik-Netz werden durch spezielle Kanten (‚Konkretisierungs-Kanten‘) verbunden. Dabei können sowohl Konzepte Relationen zueinander haben (z.B. ist der Kasusrahmen *fahren.1.2* eine Konkretisierung des Pragmatik-Konzepts *Fahrplanauskunft*) als auch Attribute. Eine Zeitangabe, d.h. ein Attribut *ZEITANGABE* eines Kasusrahmens, realisiert z.B. das Attribut *Abfahrtszeit* in ‚ich möchte **heute abend** fahren‘ oder das Attribut *Ankunftszeit* in ‚ich möchte **heute abend** dort ankommen‘. Alternativ dazu kann diese Zeitangabe aber z.B. auch den *Ankunftszeitpunkt* in ‚der IC 502 kommt **um 12 Uhr 18 an**‘ als Spezifikation des Attributs *Information* realisieren.

3.3 Dialog-Netz

Das Dialognetz repräsentiert u.a. Wissen über die Struktur von Auskunftsdialogen und deren sprachliche Realisierungsmöglichkeiten. Um ein Modell für einen Auskunftsdialog aufstellen zu können, wurden reale Zugauskunftsdialoge untersucht (siehe [Hit86]).

Kleinste Einheiten des Modells sind Dialogschritt-Typen, wie z.B. Gruß oder Bestätigungsfrage, die in etwa Sprechakten entsprechen. Der Begriff des Sprechaktes oder der Sprechhandlung kommt aus der Sprechhandlungstheorie, welche den Handlungscharakter der Sprache betont. Sprechakte werden ausgeführt, wenn man spricht (z.B. Zunge bewegen, Mund öffnen), und indem man spricht (z.B. bitten, taufen). Als Sprechakte bezeichnet Searle in [Sea71] die Handlungen, die mit der Äußerung des Sprechers verknüpft sind. Als Grundlage für die Dialogschritt-Typen sind besonders die illokutiven Akte von Bedeutung, welche der Befriedigung ganz bestimmter kommunikativer und praktischer Absichten dienen, z.B. eine Frage stellen oder eine Mitteilung machen. Die illokutiven Akte sind die Bausteine jeder Kommunikation. Sie stellen die einzelnen Züge innerhalb der kommunikativen Interaktion dar. Je nach Situation, Absicht und Ziel des Sprechers vollzieht er die entsprechenden illokutiven Akte.

Anhand der Zugauskunftsdialoge wurde ein Modell aufgestellt, in welchem die Strukturierung des Dialoges modelliert wird, d.h. wie die Dialogschritt-Typen aufeinander folgen können.

Eine erste Grobgliederung von Informationsdialogen, die auch in praktisch allen anderen Dialogtypen zu finden ist, ist die Einteilung in Einleitung, Haupt- oder Mittelteil und Abschluß. Die Einleitungsphase dient zur Kontaktaufnahme und Begrüßung der Dialogpartner sowie zur gegenseitigen Identifikation.

Beispiel:

- A: „Reiseauskunft Bahnhof Erlangen, Grüß Gott.“
- B: „Grüß Gott.“

Entsprechend wird in der Abschlußphase der Dialog durch Abschiedsformeln beendet.

Die eigentliche Informationsgewinnung findet im Mittelteil statt. Ziel des Benutzers ist eine bestimmte Information zu bekommen. Der Auskunftgebende versucht diese Information zu

geben, wozu er eventuell seinerseits noch Information vom Benutzer benötigt, die er sich dann durch Nachfragen zu beschaffen versucht.

Der Ablauf sieht im groben wie folgt aus, wobei hier nicht auf alle Schritte eingegangen werden kann.

Der Informationssuchende kann seine Frage mit einer Frageeinleitung einleiten (z.B. „*Ich hätte eine Frage.*“). Danach formuliert er seine Anfrage (z.B. „*Ich möchte morgen nach Hamburg fahren.*“).

Daraufhin kann der Informationsgebende bzw. das System durch gezielte Nach- und/oder Bestätigungsfragen die für eine Datenbankabfrage (z.B. Verbindungsgenerierung) fehlenden Parameter vervollständigen (z.B. „*Wann wollen Sie ungefähr fahren?*“).

Die vom Benutzer gewünschte Information wird dann aus einer Datenbank herausgesucht, so daß die Informationsfrage des Benutzers beantwortet werden kann (z.B. „*Sie können 13 Uhr 27 in Nürnberg abfahren.*“).

Nun kann der Benutzer entweder den Dialog beenden (z.B. „*Danke, auf Wiedersehn.*“), eine neue Anfrage stellen (z.B. „*Wann geht am Sonntag ein Zug zurück?*“) oder zur aktuellen Anfrage mehr Information anfordern (z.B. „*Wann bin ich dann in Hamburg?*“).

