

LDV-FORUM

Forum der Gesellschaft für Linguistische Datenverarbeitung GLDV

Band 4 · Nummer 2 Dez. 1986

LDV-FORUM 4(1986)2

Forum der Gesellschaft für Linguistische Datenverarbeitung e.V. (GLDV)

Redaktion

Dr. G. Knorz
TH Darmstadt
Fachbereich Informatik
Karolinenplatz 5
D-6100 Darmstadt
(Tel. (06151) 16-3952 /-2859)
Netz: knorz at DDADVS1.BITNET

Wissenschaftlicher Beirat des LDV-Forum:

I. S. Batori, R. Drewek, Ch. Habel, G. Knorz, J. Krause, H. Lehmann, H.D. Lutz, K.G. Schweisthal

Erscheinungsweise:

halbjährlich, zum 30. Juni und 30. Dezember

Bezugsbedingungen:

Der Bezug des LDV-Forum ist im Regelfall an die Mitgliedschaft in der GLDV gebunden und im Jahresbeitrag eingeschlossen. Einzelexemplare können zum Preis von DM 15,- bei der Redaktion bestellt werden.

Back-up-Exemplare bis einschließlich Ausgabe 2/85 sind zum Preis von DM 5,- (zuzügl. Porto) zu bestellen. Die Ausgabe 1/86 ist vergriffen.

Titelgestaltung:

Fred Zimmermann, FH Darmstadt



EDITORIAL

"... und dann noch das LDV-Forum! Warum machen Sie das eigentlich?!" fragte mich kürzlich ein mir wohlgesonnener Kollege mit leichter Besorgnis in der Stimme. Zum Glück ging die Sitzung gleich weiter und ich war nicht zu tiefer Gewissensforschung gezwungen. Ich hoffe aber, daß man auch *dieser* Ausgabe ansieht, daß es u.a. eine gehörige Portion Spaß sein muß, die das treibende Element meiner LDV-Forum-Aktivitäten ist.

Das vorliegende LDV-Forum ist in Fortsetzung der letzten Ausgabe dem Themenschwerpunkt **Ansätze der sprachorientierten KI-Forschung** zuzuordnen. Für das nächste Heft ist das Thema **Maschinelle Übersetzung** vorgesehen. Ich will an dieser Stelle jedoch ausdrücklich noch einmal darauf hinweisen, daß neben Beiträgen zu dem Schwerpunktthema auch solche zu anderen wissenschaftlichen Fragestellungen ausdrücklich erwünscht sind. In gleicher Weise bin ich für Anregungen zur Wahl zukünftiger Themenkreise jederzeit offen.

Das herausragende einschlägige (Tagungs-) Ereignis des zweiten Halbjahrs war zweifellos die COLING'86 in Bonn. Über die z.T. persönliche und in jedem Fall fundierte Aufarbeitung von Eindrücken von dieser Mammutveranstaltung habe ich mich außerordentlich gefreut.

Eine wesentliche Resource für jede LDV-Forum-Ausgabe sind Beiträge von Lesern zu den Rubriken *Nachrichten*, *Tagungsberichte* und *Rezensionen*. Die beiden erstgenannten Rubriken "leben" zur Zeit allerdings weitgehend von der Zusammenarbeit mit anderen Zeitschriften, insbesondere mit der *Zeitschrift für Semiotik*. Ich wünsche mir sehr, daß das LDV-Forum in dieser Hinsicht eine größere Eigenständigkeit erhalten kann. Die beiden letztgenannten Rubriken verlangen vom Autor eine (fundierte) Stellungnahme, und ich bin verschiedentlich gefragt worden, wie "deutlich" eine solche, unter Umständen negative, Wertung ausfallen dürfe. Sie werden in diesem Heft feststellen, daß ich keine Hemmungen habe, fachlich begründete und sachlich vorgetragene Kritik ohne Abstriche in das LDV-Forum zu übernehmen. "Fachlich begründet" heißt dabei für mich, daß der Autor die Kompetenz besitzen und diese in seiner Besprechung auch erkennbar machen muß, in eigener Verantwortung zu einer Veranstaltung oder einer Publikation Stellung zu beziehen. Daß eine jede solche Stellungnahme selbst wieder Gegenstand einer kontroversen Diskussion sein kann, liegt in der Natur der Sache.

Aus meinen frühen Studententagen ist mir (noch) geläufig, daß *Optimieren* niemals heißen kann "bestmöglichstes Ergebnis bei geringstem Aufwand", sondern daß man **einen** des beiden Parameter, *Qualität* oder *Aufwand*, vorgeben muß, um dann den jeweils **anderen** optimieren zu können. Wenngleich die Alltagssituation mehr Schattierungen zuläßt als es mathematische Optimierungsprozesse tun, wird die oben zitierte Einsicht auf keinen Fall widerlegt. Wenn man, wie ich dies sehe, seit

1985 beim Betrachten von zwei aufeinanderfolgenden Ausgaben bereits auf den ersten Blick deutliche Qualitätsunterschiede ausmachen kann, liegt nahe, daß

- weitere Fortschritte und Verbesserungen in vielen Details zwar möglich und wünschenswert sind, daß diese für viele Leser jedoch in der Zukunft weit weniger ins Auge fallen werden.
- die Investitionen in eine verbesserte Infrastruktur für die Zusammenstellung und Gestaltung des LDV-Forum die Arbeitsbelastung bisher noch nicht in dem wünschenswerten Maße gesenkt hat.

Die Konsequenzen liegen auf der Hand: Weitere Fortschritte werden primär (bei gleichem qualitativen Niveau des Produkts) den Arbeitsaufwand zu senken haben. Verbesserungen wird es nur in solchen Fällen geben können, wo diese aufwandsneutral sind (die Redaktion lernt ständig dazu, und vielleicht helfen Sie mit Anregungen nach ... ?!) oder wo zusätzliches (und längerfristig kalkulierbares) Engagement neu eingebracht wird. Haben Sie bitte keine Bedenken, letztere Feststellung als indirekten Sprechakt aufzufassen. Ein erstes Beispiel für eine Arbeitsteilung mit dem Ziel, den Qualitätsstandard zu sichern und weiter zu heben, ist die Einrichtung eines **Wissenschaftlichen Beirats für das LDV-Forum**, der die Begutachtung eingereichter Fachbeiträge leisten soll (siehe Impressum).

Im Übrigen würde ich mich sehr freuen, wenn wir uns Anfang März auf der GLDV-Jahrestagung sehen könnten. Auf Wiedersehen in Bonn!

Gerhard Knorz

Rubriken:

Die mit Namen gekennzeichneten Beiträge geben ausschließlich die Meinung der Autoren wieder. Einreichungen sind an die Redaktion zu richten.

Fachbeiträge:

Fachbeiträge, die zur Veröffentlichung im LDV-Forum eingereicht sind, werden von mindestens einem Mitglied des wissenschaftlichen Beirats oder (im Ausnahmefall) von einem/einer beauftragten externen Wissenschaftler/in begutachtet. Über das Ergebnis wird der Autor unverzüglich informiert. Manuskripte sind grundsätzlich bei der Redaktion einzureichen. Durch die Anmeldung von beabsichtigten Einreichungen kann die redaktionelle Planung profitieren. Außerdem sollten von der Redaktion Autorenrichtlinien für die Abfassung und Übermittlung von Beiträgen angefordert und beachtet werden. Ein Fachbeitrag hat im Regelfall eine Länge von ca. 6 bis 10 Seiten. Für Beitragsreihen gelten besondere Randbedingungen (vgl. z.B. Seite 25)

Anzeigen:

Media-Information kann bei der Redaktion angefordert werden

Herstellung:

Verlagsdruckerei Hoppenstedt, Havelstraße 9, D-6100 Darmstadt

Auflage:

400 Exemplare

Redaktionsschluß:

Redaktionsschluß für das Juni-Heft: 15. April
 Redaktionsschluß für das Dezember-Heft: 15. Oktober

Gesellschaft für Linguistische Datenverarbeitung (GLDV)

Anschrift:

Prof. Dr. Brigitte Endres-Niggemeyer
 FH Hannover
 Fachbereich BID
 Hannomagstr. 8
 D-3000 Hannover 91
 Tel. (0511) 444344

Bankverbindung:

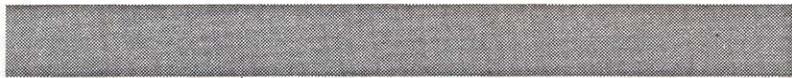
Sparkasse Bonn, BLZ 380 500 00, Kto. Nr. 120 825 633

Mitgliedsbeiträge:

Für Studierende: DM 10,00; für natürliche Personen: DM 50,00; für wissenschaftliche Institute: DM 100,00; für gewerbliche Unternehmen, Behörden und andere juristische Personen: DM 250,00

ANZEIGENFÜHRER

Springer-Verlag: <i>Literatur zur Sprachverarbeitung und -übertragung</i>	8
Klett-Verlag: <i>Stellenanzeige (DV-)Linguist(in)</i>	14
Niemeyer-Verlag: <i>Reihe Sprache und Information</i>	22
Brainware: <i>Expert-System-Shells, Lisp, Prolog</i>	23
Hoppenstedt-Verlag: <i>Stellenanzeige</i>	28
PhiloLogos: <i>Prolog II</i>	35
Bundessprachenamt: <i>Stellenanzeige Wissenschaftliche LDV-Fachkraft</i>	38
epsilon: <i>MPROLOG</i>	47
IKS, Bonn: <i>BYU Concordance</i>	50
GI: <i>BTW-Tagung '87 in Darmstadt</i>	77



WORD EXPERT PARSING UND DISAMBIGUIERUNG

Can inquiring experts be helpful?

