Kontrolle eines homogenen sprachverstehenden Systems

F. Kummert, G. Sagerer, E. Nöth*, H. Niemann*
Arbeitsgruppe Angewandte Informatik, Universität Bielefeld,
*Lehrstuhl für Informatik 5 (Mustererkennung), Universität Erlangen-Nürnberg

1. Einleitung

Rahmenbedingung bildet das Spracherkennungssystem EVAR /Nie88/, das einen natürlichsprachlichen Dialog über den Diskursbereich "Intercity-Auskunft" führen soll. Für die strukturelle Analyse und für die Interpretation einer Äußerung benötigt man linguistisches und anwendungsabhängiges Wissen. Dieses wird in einer homogenen Wissensbasis bereitgestellt, die als semantisches Netz realisiert ist /Sag88/. Neben der Wissensrepräsentation ist für ein erfolgreiches Analysesystem eine effiziente Wissensnutzung erforderlich. Dazu gehört sowohl eine schnelle linguistische Verifikation von Teilinterpretationen als auch die Erzeugung von modell- und datengetriebenen Vorhersagen zur Beschränkung des Analyseprozesses. Dies wird durch anwendungsunabhängige Inferenzregeln erreicht, die das Gerüst für die Steuerung der linguistischen Verarbeitung bilden /Nie90/.

2. Repräsentation des linguistischen Wissens

Für die Darstellung des linguistischen Wissens wird die Wissensrepräsentationssprache ERNEST /Nie90/ verwendet, die auf einem semantischen Netz basiert. Hierbei sind drei Knotentypen (Konzept, Instanz, modifiziertes Konzept) und drei Kantentypen (Spezialisierung, Bestandteil, Konkretisierung) fest definiert. Ein **Konzept** dient der Modellierung eines Begriffs, einer Objektklasse oder einer Klasse von Ereignissen. Demgegenüber stellt eine **Instanz** eine Verbindung zwischen einem Signalausschnitt und dem durch das zugrunde liegende Konzept definierten Begriff her. Ein **modifiziertes Konzept** repräsentiert Wissen, das an eine konkrete Analysesituation adaptiert wurde. Mittels des Kantentyps **Spezialisierung** (spez) werden Konzepte als Ober- und Unterbegriffe miteinander verbunden. Die Beziehung, daß ein Konzept Teil eines anderen ist, wird durch den Kantentyp **Bestandteil** (bst) ausgedrückt. Konzepte, die Begriffe aus unterschiedlichen Abstraktionsebenen darstellen, werden über den Kantentyp **Konkretisierung** (kon) in Beziehung gesetzt.

In Bild 1 ist ein Ausschnitt der linguistischen Wissensbasis dargestellt. Die Syntaxebene enthält Konzepte, die zum einen syntaktische Konstituenten wie Verbalgruppe (SY_VG) oder Präpositionalgruppe (SY_PNG) und zum anderen spezielle Zeitangaben wie Datum (SY_DATUM) modellieren. Die Semantikebene beruht auf Fillmore´s Tiefenkasustheorie. Hierbei wird angenommen, daß ein Verb für eine gewisse Bedeutung Leerstellen eröffnet, denen eine funktionale Rolle (Tiefenkasus) zugeordnet ist. Diese Theorie kann auch auf Nomina übertragen werden, so daß in dieser Ebene Konzepte für die Bedeutung von Verben (z.B. S_VR_FAHREN) und Nomina und für deren funktionale Rollen existieren (z.B. S_GOAL, S_TIME). In der Pragmatikebene wird die Bedeutung der Verb- und Nomenrahmen innerhalb des Anwendungsbereichs interpretiert (z.B. P_VR_FAHREN). Analog werden den Tiefenkasus Begriffe aus dem Diskursbereich, die sogenannten pragmatischen Bestimmungen, zugeordnet (z.B. P_ANKUNFTSORT, P_ABFAHRTSZEIT). Zusätzlich sind in dieser Ebene die Auskunftskonzepte (z.B. P_FAHRPLANAUSKUNFT) enthalten. Sie repräsentieren mögliche Benutzeranfragen und stellen die Analyseziele der linguistischen Verarbeitung dar.