4 Semantische und pragmatische Analyse

Da nicht erwartet werden kann, daß die Worthypothesengenerierung alle gesprochenen Wörter auch hypothetisiert, ist es nicht sinnvoll, mit einem Parser, der von links nach rechts arbeitet, syntaktische Satzthesen zu generieren. Aus diesem Grund liefert der Syntax-Modul nur Konstituentenhypothesen wie Nominalgruppen oder Verbalgruppen mit nur einem Nukleus (z.B. über Präpositionalgruppen wie ‚nach Hamburg‘ oder Verbalgruppen wie ‚möchten ... fahren‘). Das ganze Sprachsignal überdeckende Satzthesen oder Konstituentenhypothesen, die selbst wieder aus mehreren Konstituenten bestehen (z.B. ‚der nächste Zug | nach Hamburg‘) werden während der syntaktischen Analyse nicht generiert.

Die Aufgaben der semantischen Analyse liegen in der semantischen Interpretation der durch

die syntaktische Analyse gefundenen Strukturen. Dabei sollen sowohl semantische Mehrdeutigkeiten als auch semantische Inkonsistenzen zwischen den Wörtern, die eine solche Struktur bilden, aufgedeckt werden. Eine weitere Anforderung besteht darin, solche Strukturen, wenn sie nicht den ganzen Bereich des Sprachsignals überdecken, zu größeren Einheiten zusammenzufügen und strukturell zu interpretieren.

Die syntaktischen Konstituentenhypothesen werden semantisch verifiziert und anschließend unter Einbeziehung syntaktisch-semantischen Wissens zu komplexeren Konstituenten und Sätzen zusammengefügt. Durch die Instantiierung im Semantik-Netz (siehe Bild 2) werden die resultierenden Ketten von Worthyypothesen strukturell interpretiert.

Die semantische Verifikation der syntaktischen Konstituentenhypothesen basiert auf mehreren Komponenten (siehe auch [Ehr88b]):

- Semantische Selektionsrestriktionen zwischen Wörtern einer Konstituente werden über die Zugehörigkeit zu bestimmten semantischen Klassen definiert: Den einzelnen Lexemen werden semantische Merkmale zugeordnet, z.B. *BEWEGUNG* für ‚fahren‘. Diese semantischen Klassen sind in einem hierarchischen Klassifikationssystem angeordnet (siehe Bild 5, [Ehr86]). Im Lexikon werden für die Kombination eines Wortes mit einem anderen als Selektionsbeschränkungen Klassen angegeben, mit denen das Wort „semantisch verträglich“ sein muß. Dies gilt vor allem für die semantischen Restriktionen, die eine Präposition oder ein Adjektiv auf das zugehörige Substantiv ausüben (z.B. (*) ‚der schnelle Baum‘ oder (*) ‚während Hamburg‘). Diese Verträglichkeit ist über die Stellung der Klasse im „Klassifikationsbaum“ des Nomens definiert: gibt es eine Verbindung von X zu Y in Richtung von der Wurzel zu den Blättern, dann ist Y mit X verträglich, z.B. ist die Klasse *DING* mit *UNBELEBT* und *KONKRET* verträglich, nicht aber umgekehrt.
- Von Substantiven abhängige Wörter können Merkmale wie ‚possessiv‘ (z.B. ‚mein‘), ‚definit‘ (z.B. ‚dieser‘) oder ‚anaphorisch‘, also auf schon Gesagtes verweisend, tragen. Diese Merkmale sind nicht frei kombinierbar, so daß z.B. ‚ein nächster Zug‘ nicht akzeptiert

wird. Ein weiteres Ausschlußkriterium ist, daß im Deutschen die meisten Substantive im Singular nicht ohne Artikel auftreten dürfen (z.B. (*) ‚ich fahre mit Zug‘).

- Im Kontext ‚Intercity-Auskunft‘ werden Uhrzeiten nur als freie Angaben mit Tiefenkasus *ZEITANGABE* akzeptiert (also z.B. nicht wie in ‚haben Sie „10 Uhr dreißig“ gesagt?‘). Aktanten im Genitiv werden als für gesprochene Sprache unüblich verworfen (z.B. ‚ich bin *deiner* eingedenk‘).

Die Analyse komplexer Konstituenten, die aus mehreren einzelnen semantisch verifizierten, syntaktischen Konstituentenhypothesen aufgebaut sind, ist in EVAR auf drei verschiedene Konstruktionen beschränkt: Temporale Konstituenten wie

- „morgen | früh | gegen neun Uhr“ (siehe dazu den nächsten Abschnitt).
- Genitivische Konstruktionen wie ‚der Speisewagen *des IC's 502*‘ oder ‚der Hut *von diesem Mann*‘.
- Substantiv- und Adjektivergänzungen, z.B. ‚der Zug *nach Hamburg*‘ oder ‚meine Verabredung *um 9.00 Uhr*‘ oder ‚schneller *als der Blitz*‘.