Stephan Mehl

EWH Koblenz

Angewandte Informatik, Schwerpunkt Linguistik

Rheinau 3-4

5400 Koblenz

Unter Word Expert Parsing versteht man ein vorwiegend auf lexikalischer Semantik beruhendes Analyseverfahren, bei dem die Auflösung von Mehrdeutigkeiten als zentrale Eigenschaft von Sprachverstehen in den Mittelpunkt gerückt wird. Es wird diskutiert, welche Methoden hierzu verwendet werden und inwieweit sie den an sie gestellten Anspruch erfüllen.

1. Problemstellung

In KI-Modellen zur Simulation von Sprachverstehen steht die semantische Analyse von Sätzen und Texten als eine der zentralen Aufgaben im Mittelpunkt. Ziel ist die Abbildung natürlichsprachlicher Ausdrücke auf Wissen über die Welt (die Sprache, die Situation); diese Abbildung muß wohldefiniert sein, sie wird u.a. bestimmt durch lexikalisches und enzyklopädisches sowie durch episodisches bzw. kontextuelles Wissen (vgl. Habel 1985). Wie dieses Wissen repräsentiert werden kann, ist Gegenstand der KI-Forschung; zu klären, wie es in einer koordinierten syntaktischen, semantischen und pragmatischen Analyse verwendet werden muß, ist Aufgabe der (Computer-)Linguistik.

Die hierzu verwendeten Formalismen gehen als grundlegender Hypothese von der Existenz genereller sprachlicher Regeln aus. Die Möglichkeit, denselben Satz/Text unter Anwendung einander widersprechender Regeln verschieden zu analysieren, wird vom Prinzip her als Ausnahme betrachtet; in diesem Fall werden zunächst alternative Analyseergebnisse (z.B. Strukturbäume) angegeben, sofern keine Anhaltspunkte (z.B. Selektionsrestriktionen) dafür vorliegen, das eine oder andere auszuschließen. In der menschlichen Kommunikation dagegen erfolgt eine fast ununterbrochene Auflösung solcher Mehrdeutigkeiten. Daß dies für die maschinelle Sprachanalyse so problematisch ist, liegt u.a. an den folgenden Punkten:

- der Verschiedenheit des für die Disambiguierung notwendigen Wissens, die seine Einbettung in generalisierte Analyseregeln erschwert
- der erforderlichen Verzahnung von syntaktischer, semantischer und pragmatischer Analyse.

Die Komplexität des Disambiguierungsproblems wird üblicherweise an Sätzen wie Chomskys klassischem *Flying planes can be dangerous* oder an *Ich sah den Mann mit dem Fernrohr auf der Bank* verdeutlicht, die ohne satz-übergreifenden Kontext nicht analysierbar sind. Dabei wird zuwenig betont, daß auch bei einfacheren, offensichtlich eindeutigen Sätzen diese Eindeutigkeit erst hergestellt werden muß. So sind auch für den Satz

Der Mann saß mit einem Fernrohr auf der Bank.

mehrere Interpretationen möglich, aber nur eine wahrscheinlich. Um diese eine ermitteln zu können, müssen allerdings syntaktische Analyse (Zuord-

nung der zweiten Präpositionalphrase), lexikalisch-semantische (*Bank*) und Analyse der Tiefenkasus (*Fernrohr* als Instrument/ Begleiter/...) miteinander gekoppelt werden. Bereits bei geringfügigen Änderungen, z.B.

Ein Wachmann steht mit einem Fernrohr auf der Bank.

Der Mann saß mit gutem Ausblick auf der Bank.

Der Mann saß mit guten Aussichten auf Begnadigung.

kann sich das Ergebnis und damit vielleicht auch der Ablauf der Analyse grundlegend ändern.

Ausgehend von der Bedeutung solcher Details haben Steven Small und Chuck Rieger einen Neuansatz vorgeschlagen, in dem die skizzierte Problematik in den Mittelpunkt gerückt wird: das sog. **Word Expert Parsing** (Small/Rieger 1982, Small 1983). Sprachverstehen wird primär als Prozeß der Disambiguierung betrachtet, für die keine generellen Regeln möglich sind, sondern nur individuelle Lösungsentwürfe. Als Bezugsgröße wird die lexikalische Bedeutung eines Wortes betrachtet, die somit die zentrale Rolle im Analyseprozeß einnimmt. Daraus ergibt sich die Notwendigkeit einer veränderten Parsingstrategie (vgl. Reddig 1984). Im folgenden soll das Modell von Small und Rieger auf seine Funktionsfähigkeit hin untersucht werden: ob es neue Perspektiven bietet oder nur die genannten Probleme anders verpackt.

2. Word Expert Parsing nach Small und Rieger

2.1 Darstellung des Analyseprozesses

Aufgabe eines Word Expert Parsers ist es, darzustellen, wie sich im Verlauf des Verstehensprozesses [1] die Bedeutung von Sätzen und Texten aus der Bedeutung einzelner Wörter im Kontext ergibt. Dazu wird jedem Wort eine Prozedur zugeordnet, der **Wortexperte**, der durch Interaktion mit den Experten anderer Wörter am Aufbau von **Konzepten** mitwirkt, die die Bedeutung eines Textes wiedergeben. Statt einer Generalisierung von Sprachstrukturen sollen die Besonderheiten jedes einzelnen Wortes bei der Analyse betont werden; jede Folge von Wörtern wird als mehr oder weniger idiomatisch betrachtet.

Die Disambiguierung eines Wortes geschieht durch Prüfung der im Kontext identifizierten Wörter und Konzepte sowie durch Zugriff auf allgemeines Weltwissen. Die entsprechenden Abfragen sind im Wortexperten in Form eines Entscheidungsnetzes kodiert; ein Experte wird beim Einlesen des betreffenden Wortes aktiviert und arbeitet die Abfragen solange ab, bis nicht mehr genügend Informationen zur Verfügung stehen. In diesem Fall suspendiert er sich und wartet auf ein Kontrollsignal, das ihm anzeigt, daß andere Wortexperten die benötigte Information erarbeitet haben.

Das erforderliche Wissen wird in zwei Datenbasen bereitgestellt: einem Speicher, der die bisher aufgebauten Konzepte sowie eventuelle Erwartungen an noch fehlende Konzepte enthält, und einem "semantischen Gedächtnis", das allgemeines Weltwissen und Wissen über den Diskursverlauf (z.B. über den Fokus) liefert. Dieses semantische Gedächtnis wird nach dem bisherigen Entwicklungsstand des Modells von dessen Benutzer simuliert; hinsichtlich seiner Struktur wird lediglich angenommen, daß es eine hierarchische Repräsentation generischer Konzepte enthält, die beim Auffinden entsprechender Wörter im Text instantiiert werden. Außer ihrem Namen tragen diese "Konzepte" bei ihrer Instantiierung keinerlei Information, sondern erhalten nur einen leeren Spezifikationsrahmen, der im Laufe der Analyse von den Wortexperten mit immer genaueren Werten gefüllt wird. Das mit einem Konzept verbundene Wissen ist also überwiegend in den Entscheidungsnetzen der Wortexperten untergebracht.

