

FLEXIBLE STEUERUNG EINES SPRACHVERSTEHENDEN SYSTEMS MIT HILFE  
MEHRKOMPONENTIGER BEWERTUNGEN

G. Sagerer, F. Kummert, E.G. Schukat-Talamazzini  
Lehrstuhl für Informatik 5 (Mustererkennung)  
Universität Erlangen-Nürnberg  
Martenstr. 3  
8520 Erlangen, Bundesrepublik Deutschland

### 1. Einleitung

Rahmenbedingung bildet das Spracherkennungssystem EVAR /Nie85/, das mit einem Benutzer einen telefonischen Auskunftsdialog über einen begrenzten Aufgabenbereich (Intercity-Netz) führen soll. Für das automatische Verstehen gesprochener Sprache sind dabei zwei Spielarten der Analyse zu unterscheiden. Zum einen kennen wir eine Hierarchie sprachlicher Einheiten zunehmender zeitlicher Dauer, welche die Oberfläche einer Äußerung repräsentieren. Daneben muß eine Äußerung auch strukturell analysiert werden, d.h. ihr Aufbau aus syntaktischen Konstituenten und die Abbildung vom Gesagten zum Gemeinten muß erschlossen werden.

### 2. Bewertung

Die Bewertungsproblematik ergibt sich aus zwei widerstrebenden Forderungen innerhalb des Konzepts der zeitorientierten Analyse. Einerseits verlangen wir die Integration spezialisierter Moduln in ein Gesamtsystem, die Hypothesen mit unterschiedlichen Methoden und auf Sprachdaten mehrerer Abstraktionsebenen bewerten. Andererseits ist ein homogenes Bewertungsschema zu fordern, das der Kontrolle erlaubt, gute und weniger gute Hypothesen voneinander zu unterscheiden und die Verkettung zeitlicher Teilinterpretationen zu Interpretationen längerer Bereiche zu veranlassen.

#### 2.1 Qualitätsbewertung

Methoden des Mustervergleichs, die ein additives Qualitätsmaß liefern, wurden für die Laut- /Sch87/ und die Spektralebene /Sal86/ realisiert. Die Qualität  $B_Q(H)$  einer Hypothese wird in beiden Fällen ohne eine Längennormierung berechnet. Vermöge der Transformation

$$B(H) = B_Q(H) - m \cdot L(H)$$

wird sie auf ein Maß abgebildet, das sowohl kombinierbar als auch vergleichbar ist. Der Subtrahend stellte ursprünglich die mittlere Qualität  $m(1)$  korrekter Hypothesen der zeitlichen Länge  $l$  dar. Man weist leicht nach, daß sich die Additivität der Qualität auch auf die o.g. Bewertung vererbt, was die approximative Kombinierbarkeit (Verkettung typgleicher Hypothesen) nach sich zieht. Die Vergleichbarkeit, also die Tatsache, daß die Bewertung die Güte der Hypo-

these und nicht etwa ihre Länge (wie z.B. die Qualität) reflektiert, sieht man wie folgt ein: deuten wir  $m$  als die lokale Bewertung eines konstanten maxseg-Profils, dessen Komponentenwerte doch wohl im Mittel von den Qualitäten korrekter Hypothesen geprägt sind, entspricht  $B(H)$  bis auf einen Summanden der Shortfall-Bewertung und spiegelt damit die Chance einer Hypothese wider, Teil einer optimalen Gesamtlösung zu sein.

## 2.2 Sicherheitsbewertung

Die Qualitätsbewertung liefert i.a. noch keinen schlüssigen Hinweis auf die Sicherheit des Zutreffens einer Hypothese. So lassen sich erfahrungsgemäß Hypothesen für längere Wörter zuverlässiger anhand ihrer spektralen oder lautlichen Ähnlichkeit zum Signal als richtig oder falsch einordnen als Hypothesen, die kürzere Wörter betreffen /Sch87/. In einem ersten Ansatz wird die Sicherheit nach dem Längenkriterium bemäßt. Sie berechnet sich unter Normalverteilungsannahme aus Mittelwert und Standardabweichung der Qualitätsbewertung richtiger und falscher Hypothesen des Problembereichs. Der Bhattacharyya-Abstand der Bewertungsdichten für richtige und falsche Hypothesen nimmt die geschlossene Form

$$1/8 \frac{(m_R - m_F)^2}{(s_R^2 + s_F^2)/2} + 1/2 \ln \left( \frac{s_R^2 + s_F^2}{2 * s_R * s_F} \right)$$