Für die Analyse ganzer Sätze werden die semantisch akzeptablen Konstituentenhypothesen (einfach und komplex) unter Verwendung des im Semantik-Netz repräsentierten Wissens zu ‚Ketten‘ zusammengesetzt, wobei Lücken zwischen den einzelnen Konstituentenhypothesen einer solchen Kette akzeptiert werden. Um zu verhindern, daß anaphorische Konstituenten wie ‚er‘, die als Platzhalter semantisch ‚neutral‘ sind, überall auftreten können, werden vorher für die Konstituentenhypothesen die möglichen Referenzobjekte im Dialogkontext gesucht. Zu den rein anaphorischen Konstituenten gehören vor allem Pronomina wie ‚er‘ oder Adverbien wie ‚dort‘ und sehr viele deskriptive Konstituenten wie ‚das (vorher schon erwähnte) blaue Auto‘. Mit Hilfe des Dialogkontexts können Hypothesen ausgeschlossen werden, wenn sich im Dialogkontext keine Konstituente findet, auf die sie sich beziehen können. Dies bedeutet z.B. für den Dialoganfang, daß keine rein anaphorische Konstituente akzeptiert wird. Zum anderen können die akzeptablen Konstituenten semantisch eingeschränkt werden, indem sie die semantischen Merkmale der möglichen Referenzkonstituente zugewiesen bekommen (z.B. erhält

„der“ in Äußerung U2 alle Merkmale von „den Zug“ in S1 im Beispieldialog aus Abschnitt 2). Damit wird die Kombinationsmöglichkeit mit anderen Hypothesen eingeschränkt.

Um den Aufwand der Instantiierung im Semantik-Netz zu reduzieren, werden als mögliche Kandidaten Ketten von Konstituentenhypothesen gesucht, die sich bezüglich des Sprachsignals nicht überlappen. Dabei werden für die Suche soweit möglich jeweils mehrere „ähnliche“ Kasusrahmen und Teile von Kasusrahmen zusammengefaßt und nur eine Teilmenge der Restriktionstests ausgeführt. Damit bilden die gefundenen Ketten eine Obermenge aller im Semantik-Netz interpretierbaren Ketten aus den gefundenen Konstituentenhypothesen. Nur diese werden dann im Semantik-Netz strukturell interpretiert.

Die durch die semantische Analyse generierten Instanzen der Kasusrahmen sollen im Anwendungsbereich „Intercity-Auskunft“ interpretiert werden, d.h. zu einer Kasusrahmeninstanz muß ein Auskunftskonzept gefunden und instantiiert werden. Welche Konzepte durch eine Kasusrahmeninstanz instantiiert werden können, und die Zuordnungsmöglichkeiten der Attribute des entsprechenden Semantik-Konzepts in das Pragmatik-Netz sind durch die Konkretisierungs-Beziehungen festgelegt. Wegen der Mehrdeutigkeiten der Konkretisierungskanten gibt es im allgemeinen mehrere Möglichkeiten, eine Pragmatik-Instanz aus einer Semantik-Instanz zu generieren. Außerdem kann eine Äußerung aus mehreren Sätzen bestehen (z.B. ‚ich möchte morgen von Nürnberg nach Ulm fahren - und ungefähr um vier Uhr in Ulm sein - wann muß ich dann in Nürnberg wegfahren‘ - aus Dialog 2 aus der Erlanger Stichprobe, siehe auch [Hit86]). In einem solchen Fall muß aus mehreren Semantik-Instanzen eine Pragmatik-Instanz generiert werden. Auch dabei können mehrere alternative Pragmatik-Instanzen entstehen.

5 Interpretation der Benutzeräußerungen relativ zum Dialogmodell

Um das Dialogmodell bei der Analyse von Auskunftsdialogen sinnvoll einsetzen zu können, müssen die Benutzeräußerungen den Dialogschritt-Typen des Modells zugeordnet werden. Ist der vom Benut-

zer realisierte Dialogschritt identifiziert und interpretiert, wird daraus der nächste Schritt des Systems ermittelt.

Dieser Schritt erfordert unter anderem die Verbindung der Ebenen Syntax, Semantik und Pragmatik mit dem Dialogmodell. Es geht darum, aus den Ergebnissen der pragmatischen, semantischen und syntaktischen Analyse den Dialogschritt-Typ zu bestimmen. Dabei genügt es nicht, nur auf die Ebene der Pragmatik zurückzugreifen, da nicht alle Äußerungen des Dialogs pragmatisch analysiert werden. Vor allem die Einleitungs- und Abschlußinitiative tragen keine für den Anwendungsbereich typische Information, da sie praktisch in allen Telefondialogen verwendet werden. Sie können daher nicht pragmatisch analysiert werden. In diesem Fall muß auf die syntaktische und semantische Realisierung zurückgegriffen werden.

Um den Erwartungen des Benutzers entsprechend reagieren zu können, muß die Intention seiner Äußerungen mit berücksichtigt werden. Zu diesem Zweck werden noch folgende Merkmale zur Charakterisierung der Dialogschritt-Typen verwendet:

- **Metakommunikative Merkmale:** Ein metakommunikatives Merkmal liefert Hinweise zur Bestimmung des realisierten Sprechakts. Es werden im wesentlichen illokutive Akte bezeichnet, zum Beispiel danken, fragen, ablehnen, die dann zur Bestimmung der Intention des Benutzers verwendet werden können. Hierbei tritt die Schwierigkeit auf, daß die Zuordnung von Sprechakten zu Äußerungen keineswegs klar definiert ist. Jedoch kann die Verwendung bestimmter Wörter auf die Realisierung eines Sprechaktes hindeuten, wie z.B. performative Verben (das sind Verben, die den illokutiven Sprechakt benennen, z.B. versprechen, verurteilen), oder Grußformeln, wie z.B. „Grüß Gott“, „Hallo“. Für diese Wörter wurden die entsprechenden metakommunikativen Merkmale ins Lexikon eingetragen.
- Bei gesprochener Kommunikation ist prosodische Information ein signifikantes Merkmal, mit dessen Hilfe Satzbedeutungen spezifiziert und Intentionen ausgedrückt werden können. Dies erfolgt über die Sprechmelodie und den Satzakkzent. Bedeutungsunterscheidend wird die Sprechmelodie z.B.