Als Beispiel für die Arbeit eines Word Expert Parsers soll die Analyse des Satzes

The man eating peaches throws out a pit.

dienen (Small/Rieger 1982). Um die Bedeutung dieses Satzes darzustellen, muß der Wortexperte für *throw* ein Konzept "Handlung" (ANYACTION) instantiiieren und unter Angabe der übrigen beteiligten Konzepte bis zu *THROW-AWAY-GARBAGE* verfeinern [2]. Dazu verfügt er über den in Abb. 1 dargestellten Entscheidungsgraphen, der als weitere Bedeutungen von *throw* u.a. das Hinauswerfen einer Person, das Werfen aus einem Fenster, das Anwerfen einer Maschine sowie alle Arten des Werfens vorsieht, die nicht von

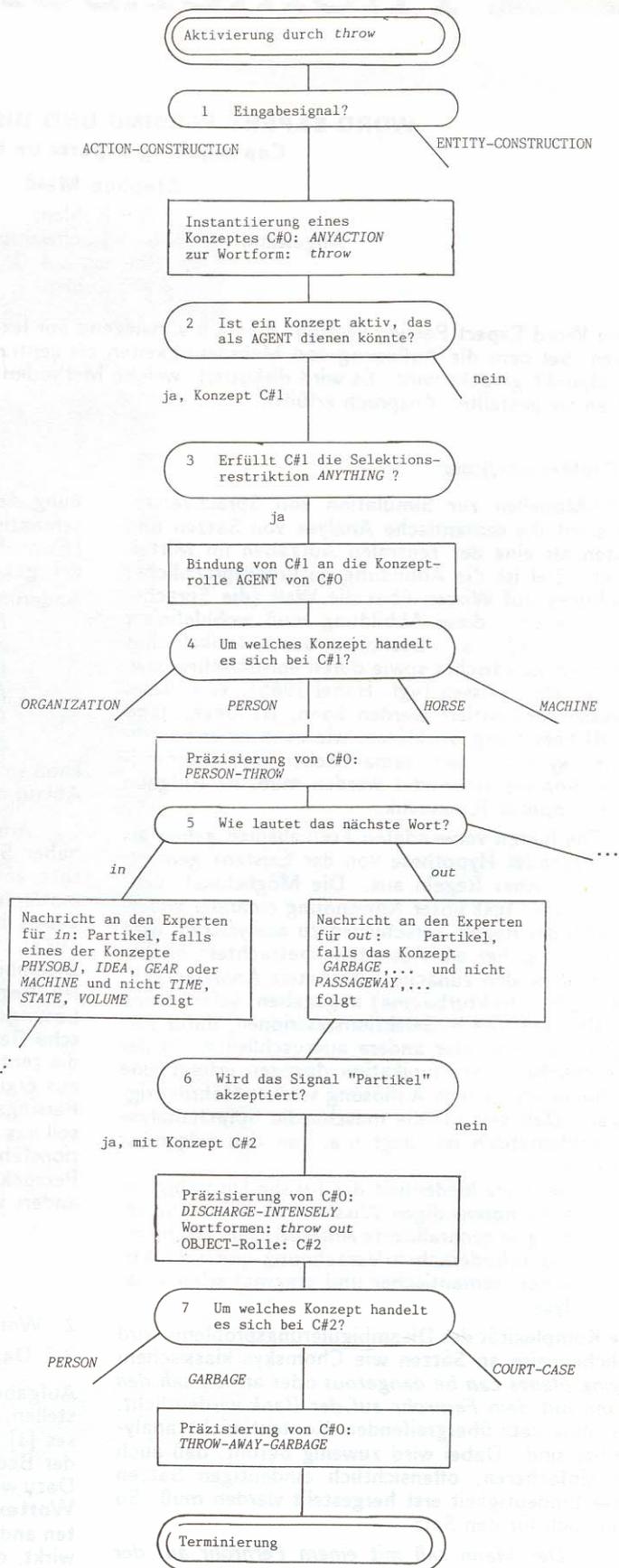


Abb. 1: "Word sense discrimination net" für *throw* (Ausschnitt, erstellt nach Informationen aus Small/Rieger 1982 und Small 1983).

einem Menschen, sondern von einem Tier, einer Maschine usw. vorgenommen werden. Für die Entscheidungen sind entweder Kontakte zu anderen Wortexperten nötig, um Aufschluß über die Satzstruktur zu gewinnen (sog. linguistische Interaktionen, Nr. 1, 5 und 6), Anfragen an die Liste der aktiven Konzepte (Nr. 2) oder an die Datenbasis für Diskurs- und Weltwissen (Diskurs- bzw. logische Interaktionen, Nr. 3, 4 und 7). Da die Verarbeitung der Wörter deterministisch von links nach rechts erfolgt, stehen einem Wortexperten zunächst nur genau diejenigen syntaktischen und semantischen Informationen zur Verfügung, die die Experten der weiter links stehenden Wörter erarbeitet haben; der Zugriff hierauf geschieht entweder über die Liste der aktiven Konzepte oder über Kontrollsignale, die die Art der im Aufbau befindlichen Konzeptstruktur angeben (vgl. Abfrage 1: ACTION-CONSTRUCTION = Verbalphrase, ENTITY-CONSTRUCTION = Nominalphrase). Sobald Informationen aus weiter rechts stehenden Satzteilen benötigt werden (Abfrage 6), muß die Abarbeitung des Diskriminationsnetzes unterbrochen und der nächste Wortexperte angestoßen werden. Die Reaktivierung an der richtigen Stelle wird durch das Setzen eines sog. Demons gesichert.

Zu a): Um zu bestimmen, ob *eating* sich auf das davor oder das dahinter stehende Nomen bezieht (vgl. *the man eating tiger*), müssen die jeweiligen Konzepte nach ihrer Erarbeitung (Stufe 2 und 7) den Rollen AGENT und OBJECT des Konzepts *EAT* zugeordnet werden. Formal geschieht dies durch eine Plausibilitätsanfrage an die Wissensbasis; da diese durch den Dialogpartner des Systems simuliert wird, liegen über die inhaltliche Realisierung dieser Entscheidungsfindung keine Anhaltspunkte vor.

Genauer beschrieben ist dagegen die Lösung des Problems b), da die betrachteten Bedeutungsvarianten von *pit* im Entscheidungsnetz des Wortexperten vorgegeben sind. Ähnlich wie bei der Verfeinerung von *throw* zu *PERSON-THROW* (vgl. Abb. 1) werden hier die semantischen Merkmale eines in der Umgebung befindlichen Konzepts überprüft; zur Auswahl stehen die Merkmale *VOLUME* (→ "Grube"), *MUSICAL-ENTITY* (→ "Orchestergraben") und *SMALL-PHYSOB* [3]. Die Auswahl des die Selektion bewirkenden Konzepts und dessen Zuordnung zu einem Merkmal bleibt jedoch wieder der Wissensbasis überlassen.

Zu c): Anhand von Selektionsrestriktionen, die ihm der Wortexperte für *throw* übermittelt (vgl. Abb. 1, Nr. 5), entscheidet der Experte für *out*, ob es sich um eine Präposition oder ein Partikel handelt. Hierzu muß in der Wissensbasis eine ISA-Zuordnung von *FRUIT-PIT* zu *GARBAGE* aufgefunden werden; damit ist gleichzeitig (Abb. 1, Nr. 7) Problem d), die Disambiguierung von *throw out*, gelöst.

Wortfolge	Konzepte
1 <u>The</u>	
2 <u>The man</u>	C#1: HUMAN-ADULT-MALE
3 (<u>The man</u>)	C#1
4 (<u>The man</u>) <u>eating</u>	C#1
5 (<u>The man</u>) <u>eating</u>	C#1, C#2: EAT
6 (<u>The man</u>) <u>eating</u> <u>peaches</u>	C#1, C#2
7 (<u>The man</u>) <u>eating</u> <u>peaches</u>	C#1, C#2, C#3: PEACH-FRUIT
8 (<u>The man</u>) <u>eating</u> <u>peaches</u> <u>throws</u>	C#1, C#2, C#3
9 (<u>The man</u> (<u>eating</u> <u>peaches</u>)) <u>throws</u>	C#1, C#2, C#3
10 (<u>The man</u> (<u>eating</u> <u>peaches</u>)) <u>throws</u>	C#1, C#3, C#2: PERSON-EAT, EAT-FRUIT
11 (<u>The man</u> (<u>eating</u> <u>peaches</u>)) <u>throws</u>	C#1, C#2, C#3, C#4: PERSON-THROW
12 (<u>The man</u> (<u>eating</u> <u>peaches</u>)) <u>throws</u> <u>out</u>	C#1, C#2, C#3, C#4
13 (<u>The man</u> (<u>eating</u> <u>peaches</u>)) <u>throws</u> <u>out</u> <u>a</u>	C#1, C#2, C#3, C#4
14 (<u>The man</u> (<u>eating</u> <u>peaches</u>)) <u>throws</u> <u>out</u> <u>a</u> <u>pit</u> .	C#1, C#2, C#3, C#4, C#5: FRUIT-PIT
15 (<u>The man</u> (<u>eating</u> <u>peaches</u>)) (<u>throws</u> <u>out</u>) (<u>a</u> <u>pit</u>).	C#1, C#2, C#3, C#4, C#5
16 ((<u>The man</u> (<u>eating</u> <u>peaches</u>)) <u>throws</u> <u>out</u> (<u>a</u> <u>pit</u>)).	C#1, C#2, C#3, C#5, C#4: DISCHARGE-INTENSELY, THROW-AWAY-GARBAGE

Abb. 2: Analyse eines Beispielsatzes

Abb. 2 zeigt den schrittweisen Aufbau von Konzepten bei der Analyse unseres Beispielsatzes. Die Klammerung deutet die Zusammenfassung einzelner Wörter bzw. Konzepte zu dominierenden Konzepten an; aktiviert ist jeweils der zu dem unterstrichenen Wort/Morphem gehörige Experte.

Als Satzbedeutung wird also eine Instanz des Konzepts *THROW-AWAY-GARBAGE* erarbeitet, dem als AGENT eine durch *EAT-FRUIT* (*PEACH-FRUIT*) näher spezifizierte Instanz des Konzepts *HUMAN-ADULT-MALE* und als OBJECT eine Instanz von *FRUIT-EAT* zugeordnet werden. Dazu mußten u.a. die folgenden Mehrdeutigkeiten bearbeitet werden:

- die Struktur von *man eating peaches* (Stufe 9)
- die Bedeutung von *pit* (Stufe 14)
- die Struktur von *throws out a pit* (Stufe 15) und
- die Bedeutung von *throw out* (Stufe 16).