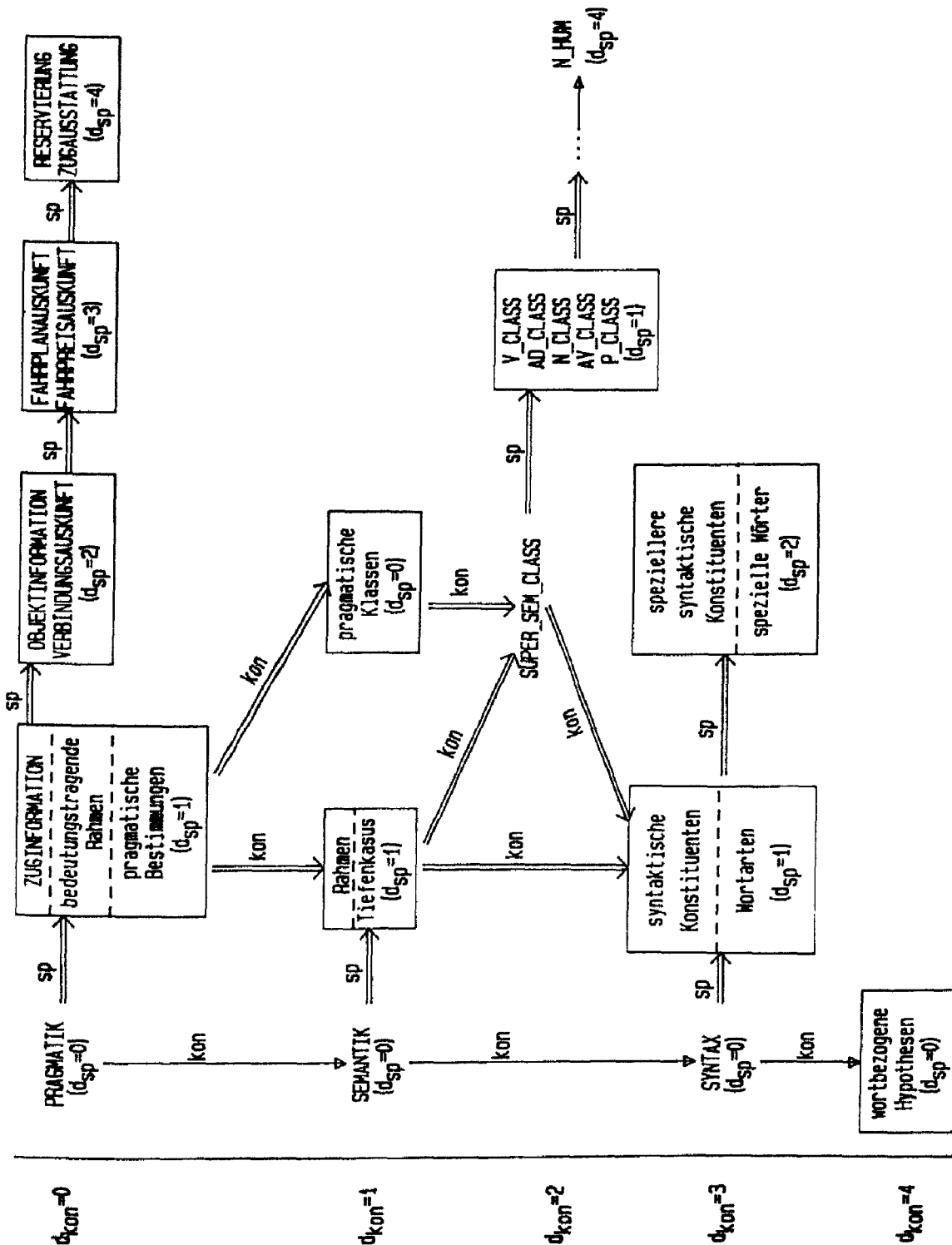
an. Als Abschätzung für die Sicherheit der Vereinigung zweier Hypothesen kann für dieses Maß das Maximum aus der Sicherheit der Einzelhypothesen angenommen werden. Diese ist eindeutig kleiner als die tatsächliche Sicherheit.

