

Computersysteme sollen spontan-sprachliche Äußerungen “verstehen”. Ein Versprecher, ein Neuansetzer des Sprechers dürfen die Stabilität (Robustheit) des Systemverhaltens nicht gefährden. Aus diesem Grunde wird weiter an der Interaktion zwischen den zwei Komponenten sprachverstehender Systeme gearbeitet. In der “**Akustik**”-Komponente ist die Segmentierung des akustischen Signals in **Wort-Hypothesen**, d.h. in Lautabschnitte mit statistisch-geschätzter lexikalischer Zuweisung (bei Diskurs-begrenztem Wörterbuch) herzustellen. Der Output solcher modular ausgeführter Systemkomponenten wird absichtlich ambig angelegt. Er kann trotz in den letzten Jahren erheblich verbesserter Zuordnungssicherheiten einen gesprochenen Satz im Normalfall ohne linguistisch oder wissensbasiert gestützte Auswahlentscheidungen nicht eindeutig abbilden. Die “**Linguistik**”, das ist hier eine die Hypothesen-Auswahl und –Interpretation steuernde Kontrolleinheit (die “**Kontrolle**”) mit einer vorcompilierten Wissensbasis als Kern, hat Varianten der ‘Akustik-Kopplung’ zur Verfügung. Zwischenergebnisse der Linguistik können z.B. einer mit ihr zeitsynchron getakteten Akustik-Einheit ein Feedback in Form einer Aktualanpassung an die vorangetriebene Interpretation ermöglichen. Beide Einheiten korrespondieren auch insofern, als die Akustik-Strecke ebenfalls entlang **semantischer** (begrifflich gestützter) Interpretationsstufen aufsteigen muß, z.B.

(1) Signalabschnitt -> Laut -> Silben-Gebilde -> lexikalischer Eintrag -> Wortgruppe.

**Links-Rechts-Kontrolle**, in beiden Komponenten, meint den mit der sprachlichen Lautäußerung schritthaltenden Verarbeitungsmodus. Der Echtzeiteffekt des Mithaltens mit dem Sprecher bleibt als technologisches Problem außer Betracht. Anzustreben ist eine Request-Funktion der Linguistik an die Akustik, abgeleitet von bis-dato-Restriktionen aus den ausgewählten Wort-Hypothesen und ihren linguistischen Relationen (Kasus-, Genus-Übereinstimmung u.a.). Sie lautet z.B.: “Finde im folgenden Signalabschnitt eine Hypothese für eine Präpositionalgruppe, welche in der Wissensbasis mit der linguistisch-pragmatischen Definition ‘Ankunftsort’ (ein für das Verstehen – im Anwendungsfall “Zugauskunft” ist das die gesuchte Verbindung – unverzichtbares Äußerungsteil) klassifiziert wird. Akustik dieser Art wird seit Jahren an den Universitäten Erlangen und Bielefeld schrittweise vorbereitet. Der semantische Interpretationszyklus (1) der Akustik läßt sich in der Linguistik im Prozess der Hypothesenkomplexion fortsetzen, der entlang der Hierarchie-Achse der linguistischen Wissensbasis in den separaten Wissensschichten (Syntax, Tiefenkasus-Semantik, Pragmatik, Dialog) abgeprüft und auf-interpretiert wird.

Der **semantische Netzwerk**-Formalismus zur Erzeugung von ERNEST-Wissensbasen [4] legt Knoten und Kanten in einer Frame-Repräsentation mit Slot-Innengliederung so an, daß eine im Kern problem-unabhängige Kontrolle möglich wird. Das der Aufbauform nach problem-**unabhängige** Netz wird problem-**abhängig** gefüllt: Hypothesen als Knoten, ihre Wechselbeziehungen als Kanten. Das bringt langfristig den Vorteil ein sprachverstehendes System mit einem Bilderkennungssystem gleicher Bauart und Kontrolle (wie in der Robotik) in einem einheitlichen Softwareaufbau verschalten zu können. Beide Aufgabenstellungen besitzen deutliche Parallelen. Das Vorläufige der Hypothesen-Auswahl verweist beidmals auf den ausgeprägten **Suchproblem**-Charakter. Der A\*-Algorithmus ist die allgemein akzeptierte Form der Kontrolle eines über konkurrierenden Hypothesen aufgespannten Suchbaums. Die Reduktion der Breitensuche zugunsten einer Tiefensuche in diesem Baum wird durch Berechnung eines Restwegmaßes auf Basis der Sicherheitswerte der Hypothesen erreicht. Der Netzwerk-Aufbau ist in der Entwurfsphase aufwendig und verlangt im Falle sprachverstehender Systeme nach diskurs-bedingter Reduktion.