2.2. Kritik

Bereits dieses eine Beispiel zeigt deutlich die Arbeitsweise des Word Expert Parsing in der von Small und Rieger entwickelten Version:

(A) Es handelt sich um ein rein semantisches Parsingverfahren; syntaktische Möglichkeiten der Strukturanalyse und der Disambiguierung werden bewußt außer Acht gelassen. Wäre der genannte Beispielsatz etwa unter Einbeziehung von syntaktischer Valenz und Kongruenz analysiert worden, so hätten die beiden strukturellen Mehrdeutigkeiten a) und c) kein Problem dargestellt: Die Numeruskongruenz hätte nur *man*, nicht aber *peaches* als Ergänzung von *throws* zugelassen, die obligatorische zweite Ergänzung würde durch eine Präpositionalphrase *out a pit* nicht abgedeckt. Dieses Verfahren findet seine Grenze dort, wo semantische Analyse ohne syntaktische Information nicht möglich ist. Angaben über die Wortstellung werden denn auch vom

Word Expert Parser z.B. für die Rollenzuweisung benötigt (vgl. Abb. 1, Nr. 5), für eine korrekte Analyse wird jedoch die Einbeziehung weiteren syntaktischen Wissens erforderlich. Vorschläge hierzu sind im Berliner KIT-Projekt gemacht worden (s.u. Kap. 3).

(B) Die Disambiguierung beruht im wesentlichen auf **semantischen Merkmalen**. Die damit verbundenen Probleme führen dementsprechend auch zu den Hauptschwierigkeiten des Word Expert Parsing:

- Welche primitiven Merkmale bzw. Konzepte werden ausgewählt? Welchen Sinn hat z.B. ein Konzept *THROW-AWAY-GARBAGE* als eigenständige Lesart des Verbs *throw out*?
- Wie lassen sich die Konzepte im semantischen Gedächtnis organisieren, so daß z.B. eine Verbindung zwischen *PEACH*, *FRUIT-PIT* und *GARBAGE* besteht?
- Wie sollen die Selektionsrestriktionen ausgewählt werden? Einerseits erscheint es zu speziell, daß für die Zuordnung des *out* die Opposition *throw out GARBAGE - out (of) PASSAGEWAY* verwendet wird; andererseits reicht selbst *GARBAGE* nicht zur Unterscheidung der Lesarten aus, wenn man Kontexte zuläßt, bei denen etwas "out the garbage" geworfen wird.

An diesem Punkt stellt sich die Frage, ob Small und Riegers "word sense discrimination net" nicht einfach nur eine prozedurale Formulierung von Selektionsrestriktionen darstellt, die üblicherweise deklarativ in einem Lexikon abgespeichert werden. Dem in Abb. 1 wiedergegebenen Beispiel entsprechen dann Einträge wie

throw 1 : AGENT = PERSON, OBJECT = PARTY → ORGANIZE-SOCIAL-EVENT

throw 2 : AGENT = PERSON, OBJECT = MOVABLE, DIRECTION = PASSAGEWAY → MOVE-QUICKLY

throw out 1 : AGENT = PERSON, OBJECT = PERSON → DISMISS

throw out 2 : AGENT = PERSON, OBJECT = GARBAGE → THROW-AWAY-GARBAGE

Im Gegensatz hierzu sind jedoch in einem Wortexperten die Informationen, die zur Analyse eines konkreten Satzes (*throw X out Y*) benötigt werden, an einer Stelle zusammengezogen und nicht auf mehrere Einträge verteilt; durch ihre Opposition an genau dieser einen Stelle lassen sich Selektionsrestriktionen festlegen, die auf die individuelle Opposition zugeschnitten sind. Am stärksten wirkt sich dieser Vorteil bei der Erkennung fester Wendungen bzw. bei der Behandlung von Ausnahmen im allgemeinen aus.

Andererseits kann damit auch nicht die Einführung beliebig spezieller semantischer Merkmale gemeint sein, denn Selektionsrestriktionen müssen ja stets mit entsprechenden Angaben im Kontext kompatibel sein. Die Vermittlung zwischen den jeweils gegebenen bzw. geforderten Konzepten ist Aufgabe der Wissensbasis. Die Funktionsfähigkeit des Wortexperten-Modells hängt also entscheidend von den Möglichkeiten der Wissensrepräsentation für ein bestimmtes Anwendungsgebiet ab.

Im folgenden soll daher das Verhalten eines Wortexpertenparsers in einer Anwendungssituation betrachtet werden. Ausgewählt wurde das Berliner KIT-Projekt, bei dem ein modifizierter Word Expert Parser u.a. für Texte aus den Bereichen "Wetterbericht" und "Hochschulwelt" eingesetzt wird (Eimermacher 1984, 1985, 1986).

3. Anwendungen und Modifikationen des WEP-Konzepts im Projekt KIT

Der im KIT-Projekt eingesetzte Word Expert Parser dient dazu, einzelne Sätze, z.B. aus einem Wetterbericht, in einen Ausdruck der logisch-semantischen Repräsentationssprache SRL zu übersetzen. Das Modell von Small und Rieger wurde hierfür in einem entscheidenden Punkt erweitert, und zwar um die Einbeziehung syntaktischen Wissens. Die wortorientierte Kontrollstruktur bleibt zwar erhalten, die semantischen Kontexterwartungen eines Wortexperten werden jedoch um syntaktische ergänzt. Dadurch läßt sich die Zuordnung der Argumente zu den Rollen eines Konzepts differenzierter steuern: bei solchen Elementen, die relativ zu dem bearbeiteten Wort positionell gebunden sind, wird überwiegend syntaktisches Wissen eingesetzt, bei solchen im freien Bereich semantisches Wissen. Darüber hinaus werden sog. Satzexperten eingeführt, die die Zuordnung von Modifikatoren sowie die Referenzauflösung unterstützen.

Trotz dieser Einschränkung des rein semantischen Parsens kommt der Syntax nicht die Aufgabe einer vollständigen Strukturbeschreibung in allen Details zu, sondern die Sicht des Word Expert Parsing als Modell eines Verstehensprozesses, bei dem nur die jeweils benötigte Auswahl von Informationen erarbeitet wird, bleibt bestehen (vgl. Eimermacher 1983:22). Die Tiefe der syntaktischen Analyse variiert je nach den gegebenen Anforderungen.

Dagegen wird ein weiterer Grundsatz von Small und Rieger, die Vereinigung allen Wissens in einem Wortexperten, teilweise zurückgenommen: Da viele Wörter (einer Wortklasse) gleiche syntaktische Kontexterwartungen stellen, werden diese nicht in jedem einzelnen Wortexperten wiederholt, sondern bilden einen eigenen sog. Wortartexperten. Umgekehrt wird das mehreren Wörtern (Flexionsformen, Ableitungen usw.) gemeinsame semantische Wissen in einem Wortstammexperten zusammengefaßt. Bei der Textanalyse entsteht dann aus Wortstamm- und Wortartexperte der benötigte Wortexperte. - Ansätze für diese Trennung existieren zwar auch im ursprünglichen Modell, das Experten für bestimmte Flexionsmorpheme (-s, -ing) vorsieht. Daraus können - im Deutschen noch stärker als im Englischen - Probleme bei der Bestimmung des zuständigen Wortexperten entstehen; Homographen wie *weiter*, *Lachen* lassen sich nicht mehr dem Wortexperten-Konzept getreu analysieren. Bei dem Berliner Modell gehen darüber hinaus jedoch noch die idiomatischen Besonderheiten einzelner Wörter verloren, die nicht einer ganzen Wortart oder einem Wortstamm gemeinsam sind.

Von besonderer Bedeutung ist schließlich die Realisierung der Wissensbasis, die Verbindung der Konzepte. Hierfür werden im KIT-Projekt Verbände eingesetzt, und zwar einheitlich sowohl im semantischen als auch im syntaktischen Bereich. Als Kontexterwartung läßt sich damit ein je nach semanti-

schen Bereich "Wetterberichte" wurden u.a. die in Abb. 3 dargestellten Elemente und Relationen aufgenommen.

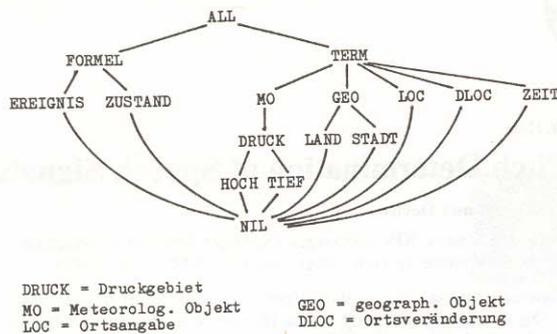


Abb. 3: Sortenverband (Eimermacher 1984:93)

Auf diesem überschaubaren, durchstrukturierten Fachgebiet hat sich WEP als erfolgreich erwiesen. Die zu bearbeitenden Mehrdeutigkeiten beschränken sich allerdings auf einen relativ kleinen Bereich der Zuordnung von Argumenten zu Rollen; Bedeutungsvarianten eines Wortes entfallen ganz bis auf einige Präpositionen, bei denen nach Orts- und Richtungsangabe differenziert wird. Über die Effizienz der Disambiguierung läßt sich daher nur wenig feststellen.