bei der Wiederholung des gerade Gesagten: ist die Äußerung mit Aussageintonation realisiert, wird sie als Bestätigung interpretiert, mit Frageintonation dagegen als Bestätigungsfrage. Zur Interpretation der Benutzeräußerungen relativ zum Dialogmodell interessiert in erster Linie die Markierung des Satzmodus, insbesondere die intonatorische Unterscheidung Frage vs. Nicht-Frage.

- Die Wortstellung im Satz, d.h. die Abfolge der Konstituenten, liefert Informationen über den Satztyp. So kann z.B. ein Fragesatz durch Erststellung des Verbs (z.B. „Muß ich irgendwo umsteigen?“) oder durch die Verwendung spezieller Fragewörter (z.B. „Wo muß ich umsteigen?“) realisiert werden. Die Intention des Sprechers stimmt nicht unbedingt mit dem Satztyp überein (z.B. ist „Es zieht!“ meistens nicht als Aussage gemeint, sondern als Aufforderung etwas dagegen zu unternehmen, eventuell das Fenster zu schließen oder ähnliches). Der Satztyp kann jedoch zumindest einen Hinweis liefern.

Wie können nun die beschriebenen Kriterien angewandt werden, um Äußerungen den entsprechenden Dialogschritt-Typen zuzuordnen? Das Dialogmodell gibt für jede Phase des Dialogs die möglichen Folgeschritte des Benutzers an. Nach der Begrüßung und Identifizierung des Systems kann der Benutzer mit den Schritten ‚Gruß‘, ‚Frageeinleitung‘ und ‚Informationsfrage‘, wie im folgenden Beispiel, fortfahren.

Beispiel:

- A: Reiseauskunft Nürnberg, Schmidt, Grüß Gott.
- B1: Guten Morgen,
- B2: ich hätte eine Frage,
- B3: ich möchte am nächsten Wochenende von Hannover zurück nach Nürnberg fahren, am Sonntag, wann geht da der letzte IC?

Obligatorisch ist die ‚Informationsfrage‘ (B3), die pragmatisch analysiert werden kann. Hier liefert der Benutzer Parameter, die zur Beantwortung der Anfrage benötigt werden. Die Dialogschritt-Typen ‚Gruß‘ (B1), meistens realisiert durch Grußformeln wie „Grüß Gott“ (mit

dem metakommunikativen Merkmal ‚grüssen‘) und ‚Frageeinleitung‘ (B2), in der das Anliegen benannt wird (z.B. „ich habe eine Frage“, „ich bräuchte eine Auskunft“), werden dagegen nicht immer realisiert.

6 Temporale Angaben

Die Bildung von Zeitangaben im Deutschen muß besonderen idiomatischen Regeln genügen, d.h. die sehr allgemeinen semantischen Selektionstests, die für die Analyse der Zeitangaben nur die Kompatibilität mit der semantischen Klasse *ZEITLICH* fordern, sind zu wenig restriktiv. Daher würden auch syntaktische Konstituenten wie ‚um übermorgen‘, ‚in dann‘ oder ‚an Zeit‘ als zulässige Zeitangaben akzeptiert. Solche ‚Fehlbildungen‘ machen bei den untersuchten Experimenten bis zu 90% aus.

Zeitangaben können auch gebildet werden, indem mehrere direkt aufeinanderfolgende Zeitangaben zu neuen Konstituenten ‚verkettet‘ werden (z.B. ‚am Montag | früh | gegen neun Uhr‘).

Für die Interpretation von Zeitangaben wird der Dialogkontext benötigt, um Zeitbereiche und Zeitpunkte, über die gesprochen wird, in der aktuellen Situation genau spezifizieren zu können. Der Zeitangabe ‚übermorgen‘ ist z.B. am 19.2.1989 ein anderer Zeitbereich als am 11.4.1989 zuzuweisen.

Die Analyse temporaler Angaben gliedert sich somit in drei Phasen:

- Die Überprüfung der einzelnen mit der semantischen Klasse *ZEITLICH* kompatiblen Konstituentenhypothesen, ob sie als temporale Angaben für sich oder in Kombination mit anderen zeitlichen Angaben gültige und im temporalen Sinn interpretierbare Konstituenten darstellen. Dabei werden den Konstituenten verschiedene Attribute zugewiesen:
 - Es wird eine Kategorie bestimmt, z.B. Datum *DAT* für ‚am 4. Januar‘ oder Wochentag *TAG* für ‚Mittwoch‘
 - Die Evaluationszeit wird angegeben, d.h. von welcher Bezugszeit aus eine Zeitangabe bestimmt werden soll. Z.B. muß ‚morgen‘ bezüglich der Sprechzeit interpretiert werden, die Zeitangabe ‚dann‘ dagegen bezüglich der zuletzt betrachteten Zeit.