Da die Erkennung der **Konstituenten** und der **Tiefenstruktur** das Kernstück des Satz-Verstehens bilden, dient die Syntaxschicht des semantischen Netzwerks dazu, diese aufzubauen. Modelliert werden notwendigerweise Konstituenten-**Primitiva**, z.B.

(2) “nach + Frankfurt” – statt: “nach + (der + hessischen + Stadt ) + Frankfurt”.

Die Links-Rechts-Kontrolle benötigt einen Aufsetzmechanismus, der die Überbrückung (**SKIP**-Mechanismus) von Redepartikeln ohne Dialogkonsequenzen (Höflichkeitsfloskeln, Versprecher, undetektierbare Signalbereiche) mit Auffindung der nächsten struktur-relevanten Stelle leistet. Die (A) Netzwerk- und die (B) Regelgrammatik-basierte Kontrolle stehen je in Relation zu alternativen Problemstellungen. (B) generiert zuverlässig die Vollform von (2), falls sie von der Akustik hochgeliefert wird. Die Ausnutzung der pragmatischen Restriktionen (“Abfahrtsort” als eindeutiger Suchbegriff einer Anfrage an die Datenbank des Bundesbahn-Fahrplanes) entzieht sich dem Zugriff des Grammatik-Formalismus. (A) kann in der Netzwerkhierarchie sehr schnell zur Pragmatik aufsteigen und Restriktionen nach unten durchdrücken (potentiell bis zur Akustik). Dafür bleibt das SKIP-Problem zu lösen. (B) will regelgerechtes, störfreies Sprechen, (A) soll spontan-sprachliche Signale mit Störungseffekten analysieren. Das Bielefeld-Erlanger Dialogsystem kann bislang die (B)-Problematik mit einer **struktur-orientierten** Kontrollstrategie weitgehend lösen. Der Funktionsnachweis wurde mit dem Intercity-Zugauskunftssystem in einem BMFT-Großprojekt erbracht [2]. Das Ergebnis wurde nicht-schritthaltend in der Einbahnstraße “erst Akustik, dann Linguistik” erzielt.

Zusammengefaßt soll die linguistische Links-Rechts-Verarbeitung:

1. die zulässige Verarbeitung unsicherer Eingabedaten gewährleisten
2. die Möglichkeit geben, von der strengen Links-Rechts-Folge systematisch abzuweichen, z.B. unsichere Hypothesen zurückzustellen
3. neue Akustik-Linguistik-Kopplungsmöglichkeiten eröffnen, z.B. synchrone Kopplung, automatische Vor-Übersetzung von Subnetzen der Linguistik in entsprechende Wort-Übergangnetze der Akustik [3].