Ähnliches gilt für den Word Expert Parser, der im Konstanzer Projekt TOPIC verwendet wird (Hahn/Reimer 1983). Zwar nimmt hier die Wissensbasis eine zentralere Rolle ein; als Organisationsprinzip für das Weltwissen wurde ein framebasierter Ansatz verwendet. Über die Möglichkeiten, die sich hieraus hinsichtlich der Auflösung von Mehrdeutigkeiten ergeben, liegen jedoch wegen des anderen Projektschwerpunkts von TOPIC (Untersuchung von Textphänomenen unter Verwendung von partiellem Parsing) noch keine näheren Angaben vor.

4. Zusammenfassung und Ausblick

Es hat sich gezeigt, daß Word Expert Parsing nicht wegen seines theoretischen Ansatzes (Sprachverstehen als Prozeß der Disambiguierung), sondern als alternative Parsingtechnik für Anwender interessant zu sein scheint. Das modulare Konzept der Wortexperten kann einerseits als Betonung idiosynkratischer Eigenschaften anstelle sprachlicher Regularitäten angesehen werden, bietet andererseits aber vor allem praktische Vorteile für die Programmierung: Die einzelnen Wortexperten können (theoretisch!) relativ unabhängig voneinander entworfen werden, sobald Richtlinien für die Schnittstellen (Art der Signale etc.) sowie eine inhaltliche Bestimmung der gemeinsamen Wissensbasis vorliegen. Die aufzunehmende Menge syntaktischen und semantischen Wissens ist individuell bestimmbar und kann genau auf einen bestimmten Verwendungszweck hin ausgerichtet werden (Pilotprojekte).

Der Beitrag dieses Ansatzes für die Lösung grundsätzlicher Probleme der semantischen Analyse über den Bereich einer Mikrowelt hinaus ist dagegen weniger leicht einzuschätzen. Einerseits wird die Frage nach der Organisation des Disambiguierungsprozesses, nach der Verzahnung verschiedener Analyseebenen klar beantwortet: Jeder Mehrdeutigkeit ist genau ein Wortexperte zugeordnet, der in der

Lage ist, das gesamte für die Disambiguierung nötige Wissen zu beschaffen. So lassen sich z.B. strukturelle Mehrdeutigkeiten, die nur mit semantischen Mitteln lösbar sind, problemlos bearbeiten. Feste Syntagmen erfordern ebenfalls keine Sonderbehandlung mehr. Sowohl die Textebene als auch die pragmatische Seite kommen allerdings durch die Wortbezogenheit deutlich zu kurz: die Darstellung der Satzbedeutung durch ein einfaches Konzept dürfte für die meisten Anwendungsbereiche denn doch zu wenig sein. Neben dem pragmatischen Aspekt halte ich jedoch die Frage nach der Wissensrepräsentation für das zentrale Problem. Der Beitrag von Small und Rieger beschränkt sich hier auf den Vorschlag einer Kopplung von Wissensbasis und Parsingmechanismus. Semantisches Parsing setzt aber detailliertere Konzepte für die Gewinnung und die Darstellung des Wissens voraus, das für die Analyse mehrdeutiger Kontexte benötigt wird. Wichtig ist insbesondere die praktische Realisierung solcher Konzepte über den Bereich einer Mikrowelt hinaus. Für die sprachorientierte KI-Forschung bleibt hier noch viel zu tun.

- [1] Auf den Anspruch der kognitiven Adäquatheit des WEP soll hier nicht eingegangen werden; vgl. Cottrell/Small 1984.
- [2] Konzepte werden im folgenden durch kursive Großbuchstaben, Wortformen durch Kleinbuchstaben dargestellt.
- [3] Pikanterweise war dieses Merkmal ursprünglich wohl nicht auf "Frucht", sondern auf "Kies" (zur Selektion der Bedeutung "Kiesgrube") gemünzt (vgl. Small/Rieger 1982: 139 ff.).

Cottrell, Garrison / Small, Steven 1984: Viewing Parsing as Word Sense Discrimination. In: Bara, Bruno / Guida, Giovanni (Hrsg.): Computational Models of Natural Language Processing. Amsterdam: North-Holland, S. 91-119.

Eimermacher, Michael 1983: Ein Word Expert Parser in PROLOG - Integration von Syntax und Semantik. KIT - Interner Arbeitsbericht 8, Berlin.

- 1984: Word Expert Parsing in Prolog. In: Rollinger, Claus-Rainer (Hrsg.): Probleme des (Text-)Verstehens. Ansätze der Künstlichen Intelligenz. Tübingen: Niemeyer, S. 89-103.

- 1985: NATAN - Natürlichsprachliche Anfrageschnittstelle für verteilte Datenbanksysteme. KIT-Report 27, Berlin.

- 1986: Wordexperts in PROLOG. In: Hellwig, Peter / Lehmann, Hubert (Hrsg.): Trends in der Linguistischen Datenverarbeitung. Hildesheim: Olms, S. 110-124.

Habel, Christopher 1985: Das Lexikon in der Forschung der Künstlichen Intelligenz. In: Schwarze, Ch.; Wunderlich, D. (Hrsg.): Handbuch der Lexikographie. Königstein: Athenäum, S. 441-474.

Hahn, Udo / Reimer, Ulrich 1983: Wortexperten-Parsing: Text-Parsing mit einer verteilten lexikalischen Grammatik im Rahmen des automatischen Textkondensierungssystems TOPIC. In: Linguistische Berichte 88, S. 56-78.

Reddig, Carola 1984: Word Expert Parsing - ein Überblick. In: Rollinger, Claus-Rainer (Hrsg.): Probleme des (Text-)Verstehens. Ansätze der Künstlichen Intelligenz. Tübingen: Niemeyer, S. 77-88.

Small, Steven 1983: Parsing as Cooperative Distributed Inference: Understanding through Memory Interactions. In: King, Margaret (Hrsg.): Parsing Natural Language. London: Academic Press, S. 247-276.

Small, Steven / Rieger, Chuck 1982: Parsing and Comprehending with Word Experts (A Theory and its Realization). In: Lehnert, Wendy G. / Ringle, Martin H. (Hrsg.): Strategies for Natural Language Processing. Hillsdale: Lawrence Erlbaum, S. 89-147.

STRUKTURIERTES WISSEN ALS GRUNDLAGE FÜR SPRACHVERSTEHENSPROZESSE

Helmar Gust

Fachbereich SLM der Universität Osnabrück
Neuer Graben (Kreishaus)

4500 Osnabrück

Sprachverstehen ist ein weitgehend wissensabhängiger Prozeß. Eine für Sprachverstehensprozesse geeignete Wissensbasis muß dabei die kontextabhängige Analyse der Eingabe und Synthese der Ausgabe ermöglichen, inkonsistentes Wissen tolerieren und die Kohärenzanalyse des Dialogs unterstützen.

Diese Aufgaben müssen zudem bei realistischen Dialogsystemen unter relativ harten Zeitbedingungen bewältigt werden. Eine naheliegende Forderung ist daher die nach einer geeigneten Strukturierung der Wissensbasis, die dafür sorgt, daß zu jeder Zeit nur ein kleiner Ausschnitt des gesamten Wissen in Betracht gezogen werden muß. Weiterhin dürfen mögliche Inkonsistenzen des Gesamtwissens die Arbeit des Systems nicht beeinträchtigen, solange sie die aktuelle Kommunikationssituation nicht betreffen. Hier wird ein Wissensbasismodell vorgestellt, das diese Probleme einer Lösung näher bringt.

1. Einleitung

Die Probleme der Strukturierung von Wissen sind nicht neu. Für KI-Systeme wurden bislang eine Vielzahl unterschiedlicher Ansätze entwickelt (Frames, Semantische Netze, Scripts, Mops, etc). Diese Strukturierungen beruhen alle auf spezifischen Wissensrepräsentationsmechanismen. In Konkurrenz zu diesen Ansätzen steht die logische Repräsentation von Wissen, die zunächst davon ausgeht, daß eine Wissensbasis ein Axiomensystem ist, d.h. eine unstrukturierte Menge logischer, als wahr anerkannter Aussagen. Der Zugriff auf die Wissensbasis erfolgt über eine Beweisprozedur, die versucht, Anfragen mit Hilfe der gegebenen Axiome zu beweisen. In diesem Kontext ist nicht a priori zu sehen, wie solche Aussagenmengen strukturiert werden können. Zudem ist ein solches Axiomensystem sehr empfindlich gegenüber Inkonsistenzen, da aus einem inkonsistenten Axiomensystem bekanntlich beliebige Aussagen folgen. Bei der Modellierung von Alltagswissen wird diese harte Konsistenzbedingung schwer realisierbar sein: Unser Wissen über die Welt, das wir zur Lösung unserer täglichen Probleme verwenden, wird kaum als ganzes konsistent sein. Trotzdem operieren wir sehr erfolgreich damit. Komplizierte Logiken, die die Handhabung von unsicherem und widersprüchlichem Wissen erlauben, sind in diesem Zusammenhang entwickelt worden (z.B. Belnap 1977, Rollinger 1984a).