- Es muß festgelegt werden, wie weit der Zeitbereich verschoben werden soll, z.B. für ‚*morgen*‘ bezüglich der Sprechzeit um einen Tag nach ‚rechts‘ auf der Zeitachse.
- Es gibt Zeitangaben, die absolut, ohne Bezug auf eine kontextuell referierte Zeitangabe bestimmt werden müssen (z.B. ‚*am 4.1.1989*‘).
- Die Suche nach Ketten temporaler Zeitangaben und ihre strukturelle Überprüfung mit Hilfe einer ATN-Grammatik (siehe [Die88]).
- Die Interpretation dieser Zeitangaben im Dialogkontext.

Das Kombinieren von einzelnen Konstituenten erfolgt nach ganz bestimmten morphosyntaktischen und semantischen, zum Teil auch pragmatischen Regeln (z.B. kann ein Zeitbereich nach dem Prinzip der schrittweisen Präzisierung von Teilintervallen bestimmt werden wie in ‚*heute → abend → nach neun*‘. Diese Analyse ist durch einen ATN-Parser realisiert, der Ketten von Konstituenten der semantischen Klasse *ZEITLICH* verarbeitet. Bild 6 zeigt einen Ausschnitt dieses ATNs zur Analyse von Zeitangaben, der z.B. ‚*um 9 Uhr | morgen | früh*‘, ‚*am Freitag | möglichst spät*‘ oder ‚*nächste Woche | am Donnerstag | um 12 Uhr*‘ akzeptiert.

Die zu analysierenden Ketten dürfen keine Lücken bezüglich des Sprachsignals besitzen (d.h. die Konstituenten müssen direkt aufeinander folgen, wobei vorher einzustellende Parameter tolerierbare Lücken und Überlappungen vorgeben). Auch für alle Teilketten muß die Zulässigkeit überprüft werden. Experimentelle Untersuchungen für 20 Testsätze ergaben, daß aus den insgesamt 140 zulässigen Zeitangaben 251 Ketten von Zeitangaben bis zur maximalen Länge von drei Hypothesen gebildet werden konnten. Diese ließen sich durch die ATN-Analyse wieder auf 135 reduzieren.

Zur Darstellung von Relationen zwischen Zeitbereichen gibt es verschiedene Möglichkeiten. In Systemen für Textverstehen (z.B. [Eng84], [Gun84]) wird meist der Ansatz verwendet, die einzelnen referierten Zeitbereiche zueinander in Beziehung zu setzen. Dafür werden im einfachsten Fall nur Vorher/Nachher-Ketten gebildet, so daß die Relation zwischen bestimmten Zeitbereichen meist nur über die Transitivität zu ermitteln ist. Ein sehr viel komplexeres System

wird in [All83b] vorgestellt, wo Zeitintervalle als Knoten in einem Netz dargestellt werden, zwischen denen durch Kanten repräsentierte Relationen bestehen, wobei zwischen zwölf verschiedenen Relationen unterschieden wird.

Da es ziemlich aufwendig ist, so ein Netz von Zeitrelationen zu verwalten, wurde in EVAR der Weg gewählt, Zeitbereiche auf einer Zeitachse zu repräsentieren. Diese Vereinfachung kann in EVAR vertreten werden, da bei der Anwendung ‚Zugauskunft‘ der wesentliche Aspekt für die Darstellung von Zeitbereichen nicht darin liegt, Sachverhalte zueinander in zeitliche Beziehung zu setzen wie beim Verstehen von komplexen Texten (‚*story comprehension*‘), wo diese Beziehungen sprachlich vor allem durch die Mittel ‚*Tempus*‘ und ‚*temporale Konjunktion*‘ realisiert werden. Fahrplanauskünfte dagegen sind stets auf einen konkret datierbaren Zeitbereich bezogen, der meist durch explizite Zeitangaben wie Uhrzeitangaben beschrieben wird.

Werden Zeitbereiche nur auf einer Zeitachse repräsentiert, müssen Beziehungen zwischen Zeitbereichen erst ermittelt werden. Daher werden in EVAR die Zeitbereiche zusätzlich durch die zeitlichen Relationen ‚*Vorgänger*‘, ‚*Nachfolger*‘ und ‚*Präzisierung*‘ in Verbindung gebracht. Die Relation beschreibt eine Detaillierungshierarchie der Elemente der Datierung (Jahr - Monat - Tag - Uhrzeit) und spiegelt damit das Prinzip der schrittweisen Verfeinerung von Teilzeitspannen der üblichen sprachlichen Darstellung von Datierung wider.

Bei der Darstellung von Zeitbereichen auf einer Zeitachse treten Probleme mit unpräzisen Zeitangaben wie „kurz vor Mitternacht“ oder „gegen halb drei“ auf. Auch solche Zeitbereiche werden als Intervall mit festen Grenzen dargestellt, die die ‚*Vorstellungen*‘ des Systems von einer solchen unpräzisen Zeitangabe (z.B. „vormittags“ als zwischen 9 und 12 Uhr) widerspiegeln. Da in einem Dialog die Möglichkeit der Korrektur einer solchen Vorstellung gegeben ist, stellt diese Art der Repräsentation keine zu starke Einschränkung dar, insbesondere wenn die ‚*unpräzisen*‘ Intervalle relativ groß gewählt werden, so daß eine Berichtigung durch den Benutzer nur sehr selten zu erfolgen hat.