3. Linguistische Analyse

Die Instantiierung eines Auskunftskonzepts, d.h. die Interpretation des Sprachsignals mittels einer Abbildung auf diejenigen Konzepte, die mit dem Auskunftskonzept verbunden sind, ist das Ziel der Analyse. Die zugehörige Instanz enthält dann die Information, die für eine Datenbankabfrage benötigt wird (z.B. Abfahrtsort, Ankunftsort und Abfahrtszeit). Startpunkt der Analyse bilden Worthypothesen zu Wörtern, die im Anwendungsbereich bedeutungsvoll sind, da diese von der Hypothesengenerierung besser detektiert werden. Für jede dieser Hypothesen wird eine Instanz zur zugehörigen Wortart gebildet (z.B. "Zugverbindung" -> I(NOMEN)), die als konkurrierende Teilinterpretationen den Suchbaum initialisieren. Die weitere Analyse wird mittels des A*-Algorithmus gesteuert, der zu jedem Zeitpunkt den aktuell bestbewerteten Suchbaumknoten zur Bearbeitung auswählt. Die Bewertung eines Knotens erfolgt über einen Vektor, dessen Komponenten die strukturelle Zulässigkeit, die Übereinstimmung mit dem Sprachsignal, die Sicherheit für das Vorhandensein einer richtigen Hypothese und die Dringlichkeit der

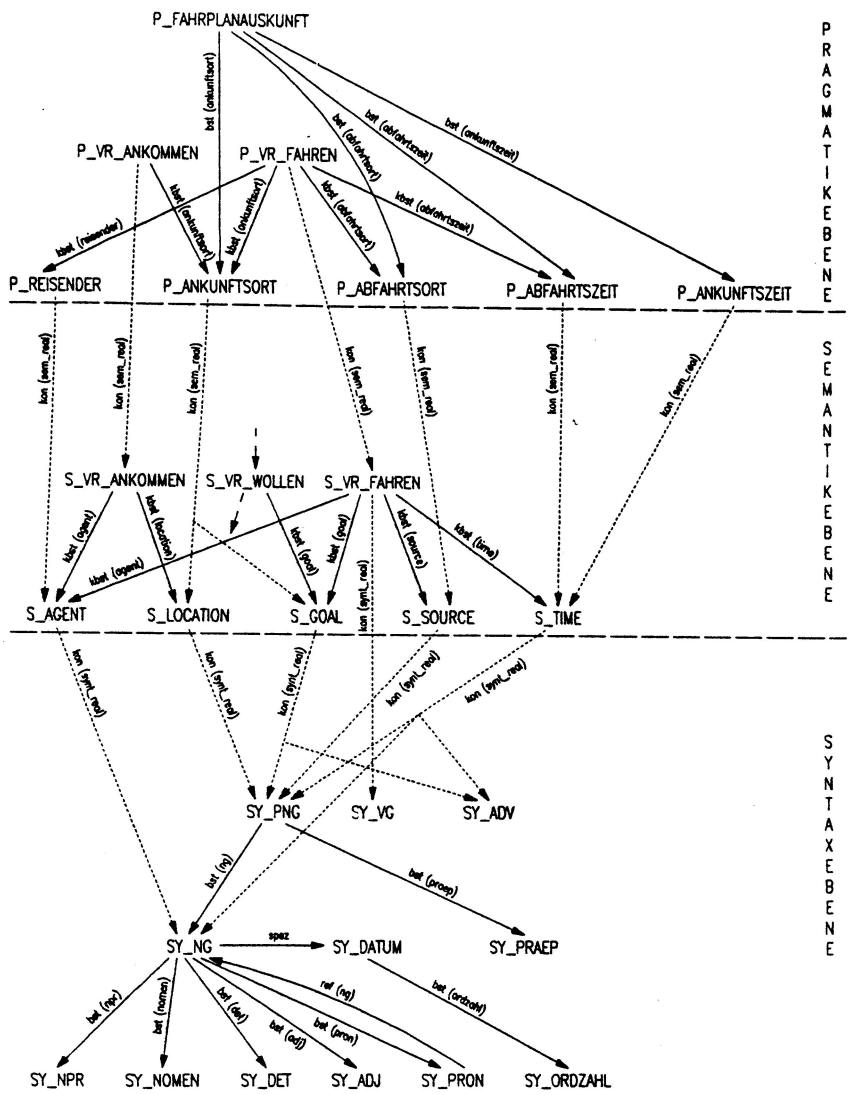

Bild 1: Ausschnitt der linguistischen Wissensbasis

weiteren Verarbeitung der mit dem Knoten verbundenen Teilinterpretation widerspiegelt. Eine ausführliche Darstellung des Bewertungskalküls findet sich in /Sag88/.