Die Problematik der Repräsentation von Wissen als unstrukturierte Menge logischer Aussagen ist auch in einem anderen Bereich, der Wissenschaftstheorie, erkannt worden (Stegmüller 1979; Sneed 1971). Die dort angebotene Lösung trägt auch für die Repräsentation von alltäglichem Wissen einiges aus. Dieser Ansatz erlaubt es, die logische Repräsentation von Wissen beizubehalten, ohne die Nachteile einer unstrukturierten Wissensbasis bzw. einer komplizierten 'Widerspruchslogik' dafür in Kauf nehmen zu müssen.

Hier wird versucht, eine analoge Strukturierung logischer Wissensbasen einzuführen. Diese Strukturierung erlaubt die Abschwächung der Konsistenzforderung auf relevante Teilmengen der Wissensbasis. Verbunden damit ist die Möglichkeit effizienter Zugriffsmechanismen. Durch Restrukturierungen können komplizierte Lernprozesse modelliert werden.

2. Wissensnetze

Das Wissen wird in einem gerichteten Graphen, einem Netz, organisiert. Ein solcher Graph besteht aus einer Menge von Knoten, auf der eine Nachfolgerrelation definiert ist. (Zunächst seien nur Bäume zugelassen, eine Erweiterung auf gerichtete azyklische Graphen ist ohne größere Probleme möglich.) Die Knoten entsprechen zunächst nicht näher spezifizierten Wissenspaketen, die wir im folgenden auch als Wissensmodule bezeichnen wollen. Jeder Wissensbasiseintrag (in unserem Kontext eine logische Aussage) ist genau einem solchen Knoten zugeordnet. Die Kanten des Graphen haben zunächst keine inhaltliche Interpretation (im Gegensatz zu anderen Wissensbasisstrukturierungen, wie z.B. den Semantischen Netzen).

Jedem Knoten eines solchen Wissensnetzes können drei Wissensbereiche (eine Partition der Wissensbasis) zugeordnet werden:

sichtbares Wissen } verfügbares Wissen
erreichbares Wissen }

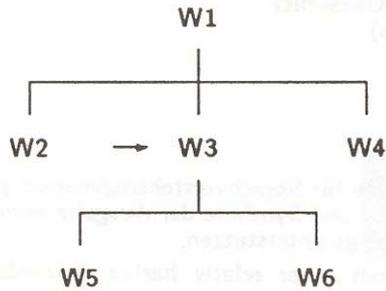
nicht verfügbares Wissen

Eine Kommunikationssituation ist unter anderem durch die Auszeichnung eines Knotens des Wissensnetzes, den aktuellen Fokus, gekennzeichnet. Das sichtbare und erreichbare Wissen des aktuellen Fokus bildet zusammen das aktuell verfügbare Wissen. Der Rest der Wissensbasis ist nicht verfügbar. Diese Unterteilung ist dynamisch und ändert sich abhängig von der Kommunikationssituation dadurch, daß sich der aktuelle Fokus ändert.

Der Bereich für die Überprüfung der Konsistenz ist das sichtbare Wissen. Die Wissensbasis kann daher insgesamt inkonsistent sein, ohne daß dies zu irgendeiner Zeit ins Gewicht fällt, solange das jedem Knoten des Netzes zugeordnete sichtbare Wissen konsistent ist. Eine Wissensbasis, die diese Eigenschaft besitzt möchte ich im folgenden als schwach konsistent bezeichnen.

Von jedem Knoten ist nur ein kleiner Bereich der Wissensbasis sichtbar. Daher können effiziente Zugriffsmechanismen entwickelt werden.

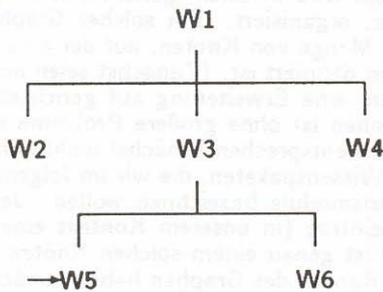
Beispiel:



sichtbar: W1 W3
 erreichbar: W5 W6
 aktuell: W1 W3 W5 W6

Erreichbares Wissen kann sichtbar werden, wenn der Fokusknoten entlang der Kanten im Netz nach unten wandert.

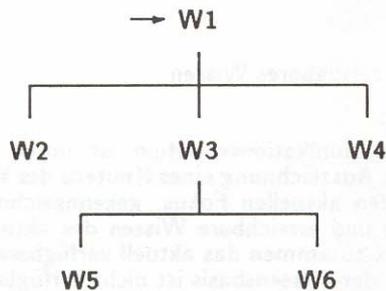
Beispiel:



sichtbar: W1 W3 W5
 erreichbar: -
 aktuell: W1 W3 W5

Sind Sachverhalte mit dem sichtbaren Wissen des aktuellen Fokus inkonsistent, so kann diese Inkonsistenz eventuell verschwinden, wenn der Fokus entlang der Kanten im Netz nach oben wandert. Dadurch wird das sichtbare Wissen verkleinert, jedoch andere Bereiche der Wissensbasis werden erreichbar.

Beispiel:



sichtbar: W1
 erreichbar: W2 W3 W4 W5 W6
 aktuell: W1 W2 W3 W4 W5 W6

Hieraus lassen sich nun einfache Regeln für die Dynamik des Fokusknotens formulieren:

1. Wird für die Analyse der Eingabe (Synthese der Ausgabe) erreichbares Wissen benötigt, so wandert der Fokusknoten zu dem entsprechenden Wissensmodul (nach unten).

2. Kann die Eingabe im aktuellen Fokus nicht analysiert werden (treten Inkonsistenzen auf), so wandert der Fokus nach oben, bis keine Inkonsistenzen mehr auftreten, und danach eventuell in einem anderen Bereich der Wissensbasis wieder zu spezielleren Wissensmodulen. (z.B. von W3 über W1 nach W4).
3. Kann auch dadurch keine Konsistenz erreicht werden, so muß das Wissensnetz verändert werden.
4. Eingegebene Sachverhalte werden, wenn die Analyse erfolgreich war, im dann aktuellen Fokusknoten eingetragen.

Auf der Menge des sichtbaren Wissens kann eine einfache Relevanzbewertung eingeführt werden, derart, daß die Relevanz der Wissensbasiseinträge mit größerem Abstand vom Fokus nach oben abnimmt. Mit diesem Ansatz wurde ein kleines wissensbasiertes System implementiert, bei dem allgemeineres Wissen im Netz weiter oben und spezielleres Wissen weiter unten angeordnet wurde.

3. Der Arbeitszyklus des Systems

Während der Übergang des Fokusknotens zu spezielleren Wissensmodulen ohne weiteres möglich ist (erreichbares Wissen ist auch aktuelles Wissen), ist der Übergang zu allgemeineren Wissensmodulen unter Umständen mit Aufwand verbunden. Das bezieht sich nicht auf den Zugriff auf das sichtbare Wissen, sondern auf die Änderung des Fokusknotens. Das den allgemeineren Wissensmodulen zugeordnete Wissen ist direkt verfügbar, mit der Änderung des Fokusknotens werden aber andere Bereiche der Wissensbasis erreichbar.

Während des Arbeitszyklus wird der Weg protokolliert, so daß ein Zurückgehen zu vorher benutzten Knoten ohne großen Aufwand möglich ist. Nach Beendigung des Arbeitszyklus wird dieses Protokoll vergessen. Als Arbeitszyklus wird hier zunächst die Verarbeitung einer Eingabe und die Erzeugung der entsprechenden Ausgabe angenommen. Der Fokus am Ende des Arbeitszyklus bleibt erhalten und wird als Ausgangspunkt für den nächsten Arbeitszyklus verwendet.

Größerer Aufwand ist erst dann nötig, wenn der Ausgangsfokus eines Arbeitszyklus bereits zu speziell war. Dies deutet auf eine Inkohärenz im Dialog hin. In diesem Fall muß der Fokus zu allgemeineren Wissensmodulen hin verlegt werden. Allerdings ist jetzt kein Protokoll der Aktionen mehr vorhanden, die während der letzten Arbeitszyklen durchgeführt wurden, so daß das Risiko besteht, daß im letzten Arbeitszyklus Änderungen der Wissensbasis erfolgten, die auf einer falschen Analyse beruhten. Der Aufwand bezieht sich darauf, diese noch zu erkennen und wenn nötig rückgängig zu machen.