Abb.1. soll die Grundideen des Algorithmus verdeutlichen. Die Zeitachse markiert die Lieferfolge des Akustik-Outputs. Die Hypothesen werden in den Hauptprozeßrichtungen **Zielschätzung** vs. Prädiktion top-down- oder bottom-up zum aktuellen Interpretations-Netz hinzugefügt. Jeder Suchbaumknoten bildet ein mit dem Netz der Wissensbasis konsistentes Subnetz ab. Unter den Knoten des Subnetzes (SY\_, S\_, P\_ markieren Knoten der Syntax-, Semantik-, Pragmatik-Schicht) ist stets ein Zielknoten ausgezeichnet. Die End-Kontrolle im semantischen Netz besteht in der Instanziierung eines der möglichen Zielkonzepte. Dies sind im Fall der Intercity-Zugauskunft die Knoten P\_Fahrplanauskunft, P\_Verbindungsauskunft (im Subnetz von Abb.1. noch nicht erreicht). Dahinter verbergen sich die verschiedenen Typen anforderbarer Auskünfte. Sukzessive Zielfortsetzung als Evaluationsprozeß entlang modellierter Pfade bildet einen Schwerpunkt der Netzwerk-Kontrolle. Auch hier bestehen starke Parallelen zum bildverstehenden Computersystem, das häufig mit unscharfen Erkennungszielen auskommen muß. Der Zielerreichungspfad wird von jedem Subnetz aus durch Abschätzen des nächsten erreichbaren Zwischenziel-Knoten aufgebaut. Der bestbewertete Modell-Knoten wird partiell (oder komplett) instanziiert und dem Subnetz angefügt. Korrespondierend wird ein neuer Suchbaum-Knoten angelegt, d.h. ein Knoten des problem-unabhängig implementierten A\*-Kontrollbaumes. Suchbaumverzweigungen, resultierend aus schlechter bewerteten oder durch Steuerungsparameter unterpriorisierten Hypothesen-Knoten, werden in einer A\*-Warteschleife zum späteren Rückgriff festgehalten. Die Sperrlinie zwischen der Syntax und den höheren Wissensschichten (**Syntax-Sperre**)

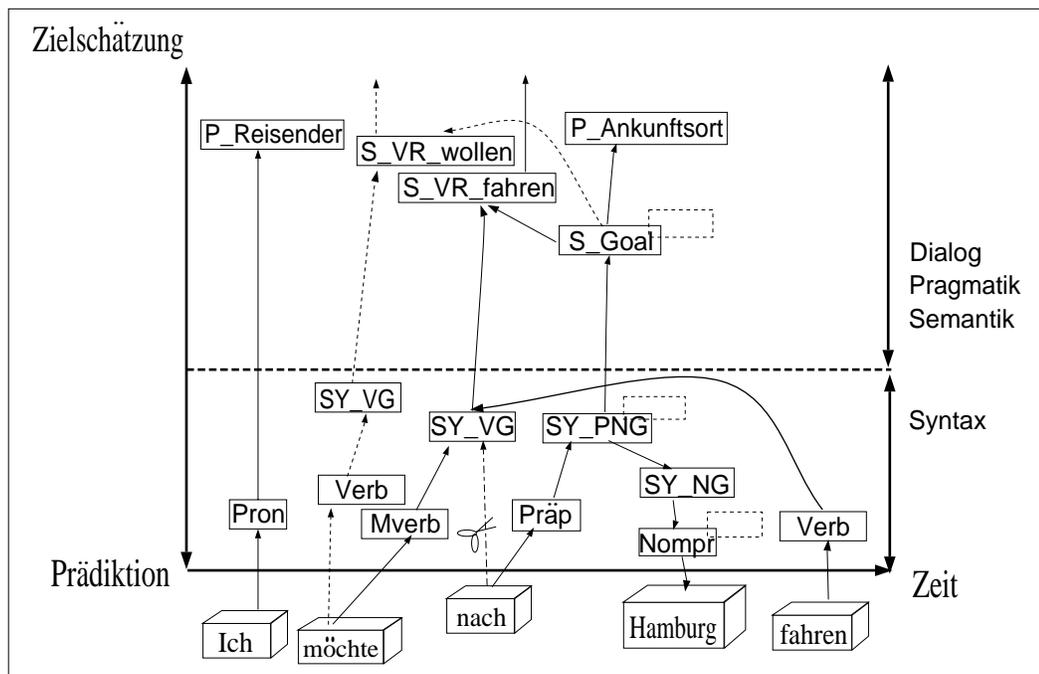


Abb.1.