Um die restriktiven Vorhersagen aus der Pragmatikebene zu nutzen, werden die Instanzen aus der Initialisierung pragmatischen Bestimmungen zugeordnet. Dazu werden die Kantenverbindungen innerhalb der linguistischen Wissensbasis ausgenutzt. Bild 2 zeigt den Inhalt eines Suchbaumknotens, in dem die initiale Hypothese "Hamburg" der pragmatischen Bestimmung P_ANKUNFTSORT zugeordnet wurde. Durch diese Zuordnung wird das allgemeine linguistische Wissen datenabhängig modifiziert. Dies wird in Bild 2 durch die modifizierten Konzepte Q(SY_PNG), Q(S_GOAL) und Q(P_ANKUNFTSORT) ausgedrückt. So enthält Q(SY_PNG) die Position der Hypothese "Hamburg" im Sprachsignal, wodurch der Endpunkt der

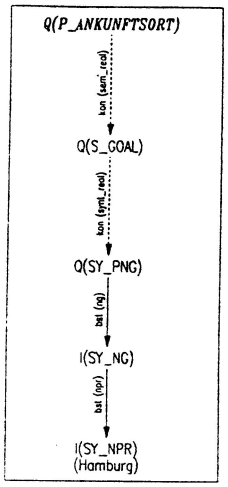

Bild 2: Inhalt eines Suchbaumknotens nach der Zuordnung einer pragmatischen Bestimmung

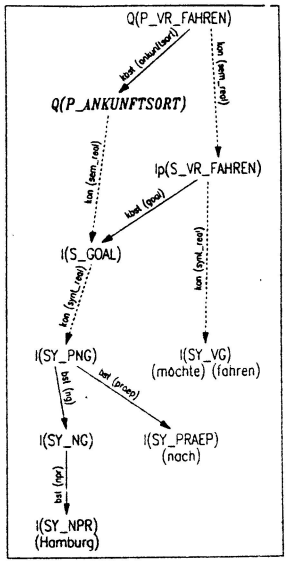

Bild 3: Inhalt eines Suchbaumknotens nach Einbettung in den Kontext eines Verbrahmens

fehlenden Präposition festgelegt ist. Gleichzeitig werden die Erwartungen aus der Pragmatik- und Semantikebene über modellgetriebene Modifikationen an Q(SY_PNG) weitergegeben. Dadurch ist nur noch die Präposition "nach" für Q(SY_PNG) zulässig, wodurch die Anzahl konkurrierender Worthypothesen zusätzlich eingeschränkt wird. Wird für diese Restriktionen keine passende Worthypothese gefunden, so wird die Teilinterpretation verworfen. Im anderen Fall wird versucht die Teilinterpretation im Kontext eines Verb- oder Nomenrahmens zu verifizieren. Über die Kanten im semantischen Netz können die passenden Rahmen einfach gefunden werden. Unter Einbeziehung der bereits analysierten Daten und des Modellwissens wird versucht eine Nominal- bzw. Verbalgruppe, als syntaktische Realisierung eines Nomen- bzw. Verbrahmens zu instantiieren. Bild 3 zeigt den Zustand eines Suchbaumknotens nach diesem Schritt. Um eine zielgerichtete Analyse zu gewährleisten, wird anschließend die Teilinterpretation aus Bild 3 in Bezug zu einem passenden Auskunftskonzept gesetzt. Dadurch werden modellgesteuert zuerst diejenigen pragmatischen Bestimmungen angefordert, die für die gewünschte Anfrage obligatorisch sind. Dies ermöglicht eine schnelle Verifikation der bereits analysierten Daten, da bei einem Nichtzutreffen der Vorhersagen die Hypothese verworfen werden kann. Bild 4 zeigt den Inhalt eines Suchbaumknotens, der die vollständige Interpretation einer Äußerung widerspiegelt. Mit Hilfe der getroffenen Zuordnungen (Abfahrtsort: Nürnberg, Ankunftsort: Hamburg, Abfahrtszeit: nächster Dienstag vormittag) kann nun der Benutzerwunsch nach einer Fahrplanauskunft befriedigt werden.