Diese Organisation der Wissensbasis kann auch für die lexikalische Analyse der Eingabe ebenso wie für die Generierung der Ausgabe weitreichende Konsequenzen haben: Wenn bei einem Dialogsystem die lexikalischen Einträge in der Wissensbasis stehen und der Fokus auch in den Grammatikregeln mitgeführt wird, dann ist eine Beeinflussung des Fokus bereits durch die grammatische Analyse möglich. Andererseits reduziert diese Strukturierung der lexikalischen Informationen mögliche Mehrdeutigkeiten erheblich, da nur der aktuelle Bereich der Wissensbasis zur Analyse verwendet werden kann.

4. Realisierungsmöglichkeiten in Prolog.

Es gibt zwei prinzipiell verschiedene Realisierungsmöglichkeiten der hier beschriebenen Wissensbasisstruktur in Prolog, die folgendermaßen charakterisiert werden können:

1. Möglichst einfach und Prolog-nah.
2. Möglichst flexibel.

Im ersten Fall gehen wir von einer eindeutigen Indizierung der Wissensmodule aus. Das Wissensnetz wird als Menge teilinstanzierter Listenstrukturen repräsentiert. Die Komponenten der Listen sind die Indizes der Vorgänger im Netz. Die Listen haben ein offenes (variables) Ende. Die Vorgängerrelation kann dann wie folgt definiert werden:

T' ist ein Vorgänger von T, wenn die zugeordneten Listen unifizierbar sind und die Liste von T' kürzer als die von T ist.

Beispiel (':' markiert den Rest einer Liste und '*' ist eine Variable):

W3 = (1 3 :*)
 W5 = (1 3 5 :*)
 W3 ist ein Vorgänger von W5

Jeder Wissensbasiseintrag enthält die Struktur des Wissensmoduls, zu dem er gehört, als Argument. Der Fokus wird durch eine entsprechende Struktur repräsentiert

Bei dieser Realisierung gehören genau diejenigen Wissensbasiseinträge zum aktuellen Wissen, deren Wissensmodulmarkierungen mit dem Fokus unifizierbar sind. Die bei der Unifikation entstehende Struktur wird der neue Fokus. Daraus folgt:

1. Gehörte der Wissensbasiseintrag zum sichtbaren Wissen, so ändert sich der Fokus nicht.
2. Gehörte der Wissensbasiseintrag zum erreichbaren Wissen, so ist der neue Fokus gleich dem Wissensmodul, zu dem der Eintrag gehört.

Damit werden die Fokusbewegungen zu speziellen Wissensmodulen automatisch durch die Unifikation in Prolog durchgeführt. Da Prolog automatisch Backtracking durchführt, wenn die Analyse sich in einer Sackgasse befindet, wird auch der Übergang zu allgemeineren Wissensmodulen innerhalb des Arbeitszyklus automatisch bis zum Ausgangsfokus durchgeführt. Erst der Übergang zu noch allgemeineren Wissensmodulen muß explizit unter Berücksichtigung der schwachen Konsistenzbedingung durchgeführt werden.

Da Prolog die Wissensbasis sequentiell durchsucht, läßt sich die Relevanzbewertung dadurch erreichen, daß spezielleres Wissen vor allgemeinerem Wissen eingetragen wird.

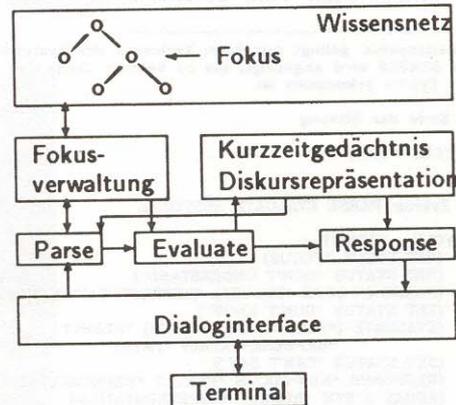
Bei dieser Realisierung wird die gesamte Wissensbasis durch ein einziges Prologprädikat dargestellt. Dies führt bei Prologsystemen, die nicht über ein ausgeklügeltes Verfahren für eine indizierte Zugriffsmöglichkeit auf Klauseln (die Indizes sind teilinstanzierte Strukturen) verfügen, zu einer sehr ineffizienten Zugriffstrategie. Die Einschränkung des Suchraums durch den Fokusmechanismus ist hier nur formaler Natur und schlägt sich nicht im Verhalten des Systems nieder.

Ein effizienteres System liefert die explizite Verwaltung des Wissensnetzes. Hier werden die Knoten des Wissensnetzes eigenständige Prologprädikate (eventuell sogar Moduln, falls das Prologsystem Mo-

duln unterstützt). Die Beziehungen zwischen den Knoten müssen explizit angegeben und von einem kleinen Prologprogramm verwaltet werden.

An der Universität Osnabrück wurden im Rahmen des Projektes SGK (Simulation grammatischer Kreativität) und in Zusammenarbeit mit der Arbeitsgruppe LAKOS (Wachsmuth 1985) Konzepte für wissensbasierte Dialogsysteme zur Simulation von Sprachprozessen entwickelt. Die Struktur dieser Systeme ist im folgenden wiedergegeben:

Das Dialogsystem LAKOS



Das folgende Prologprogramm ist eine Art Dialogsystem-Shell, d.h. ein Rahmenprogramm, das durch anwendungsspezifische Erweiterungen zu einem vollständigen Dialogsystem wird. Es ist in MLOG geschrieben (Lisp-Notation, '*' markiert Variablen, ':' markiert den Rest einer Liste).

1 (LAKOS *KNOWLEDGE-BASE) ←

Vorbereitung

```
(PL — LAKOS VERSION 1.5 —)
(KILL STM)
(ADDAX STM)
(KILL FOCUS)
(ADDAX FOCUS
 *KNOWLEDGE-BASE)
```

Dialog-(Backtracking-)Schleife

```
LOOP
PL
(R *INPUT *TYPE)
(CYCLE *TYPE : *INPUT).
```

```
-----
; Rudimentäre Begründungskomponente
; a) Eine konsistente Begründung konnte gefunden werden
1 (REASON (! BECAUSE *PATH) *RESULT *RESULT *PATH) .

; b) Im aktuellen Fokus ist keine konsistente Begründung
; möglich.
2 (REASON (??? *LAST-ANSWER *RESULT *PATH)
 *RESULT *LAST-ANSWER *PATH) .

-----
; Evaluiert eine Liste von Fragen oder Bedingungen
1 (EVALUATE-LIST () ()) .
2 (EVALUATE-LIST (( *ANSWER : *PATH) : *FACTS)
 (*FOCUS *QUESTION *ANSWER) : *CONDITIONS)) <-
(EVALUATE (*FOCUS *) *QUESTION *ANSWER *PATH)
(EVALUATE-LIST *FACTS *CONDITIONS) .

; Evaluiert eine Frage oder Bedingung
; a) Erklärungsfrage
1 (EVALUATE (*FOCUS *FOCUS) *ANSWER (? WHY) *PATH) <-
(STM *LAST-ANSWER *LAST-QUESTION)
(EVALUATE (*FOCUS *) *RESULT *LAST-QUESTION *PATH)
(REASON *ANSWER *RESULT *LAST-ANSWER *PATH) .
```

```

; b) Die Antwort steht im Kurzzeitgedächtnis
2 (EVALUATE *FOCUS *ANSWER *QUESTION ()) <-
  (STM *ANSWER *QUESTION) .

3 (EVALUATE *FOCUS (! TRUE : *QUESTION)
  (? : *QUESTION) ()) <-
  (STM (! *POSITION : *QUESTION) *) .

; c) In der Wissensbasis steht ein Faktum
4 (EVALUATE *FOCUS (! : *RESULT) (? : *QUESTION) ()) <-
  (ACCESS *FOCUS FACT *RESULT *QUESTION) .

; d) In der Wissensbasis steht eine Regel
5 (EVALUATE (*FOCUS *NEW-FOCUS) (! : *RESULT)
  (? : *QUESTION) *PATH)) <-
  (ACCESS (*FOCUS *NEW-FOCUS) RULE *RESULT *QUESTION
  : *CONDITIONS)
  (EVALUATE-LIST *PATH *CONDITIONS) .

```

Arbeitszyklus: gelingt nur beim Verlassen des Systems
 ; Mit STATUS wird angezeigt, bis zu welcher Stelle der
 ; der Zyklus gekommen ist.

; a) Ende der Sitzung

```

1 (CYCLE ! BYE) <-
  /

```

; b) Zyklus: PARSE EVALUATE RESPONSE

```

2 (CYCLE : *INPUT) <-
  (GET-FOCUS *FOCUS)
  (SET STATUS "DON'T UNDERSTAND")
  (PARSE (*FOCUS *FOCUS1) *REPRESENTATION *INPUT)
  (SET STATUS "DON'T KNOW")
  (EVALUATE (*FOCUS1 *NEW-FOCUS) *RESULT
  *REPRESENTATION *PATH)
  (SET STATUS "CAN'T SAY")
  (RESPONSE *NEW-FOCUS *RESULT *REPRESENTATION *PATH)
  (ADDAX 1 STM *RESULT *REPRESENTATION)
  (TRUE (KILL 0 STM))
  (KILL FOCUS)
  (ADDAX FOCUS : *NEW-FOCUS)
  /
  FAIL .