Im semantischen Netz (Pragmatik-Netz) wird jede Zeitangabe durch genau eine Instanz des Konzepts ‚*Zeitbereich*‘ oder seiner Spezialisierungen repräsentiert. Attribute geben den Detaillierungsgrad (‚*Datum*‘, ‚*Uhrzeit*‘, ‚*Woche*‘, ‚*Tageszeit*‘) und Relationen (‚*Einbettung*‘, ‚*Ver-*

feinerung, *Vorgänger*, *Nachfolger*) zu anderen Zeitangaben an (z.B. zur *Sprechzeit*, die jeweils zu Beginn des Dialogs initialisiert wird).

Bezugsgrößen für die Interpretation von temporalen Angaben sind im Dialogkontext repräsentiert, das heißt die zeitlichen Relationen (*Einbettung*, *Verfeinerung*, *Vorgänger*, *Nachfolger*) von temporalen Instanzen im Analysekontext müssen zu Instanzen im Dialogkontext hergestellt werden, nicht zu den konkurrierenden Hypothesen im Analysekontext selbst. Ob alle während der semantischen Analyse generierten temporalen Angaben auch interpretiert werden können, hängt von der aktuellen Situation ab, z.B. von der Uhrzeit, zu der der Dialog stattfindet (unter der Voraussetzung, daß nur über zeitliche Größen in der Zukunft gesprochen wird, ist z.B. *heute mittag* nur vormittags korrekt interpretierbar). Werden in einer Äußerung mehrere Zeitangaben artikuliert, muß die Interpretation in zwei Phasen erfolgen (z.B. in *Wann fährt denn heute noch ein Zug nach Hamburg?* kann der potentielle Zeitbereich für *wann* erst bestimmt werden, wenn der für *heute noch* schon spezifiziert ist). Solche unbestimmten Zeitangaben (zur Zeit nur Frageadverbien) werden daher immer als interpretierbar gekennzeichnet. Der dazugehörige Zeitbereich wird erst nach der vollständigen pragmatischen Analyse ermittelt, wenn schon feststeht, welche der Instanzen für den folgenden Dialogverlauf als Repräsentant für die Benutzeräußerung gewählt wurde.

An einem Beispiel soll die Interpretation von Zeitangaben verdeutlicht werden: Als *Sprechzeit* wird Mittwoch der 1.3.1989, 23 Uhr 26 angenommen. Als Initialisierung wird eine Instanz vom Konzept *Sprechzeit* generiert:

Konzept: Sprechzeit
Instanz-Nr: 3003
Woche: So,26.2.1989 - Sa,4.3.1989
Datum: Mi,1.3.1989
Uhrzeit: (23 26 h)

Für die Analyse der Äußerung U3 des Beispieldialogs aus Abschnitt 2 sind die Zeitangaben TZ1 („morgen“) und TZ2 („10 Uhr 15“) bekannt und werden wie die *Sprechzeit* im Dialogkontext repräsentiert:

TZ1 (morgen)
Konzept: Zeitbereich
Instanz-Nr: 3007
Datum: Do,2.3.1989
Uhrzeit: ((0 0 h) (24 0 h))

TZ2 (zehn Uhr fuenfzehn)
Konzept: Zeitpunkt
Instanz-Nr: 3010
Vorgänger: \$Sprechzeit/3003
Datum: Do,2.3.1989
Uhrzeit: (10 15 h)

Als Beispiel für mögliche Hypothesen über Zeitangaben werden hier *dann* und die gesprochene Konstituente *abends* gewählt. Die Interpretation im Analysekontext würde folgendermaßen aussehen, wobei die Zeitangaben über die Attribute, hier nur *Einbettung* und *Verfeinerung* nur zu den im Dialogkontext bekannten Instanzen in Beziehung gesetzt werden, nicht aber untereinander (*dann* zu *abends*):

k23 (dann)
Konzept: Zeitbereich
Instanz-Nr: 3013
Einbettung: \$Zeitbereich/3007
Verfeinerung: \$Zeitpunkt/3010
Datum: Do,2.3.1989
Uhrzeit: ((10 15 h) (24 0 h))

k13 (abends)
Konzept: Zeitbereich
Instanz-Nr: 3017
Einbettung: \$Zeitbereich/3007
Datum: Do,2.3.1989
Tageszeit: ((17 0 h) (23 0 h))
Uhrzeit: ((17 0 h) (23 0 h))

Als Zeitbereich für *dann* wird Donnerstag, der 2.3.1989 zwischen 10.15 Uhr und 24.00 Uhr bestimmt in Abhängigkeit von der zuletzt betrachteten Zeit TZ2 (*morgen* - zehn Uhr fünfzehn). Damit ist *dann* in *Zeitbereich/3007* (*morgen*) eingebettet, und *Zeitbereich/3010* (*morgen* - zehn Uhr fünfzehn) stellt eine Verfeinerung von *dann* dar. *abends* wird ebenfalls bezüglich TZ2 interpretiert. Der entsprechende Zeitbereich wird als Donnerstag, der 2.3.1989 zwischen 17 Uhr und 23 Uhr bestimmt.