4. Ergebnisse

Mit der oben geschilderten Vorgehensweise wurden Versuche an 81 gesprochenen Sätzen durchgeführt. Die Worthypothesengenerierung arbeitete dabei sprecherunabhängig auf einem Lexikon mit 1071 Einträgen. Als Hardware stand eine DEC-RISC-Station 5000 mit 24 MByte Arbeitsspeicher zur Verfügung. Für die linguistische Analyse wurden die 100 bestbewerteten Worthypothesen des Akustik-Moduls verwendet, die um die fehlenden Hypothesen zu den gesprochenen Wörtern ergänzt wurden. Dabei wurden die nachgezogenen Hypothesen jedoch nicht in ihrer Bewertung angehoben, sondern es wurde deren tatsächliche akustische Qualität für den A*-Algorithmus verwendet. Unter diesen Rahmenbedingungen wurden 43 Sätze vollständig korrekt analysiert, während bei 14 Sätzen die Äußerung unvollständig interpretiert wurde (z.B. "morgen" anstatt "morgen früh"). Bei 15 Sätzen wurde ein Äußerungsteil falsch

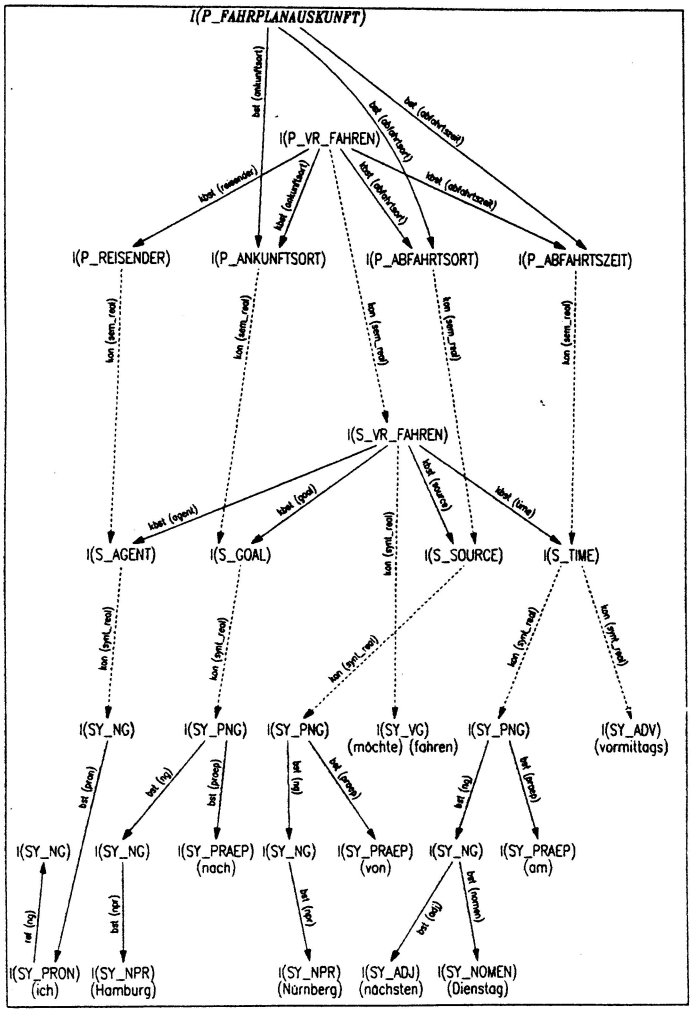

Bild 4: Inhalt eines Suchbaumknotens mit vollständiger Interpretation einer Äußerung

interpretiert und bei den restlichen 9 Sätzen brach die Analyse ohne Ergebnis ab. Im Durchschnitt wurden hierbei 1690 Suchbaumknoten, 1982 Instanzen und 1923 modifizierte Konzepte erzeugt. Die durchschnittliche Rechenzeit betrug 133 CPU-Sekunden.

/Nie85/ H. Niemann, A. Brietzmann, U. Ehrlich, S. Posch, P. Regel, G. Sagerer, E.G. Schukat-Talamazzini: *A Knowledge Based Speech Understanding System.* International Journal of Pattern Recognition and Artificial Intelligence, 2(2): S. 321-350, 1988

/Nie90/ H. Niemann, G. Sagerer, S. Schröder, F. Kummert: *ERNEST: A Semantic Network System for Pattern Understanding.* IEEE Transactions on Pattern Analysis and Machine Intelligence: S. 883-905, 1990

/SAG88/ G. Sagerer, F. Kummert: *Knowledge Based Systems for Speech Understanding.* In H. Niemann, M. Lang, G. Sagerer (Hrsg.): *Recent Advances in Speech Understanding and Dialog Systems*, S. 421-458, NATO ASI Series F, Vol. 46, Springer-Verlag, Berlin, 1988