soll die Gefahr eindämmen, im Falle zu vieler und zu unsicherer Hypothesen aufwendige Pfadbildungen bis in höhere Schichten vornehmen zu müssen. Das Sperr-Kriterium beruht auf der Konstituenten-Erkennung. Geprüft wird die Komplett-Instanziierung aller modellierten Mindest-Bestandteile. In Abb.1. erfolgt für die 'ich'-Hypothese die Zielschätzung bis zur P\_Schicht, weil die Konstituente bereits komplett ist. Die Kontrolle stoppt die Pfad-Fortsetzung, da ein kontextbildender Verbrauchen (gemäß der Tiefenkasustheorie von Fillmore) [1] fehlt. 'Ich möchte nach Hamburg' ist ein diskurs-relevanter Satz. Die Hypothese 'möchte' ist deshalb suchbaumverzweigend (markiert mit schraffiertem Pfeil) als Modalverb und als Vollverb modelliert. Auf dem Vollverb-Pfad ist die Konstituente komplett, weshalb die Sperrlinie bishin zur Erzeugung des Verbrauchs S\_VR\_wollen überwunden wird. Unvollständige Verbrauchen haben i.S. von unaufgelöstem Kontext eine allgemeine Sperrfunktion. Auf dem Modalverb-Pfad ist dagegen bereits die Syntax-Sperre wirksam. Im folgend zu erläuternden Links-Rechts-Automat kippt ein Zustands-Flag um, das zur Ergänzung der unfertigen Konstituente auffordert. Der Pfad der 'nach'-Hypothese stoppt an der Syntax-Sperre. Die zu unterscheidenden (suchbaumverzweigenden) Varianten von Präpositionalgruppen umfassen den in Abb.1. hervorgehobenen zeit-kontinuierliche Fall, der sich hier förmlich als Umkehrpunkt zur modell-getriebenen **Prädiktion** anbietet. In [5] wurde dargestellt, welche daten- und modell-getriebenen Kontrollprozesse bereits im Rahmen der struktur-orientierten Analyse alternierend kombiniert werden können. Die schritthaltende Analyse kommt hier an den Punkt, ihre arteigene **Nachfolger-Restriktion** einzusetzen, nämlich die Anforderung einer restringierten Fortsetzungshypothese an die Akustik. Das Subnetz wird durch einen Modell-Pfad nach unten verlängert, wobei Suchbaumverzweigungen vorkommen. Im Fall der 'nach'-Hypothese wird durch das gesetzte Sperr-Flag eine Suchbaumverzweigung erzwungen. Es ist zu testen, ob die unfertige Verbal-Konstituente ausgebaut werden kann (die 'Schere' verweist auf das spätere Abschneiden des linguistisch unzulässigen Pfades). Abb.2. verdeutlicht die Hauptachse der Kontrolle in Form eines **Pipeline**-Automaten von 5 Prüfstationen. Die **Homogenität** des Netzwerkaufbaus, d.h. der gleiche formale Aufbau von Hypothesen beliebigen Komplexitätsgrades, bewirkt, daß jede Prüfstation mit ihnen arbeiten kann. Insbesondere können die auf-interpretierten Hypothesen in die Pipeline zur Weiterverarbeitung zurückfließen. Die Ausgänge der Stationen (bis auf die 2.) stoßen die modularen

Grundprozesse der struktur-orientierten Kontrolle an. Zur Lösung des SKIP-Problems kann temporär auf Struktur-Orientierung umgeschaltet werden. Eine neue Hypothese wird genau dann angefordert, wenn alle entlang der Pipeline ausführbaren Komplexionen durchgeführt sind. Wenn die Nachfolger-Restriktion unerfüllt bleibt, springt die Kontrolle zur SKIP-Verarbeitung über. Der vom Vorgänger-Wort aufgespannte Pfad von Suchbaum-Knoten wird mit heruntergesetztem Sicherheitsmaß in die A\*-Warteschlange zurückgeführt. Nach erfolgreicher SKIP-Verarbeitung wird versucht, übersprungene Signalabschnitte mittels streng linksseitig aufgebauter Prädiktionspfade zu re-integrieren. Jeder Suchbaumknoten enthält eine mit der Wissensbasis konsistente Teil-Interpretation des Signals. Der Interpretationsfortschritt ist evaluierbar [3]. Das Resultat kann im Dialog verbessert werden.