Gerade im Anwendungsbereich *Zuginformation* ist die Interpretation von Zeiten ein wichtiger Aspekt, der zur Beantwortung der meisten Anfragen benötigt wird. Um z.B. eine Anfrage vom Typ *Fahrplanauskunft* bearbeiten zu können, wird ein Zeitintervall für die Abfahrts- oder Ankunftszeit benötigt. In diesem Intervall muß das Datum, die Uhrzeit und der Wochentag genau genug spezifiziert sein. Erst dann können

die vom Benutzer gewünschten Verbindungen genau genug eingegrenzt und gefunden werden.

7 Ausblick

Eine Testumgebung des Dialogsystems EVAR erlaubt eine vollständige Verarbeitung vom Sprachsignal bis zur Antwortgenerierung, auf die hier nicht weiter eingegangen wurde. In dieses System wurden neben den beschriebenen Verarbeitungsschritten Methoden zur Erkennung gesprochener Äußerungen integriert (siehe [Kum91], [Sch87], [Kuh90], [Noe91] und [Kun91]).

Die Testumgebung umfaßt zur Zeit zwei Dialogschritt-Typen:

- die initiale Informationsfrage des Benutzers und
- die Antwort des Systems.

Wird die Äußerung als eine Anfrage nach einer Intercity-Verbindung interpretiert, so kann eine passende Verbindung (eventuell auch mehrere) gesucht werden, die dann als Antwort des Systems über ein Sprachsynthesegerät ausgegeben wird.

Beim derzeitigen Stand der automatischen Spracherkennung erscheint es noch nicht möglich, ein Dialogsystem für gesprochene Sprache zu entwickeln, das sowohl Realzeitanforderungen als auch die Anforderung nach möglichst wenigen Einschränkungen des Sprachumfangs erfüllt. Insbesondere bleibt die Bedeutungsanalyse zeitintensiv wegen der vielen konkurrierenden Worthypothesen und ihrer Kombinations- und der daraus resultierenden Interpretationsmöglichkeiten. Daher müssen zur Vermeidung der kombinatorischen Explosion verstärkt situations- und kontextabhängige Erwartungen eingesetzt werden.

Ein weiterer wichtiger Punkt vor allem für gesprochene Sprache sind elliptische, d.h. syntaktisch oder semantisch unvollständige Konstruktionen. In gesprochenen Informations-Dialogen ist ein Großteil der Äußerungen nach der dialoginitiiierenden Äußerung nur noch fragmentarisch. Elliptische Konstruktionen können aber nur verwendet werden, wenn der Gesprächspartner die fehlenden Teile auch ergänzen kann. Dies

soll dazu ausgenützt werden, solche Konstruktionen aus dem aktuellen Dialogkontext vorauszusagen. Damit kann die kombinatorische Explosion vermieden werden, die dadurch entstehen würde, wenn an jeder Stelle das Fehlen eines Wortes akzeptiert würde.

Im Hinblick auf die Verwendung prosodischer Information, sind neben den schon beschriebenen weitere Einsatzgebiete vorstellbar. Durch besondere Betonung (Satzakzent) oder Abgrenzung durch Pausen können bestimmte Phrasen einer Äußerung als besonders wichtig hervorgehoben werden. Allerdings ist der Zusammenhang zwischen spezifischen Aspekten der Dialogstruktur und prosodischen Mitteln noch nicht ausreichend erforscht (siehe [Hir87c]). Für EVAR liegen erste Resultate zur Integration prosodischer Analyseergebnisse für die semantische Analyse vor (prosodische Verifikation von Konstituenten). Die darauf aufbauenden prosodischen Interpretationen von linguistischen Strukturen sollen in Zukunft für die pragmatische, vom situationellen Kontext abhängige Analyse (Focus-Bestimmung, d.h. die Bestimmung der ‚wichtigsten‘ Information einer Äußerung) eingesetzt werden.

Danksagung: Die obigen Arbeiten wurden vom BMFT im Rahmen des Verbundvorhabens „Sprachverstehende Systeme“ sowie von der DFG im Rahmen des Schwerpunktprogramms „Modelle und Strukturanalyse bei der Auswertung von Bild- und Sprachsignalen“ gefördert. Nur die Autoren sind für den Inhalt dieses Beitrages verantwortlich.

Literatur

- [All83a] J. Allen: *Maintaining Knowledge about Temporal Intervals*. Communication of the ACM, 26: S. 823 – 843, 1983.
- [All83b] J. Allen, J. Koomen: *Planning Using a Temporal World Model*. In *Proc. of the 8th IJCAI*, Karlsruhe, 1983.
- [Bra77] R. Brachman: *A Structural Paradigm for Representing Knowledge*. BBN Rep. No 3605. Revised version of Ph.D. Thesis. Harvard University, 1977.
- [Bri84] A. Brietzmann: *Semantische und pragmatische Analyse im Erlanger*