## 2.3 Prioritätsbewertung

Mit Hilfe des Prioritätsmaßes soll die Dringlichkeit von Analyseschritten dynamisch beurteilt werden. Die Grundlage dieser Bewertungskomponente kann somit nur das linguistische Modell sein. Dieses ist in Form eines hierarchischen semantischen Netzwerks kodiert, dessen Grundlagen in /Sag85/ beschrieben sind. Da das Ziel der Analyse zunächst eine Interpretation des Sprachsignals in Bezug auf den Aufgabenbereich ist, sind Teilinterpretationen zu bevorzugen, die eine pragmatische Relevanz besitzen. Damit werden bei bottom-up Verarbeitungsschritten Teilinterpretationen bevorzugt, die eine solche Charakteristik besitzen. Für top-down Expansionen ist es wichtig, möglichst schnell eine Vorhersage konkretisieren zu können. Daher sind Vorhersagen zu bevorzugen, die eine "kurze Verbindung" gemäß den Hierarchien des Netzwerks zur Worthypothesenebene besitzen. Des weiteren sind spezielle Interpretationen allgemeineren vorzuziehen. Diese Anforderungen an ein Prioritätsmaß lassen sich mit Hilfe des Netzwerkformalismus beschreiben. Dabei wird nur die Hierarchisierung des Netzwerkes gemäß der Kanten verwendet.

### 3. Linguistisches Modell

Ausgehend von bestehenden Moduln in EVAR /Nie85/ wurde eine homogene hierarchische Wissensbasis als semantisches Netzwerk geschaffen /Kum86/. Es repräsentiert die Syntax und Semantik der deutschen Sprache und die sprachliche Kompetenz des Systems, wie sie durch den Diskursbereich "Interocity-Auskunft" gegeben ist. Das folgende Bild zeigt eine Netzwerkübersicht.



#### 4. Steuerung der strukturellen Analyse

Produktionen zur Erzeugung von Instanzen /Sag87/ legen den Kern der Verarbeitungsstrategie fest. Diese top-down Suche mit bottom-up Instantiierung wird durch eine erwartungs- und datengetriebene Modifikation von Konzepten erweitert, was ein flexibles Ausnutzen von Modelleinschränkungen und Analysewissen erlaubt. Zur Steuerung der Analyse dient ein erweiterter A\*-Algorithmus, der auf dreistelligen Bewertungen arbeitet.

Nach der Schätzung von möglichen Zielkonzepten wird der Suchbaum initialisiert. Jeder Knoten des Suchbaums enthält ein expandiertes Modell eines Zielkonzepts, welches den augenblicklichen Stand der top-down Suche und bottom-up Instantiierung von der Wurzel des Suchbaums bis zu diesem Knoten widerspiegelt. Die Auswahl eines Knotens zur Bearbeitung erfolgt über die Güte (siehe Abschnitt 2.1) und die Sicherheit (siehe Abschnitt 2.2) der Interpretation des Knotens. Dabei ist primär die Güte ausschlaggebend. Nur bei ähnlicher Güte mehrerer Knoten entscheidet die Sicherheitsbewertung über das weitere Vorgehen. Die Sicherheit beeinflusst somit nicht die Güte eines Knotens, sondern er wird in seiner weiteren Bearbeitung nur verzögert bzw. beschleunigt. Die Priorität wird bei der Auswahl von Suchraumknoten nicht betrachtet. Sie entscheidet nach der Auswahl eines Knotens, welche die günstigste Strategie für die weitere Bearbeitung dieses Knotens ist. Sie orientiert sich dabei an dem Ziel, möglichst schnell und einfach die strukturelle Analyse am linguistischen Modell über Hypothesen verifizieren zu lassen. Durch die Prioritätsbewertung (siehe Abschnitt 2.3) werden somit gemäß den Produktionen zur Instantiierung gleichrangige (modifizierte-) Konzepte für die weitere Verarbeitung geordnet. Nachdem ein Knoten gemäß Güte und Sicherheit und ein (modifiziertes-) Konzept in diesem Knoten gemäß Priorität ausgewählt wurden, wird situationsabhängig eine der folgenden Aktionen angestoßen:

- Expansion des Konzepts (= top-down Suche)
- Instantiierung eines nicht primitiven Konzepts ( $g_{kon} \neq 0$  oder  $g_{bst} \neq 0$ )
- Instantiierung eines primitiven Konzepts ( $g_{kon} = 0$  und  $g_{bst} = 0$ )

Bei der Expansion eines Konzepts werden gemäß den Produktionen zur Instantiierung von Konzepten die Konkretisierungen und Bestandteile in die strukturelle Analyse miteinbezogen. Durch die Betrachtung von Modalitäten kann der Suchbaum dabei aufgespalten werden. Über invertierbare Attribute und invertierbare Restriktionen an Kanten kann Wissen, das während der top-down Expansion gesammelt wurde, von den neu betrachteten Konzepten verwendet werden, um Information lokal im Suchbaumknoten einzuschränken. Diese wird in einem modifizierten Konzept abgelegt.

Bei der Instantiierung eines nicht primitiven Konzepts wird die Verzeigerung im Instanzenspeicher mit der Umgebung vorgenommen, es werden Attribute berechnet und die Relationen getestet. Die erzeugte (diskontinuierliche) Wortkette wird

als Hypothese zur Bewertung der Güte an die zeitorientierte Analyse übergeben. Ist damit das Zielkonzept eines Knotens instantiiert, so entscheidet die Kontrolle, ob die Analyse abgebrochen wird (falls ein Auskunftskonzept mit genügend großer Sicherheit instantiiert wurde), ob die Analyse erweitert fortgesetzt wird (falls ein Auskunftskonzept nicht mit genügend großer Sicherheit instantiiert wurde) oder ob ein neues Zielkonzept geschätzt werden muß (falls kein Auskunftskonzept instantiiert wurde).

Für die Instantiiierung primitiver Konzepte wird ebenfalls eine Hypothese an die zeitorientierte Analyse übergeben. Je nach Wissensstand wird eine Einschränkung für die möglichen syntaktischen, semantischen und pragmatischen Klassen sowie für den Zeitbereich mitgeliefert. Das bedeutet, daß die Instantiiierung eine Suche über die Worthypothesen darstellt. Nach der Instantiiierung wird versucht die neu gewonnene Information zur Vorhersage für andere Konzepte des aktuellen Knotens zu verwenden. Dazu wird über invertierbare Attribute und Relationen die intensionale Beschreibung dieser Konzepte so weit wie möglich eingeschränkt.

#### 5. Zusammenfassung

Die angestrebten drei Bewertungsmaße Qualität, Sicherheit und Priorität konnten formal entsprechend den geforderten Eigenschaften festgelegt werden. Darüber hinaus wurde eine homogene hierarchische Wissensbasis geschaffen, welche einem modifizierten  $A^*$ -Algorithmus erlaubt, flexibel auf top-down Erwartungen und bottom-up Hypothesen zu reagieren.

- /Kum86/ F. Kummert, Ansätze für einen flexiblen Kontrollalgorithmus für ein sprachverstehendes System, Diplomarbeit am Lehrstuhl 5 (Mustererkennung), IMM Universität Erlangen-Nürnberg, 1986.
- /Nie85/ H. Niemann, A. Brietzmann, R. Mühlfeld, P. Regel, G. Schukat: The Speech Understanding and Dialog System EVAR. In R. DeMori, C.Y. Suen: New Systems and Architectures for Automatic Speech Recognition and Synthesis, NATO ASI Series F, Springer Verlag, 271-302, 1985.
- /Sag85/ G. Sagerer: Darstellung und Nutzung von Expertenwissen für ein Bildanalysesystem, IFB 104, Springer Verlag, 1985.
- /Sag87/ G. Sagerer, S. Schröder, H. Niemann: An associative network as system shell for knowledge based image understanding. In Proc. 2nd CAIP 87, Wismar 1987.
- /Sal86/ R. Salzbrunn: Mehrstufige Dynamische Programmierung zur Wortverifikation, Diplomarbeit am Lehrstuhl 5 (Mustererkennung), IMM Universität Erlangen-Nürnberg, 1986.
- /Sch87/ E.G. Schukat-Talamazzini: Generierung von Worthypothesen in kontinuierlicher Sprache, IFB 141, Springer Verlag, 1987