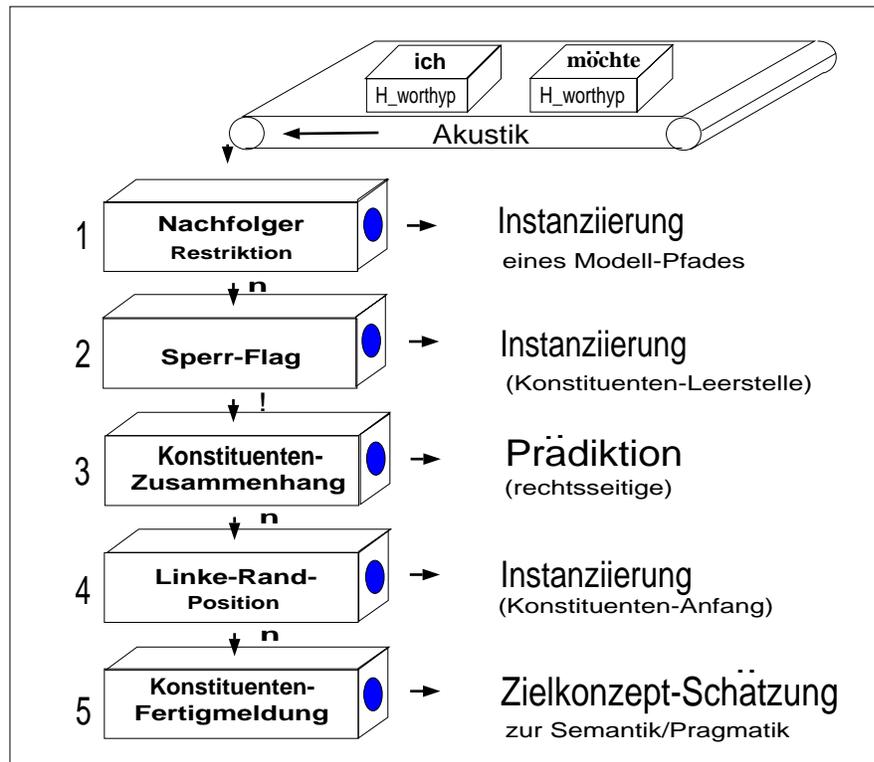


Abb.2.

## Literatur

- [1] J. C. Fillmore. Plädoyer für kasus (original: The case for case). In W. Abraham, editor, *Kasustheorie*. Athenäum-Verlag, Wiesbaden, 1971.
- [2] F. Kummert, G. Fink, and G. Sagerer. Robuste Verarbeitung fehlerhafter Segmentierungsergebnisse. In S. Fuchs and R. Hoffmann, editors, *Mustererkennung 92, 14. DAGM-Symposium Dresden, Informatik aktuell*, pages 269–273. Springer-Verlag, Berlin, 1992.
- [3] G. Sagerer, F. Kummert, G. A. Fink, and B. Seestaedt. Automatische extraktion von sprachmodellen für hidden-markov-modelle aus einem semantischen netzwerk. In R. Hoffmann, editor, *Elektronische Sprachsignalverarbeitung*, pages 151–160, Dresden, 1991.
- [4] S. Schröder, F. Kummert, and G. Sagerer. Ernest – die immd5 netzwerk-umgebung. Unveröffentlichte technische Unterlage, Lehrstuhl für Informatik 5 (Mustererkennung), Universität Erlangen-Nürnberg, 1989.
- [5] B. Seestaedt, F. Kummert, and G. Sagerer. Speech understanding by alternating control: a semantic network based approach. In *6th European Signal Processing Conference*, Brussels, Belgium, 1992.