- Spracherkennungsprojekt. Dissertation. Arbeitsberichte des IMMD. Band 17 Nr. 5.* Universität Erlangen-Nürnberg, Erlangen, 1984.
- [Die88] M. Diezel: *Analyse von temporalen Angaben.* Studienarbeit, Lehrstuhl für Informatik 5 (Mustererkennung), Universität Erlangen-Nürnberg, Erlangen, 1988.
- [Ehr86] U. Ehrlich: *Ein Lexikon für das natürlich-sprachliche Dialogsystem EVAR Arbeitsberichte des IMMD. Band 19 Nr. 3.* Universität Erlangen-Nürnberg, Erlangen, 1986.
- [Ehr88a] U. Ehrlich: *Semantic and Pragmatic Analysis in the Speech Understanding System EVAR.* In *Computer Processing of Language Data (ROJP). Proc. of the 4th Conference*, S. 43–58, Portoroz (YU), 1988.
- [Ehr88b] U. Ehrlich, H. Niemann: *Using Semantic and Pragmatic Knowledge for the Interpretation of Syntactic Constituents.* In G. S. H. Niemann, M. Lang (Editor): *NATO ASI Series F*, S. 485–490, Springer Verlag, Berlin, 1988.
- [Ehr90] U. Ehrlich: *Bedeutungsanalyse in einem sprachverstehenden System unter Berücksichtigung pragmatischer Faktoren.* Sprache und Information, Max Niemeyer Verlag, Tübingen, 1990.
- [Eng84] K. Engelberg, C. Hauenschild, S. Knöpfler, P. Pause: *ConTra - Ein prozedurales Modell des Textverstehens für die Übersetzung.* Sonderforschungsbereich 99, Universität Konstanz, Konstanz, 1984.
- [Fil71] J. Fillmore: *Plädoyer für Kasus (Original: The Case for Case).* In W. Abraham (Editor): *Kasustheorie*, Athenäum-Verlag, Wiesbaden, 1971.
- [Gri86] R. Grishman: *Computational Linguistics - An Introduction.* Cambridge University Press, Cambridge, 1986.
- [Gun84] S. Günther: *Zur Repräsentation und Verarbeitung zeitlichen Wissens.* In R. Rollinger (Editor): *Probleme des (Text-)Verstehens*, S. 143–155, Max Niemeyer Verlag, Tübingen, 1984.
- [Hay86] P. Hayes, A. Hauptmann, J. Carbonell, M. Tomita: *Parsing Spoken Language: a Semantic Caseframe Approach.* In *Proc. of the 11th COLING*, S. 587–592, Bonn, 1986.
- [Hir87a] J. Hirschberg: *ohne Titel.* In Y. Wilks (Editor): *TINLAP-3*, S. 86–92, New Mexico, 1987.
- [Hir87b] J. Hirschberg, D. Litman: *Now Let's Talk about NOW: Identifying Cue Phrases Intonationally.* In *Proc. of the 25th Annual Meeting of the ACL*, S. 163–171, Stanford, 1987.
- [Hir87c] J. Hirschberg, D. Litman, J. Pierrehumbert: *Intonation and the Intentional Structure of Discourse.* In *Proc. of the 10th IJCAI*, S. 636–639, Milano, 1987.
- [Hit86] L. Hitzenberger, R. Ulbrand, H. Kritzenberger, P. a. Wenzel: *FACID Fachsprachlicher Corpus informationsabfragender Dialoge DICOS.* FG Ling. Informationswissenschaft Universität Regensburg, Regensburg, 1986.
- [Kuh90] T. Kuhn, S. Kunzmann, E. Noeth, S. Rieck, E. Schukat-Talamazzini: *Iterative optimization of the Data Driven Analysis in Continuous Speech.* NATO ASI Speech Recognition and Understanding, Springer-Verlag, Berlin, 1990. In Vorbereitung.
- [Kum91] F. Kummert: *Flexible Steuerung eines sprachverstehenden Systems mit homogener Wissensbasis. Dissertation. Technische Fakultät der Universität Erlangen-Nürnberg.* Erlangen, 1991.
- [Kun91] S. Kunzmann: *Die Worterkennung in einem Dialogsystem für kontinuierlich gesprochene Sprache. Dissertation.* Max Niemeyer Verlag, Tübingen, 1991. In Vorbereitung.
- [Nie88] H. Niemann, A. Brietzmann, U. Ehrlich, S. Posch, P. Regel, G. Sagerer, R. Salzbrunn, E. Schukat-Talamazzini: *A Knowledge Based*

Speech Understanding System. International Journal on Pattern Recognition and Artificial Intelligence, 2: S. 321–350, 1988.

- [Noe91] E. Noeth: *Prosodische Information in der automatischen Spracherkennung Berechnung und Anwendung. Dissertation*. Max Niemeyer Verlag, Tübingen, 1991. In Vorbereitung.
- [Sch87] E. G. Schukat-Talamazzini: *Generierung von Worthypothesen in kontinuierlicher Sprache*. Informatik-Fachberichte, Springer-Verlag, Berlin, 1987.
- [Sea71] J. R. Searle: *Sprechakte*. Suhrkamp, Frankfurt/M, 1971.
- [Shi86] M. Shigenaga, Y. Sekiguchi, T. Yagisawa, K. Kato: *A Speech Recognition System of Continuously Spoken Japanese Sentences and an Application to a Speech Input Device*. In *ICASSP*, S. 1577–1580, Tokyo, 1986.
- [Tes66] L. Tesnière: *Elements de syntaxe structurale (2nd edition)*. Klincksieck, Paris, 1966.