```

; Fehlermeldung

```

3 (CYCLE : *INPUT) <-
  (STATUS : *STATUS)
  PL
  (P " " : *STATUS)
  (PL !)
  /
  FAIL .

```

Fokustransformation, falls die Verarbeitung im aktuellen
 ; Fokus nicht gelingt: Verschiebung zu Vorgängern.

```

1 (GET-FOCUS *FOCUS) <-
  (FOCUS : *FOCUS) .
2 (GET-FOCUS *FOCUS) <-
  (DELAX FOCUS *OLD-NODE *NODE : *PREDECESSORS)
  (ADDAX FOCUS *NODE : *PREDECESSORS)
  (GET-FOCUS *FOCUS) .

```

PARSE und RESPONSE sind anwendungsspezifische Prädikate. PARSE übersetzt die konkrete Eingabe in eine interne Repräsentation, RESPONSE übersetzt die interne Repräsentation der Antwort wieder zurück.

ACCESS, das Programm für die Verwaltung der Wissensbasis, ist sehr einfach. Es legt für jeden Knoten die zugeordneten Wissensbereiche fest.

```

1 (ACCESS (*FOCUS *NEW-FOCUS) : *KNOWLEDGE) ←
  (VISIBLE *FOCUS : *KNOWLEDGE) .
2 (ACCESS (*FOCUS *NEW-FOCUS) : *KNOWLEDGE) ←
  (ACCESSIBLE (*FOCUS *NEW-FOCUS) :
  *KNOWLEDGE) .

1 (VISIBLE (*NODE : *PREDECESSORS) : *KNOWLEDGE) ←
  (*NODE : *KNOWLEDGE) .
2 (VISIBLE (*NODE : *PREDECESSORS) : *KNOWLEDGE) ←
  (VISIBLE *PREDECESSORS : *KNOWLEDGE) .

1 (ACCESSIBLE ((*NODE : *PREDECESSORS)
  (*NEW-NODE *NODE : *PREDECESSORS))
  : *KNOWLEDGE) ←
  (*NODE SUCCESSOR *NEW-NODE)
  (*NEW-NODE : *KNOWLEDGE) .

```

```

2 (ACCESSIBLE ((*NODE : *PREDECESSORS) *NEW-FOCUS)
  : *KNOWLEDGE) ←
  (*NODE SUCCESSOR *NEW-NODE)
  (ACCESSIBLE ((*NEW-NODE *NODE :
  *PREDECESSORS) *NEW-FOCUS) : *KNOWLEDGE) .

```

Die Struktur der Wissensbasiseinträge hat die Form von Frage-Antwort-Paaren, dabei können bei der Antwort folgende Typen unterschieden werden: Positive und negative Antwort, sowie Antworten auf Entscheidungsfragen bzw. Ergänzungsfrage. Die Wissensbasiseinträge haben die Struktur:

- {wissensmodul FACT antwort frage}
- {wissensmodul RULE antwort frage bedingungen}
- {wissensmodul SUCCSESSOR folgemodul}

Die Bedingungen bei einer Regel sind Tripel aus einem Fokus, in dem sie evaluiert werden sollen, einem Antwortmuster und einem Fragemuster. Diese Struktur der Wissensrepräsentation ist für das vorgestellte Konzept nicht essentiell und birgt sicherlich Probleme in sich. Es wurde hier gewählt, um auf einfache Weise negative Antworten zuzulassen, ohne auf die auch problematische 'negation-as-failure'-Regel zurückgreifen zu müssen, die sonst in Prolog üblich ist. Andere Wissensbasiseinträge sind ohne weiteres möglich und werden im folgenden auch benutzt.

5. Anwendung auf die Nominalkomposition

Im Projekt SGK in Osnabrück wird diese Strukturierung der Wissensbasis zur Reduzierung der Mehrdeutigkeiten bei der Analyse von Nominalkomposita verwendet. Die Analyse von Nominalkomposita ist in den überwiegenden Fällen ein wissensabhängiger Prozeß (Kanngießer 1985). Das benötigte Wissen ist teils allgemeiner Natur, teils jedoch sehr spezielles Wissen über den aktuellen Kommunikationskontext. Insbesondere ad-hoc-Kompositionen sind häufig nur mit Hilfe der letzteren Art von Wissen zu analysieren.

Das allgemeine Regelschema für die Komposition lautet: Es gibt eine Relation zwischen der Bedeutung der Komponenten und der Bedeutung des Kompositums. Dies allgemeine Regelschema ist unbefriedigend, solange nicht näher spezifiziert werden kann, welcher Art diese Relation ist. In (Kanngießer 1985) sind für die Nominalkomposition einige speziellere Regeln angegeben. Ein Großteil der komplizierteren Nominalkomposita wird nach folgendem vereinfachten Regelschema gebildet:

$$\exists w \in W, \exists r \in w(des(b)) :$$

$$des(ab) = \{x | x \in des(b) \wedge \exists y \in des(a) : r(x, y)\}$$

Dabei ist W eine Menge von Wissenskategorien (siehe (Kanngießer 1985)) und $w(des(b))$ die zu einer solchen Wissenskategorie gehörenden Wissens-elemente. So ist ein Fahrraddieb jemand, der ein Fahrrad gestohlen hat. Die Relation 'stehlen' gehört zu den intrinsischen Bedeutungselementen von 'dieb'.

Paraphrasieren läßt sich dies Regelschema wiederum vereinfacht so: Es gibt eine geeignete Relation 'r', in der die Bedeutungen der Komponenten zueinander stehen. Das 'geeignet' in dieser Paraphrasierung kann zwar noch etwas spezifiziert werden, dies ändert jedoch nicht viel an der möglichen Inflation von Bedeutungen von Komposita. Bei lexikalisierten Komposita geht diese bedeutungskonstituierende Relation häufig sogar verloren. Wer weiß z.B. schon

(außer den Herstellern), daß eine Kalbsleberwurst nicht etwa eine Wurst aus Kalbsleber, sondern eine Wurst aus Schweineleber ist, die etwas Kalbfleisch enthält. Um die Anzahl der möglichen Bedeutungen zu begrenzen ist ein Wissensbasismodell angebracht, bei dem in Abhängigkeit vom aktuellen Kommunikationskontext nur ein möglichst begrenzter Ausschnitt der Wissensbasis zur Analyse herangezogen werden muß.

So kann eine Plastikscherer je nach Kontext z.B. eine Schere zum Schneiden von Plastik oder aber eine Schere aus Plastik oder vielleicht sogar eine Schere eingewickelt in Plastikfolie sein. Hingegen dient eine Regalschere wohl weniger zum Schneiden von Regalen noch besteht sie aus Regalen. Möglicherweise ist ihr angestammter Platz in einem Regal, oder es handelt sich um eine scherenförmige Verstrebung eines Regals.

In der obigen Kompositionsregel wird an einer Stelle explizit auf Wissen zugegriffen:

$$\exists r \in w(des(b))$$

$w(des(b))$ beschreibt die Menge der Relationen, die die Bedeutung eines Kompositums nach dem obigen Regelschema konstituieren können. Diese Menge kann sinnvollerweise auf die Menge derjenigen Relationen, die im aktuellen Fokus sichtbar sind, beschränkt werden. Implizit gibt es noch zwei weitere Stellen, an denen das Fokuskonzept zur Reduktion von Mehrdeutigkeiten beiträgt: $des(a)$ und $des(b)$ können selbst mehrdeutig sein. Da diese beiden Ausdrücke aber auch im aktuellen Fokus berechnet werden müssen, gilt hierfür (rekursiv) das gleiche Argument.

In einem Kontext, in dem es um das Bearbeiten verschiedener Materialien geht, sind die obigen Kompositionen mit 'Scherer' einfach zu analysieren: Die Relation 'dazu dienen, etwas zu schneiden' ist hier ein intrinsisches Bedeutungselement von 'Scherer'. Dies kann vermittels LAKOS repräsentiert werden:

bearbeiten IB ((dienen *x (schneiden *x *y))
(material *y) (schere *x)).

Den Eintrag interpretieren wir im Sinne einer Lambda-Abstraktion für Relationen: Angewendet auf einen Term der Sorte 'schere' erhalten wir ein Prädikat, das auf einen Term der Sorte 'material' angewendet werden kann, und dann eine Formel liefert. Die Lambda-Reduktion wird durch die Unifikation bewerkstelligt.

Im Kontext 'möbel' spielen Beschläge zur mechanischen Stabilisierung eine wesentliche Rolle:

möbel SW ((dienen *x (stabilisieren *x *y))
(möbelstück *y) (beschlag *x)).

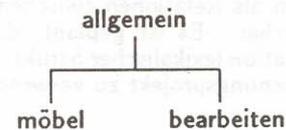
Allgemein gilt, daß auf fast alle Dinge die Relationen 'besteht aus' und 'hat die Form von' anwendbar sind:

allgemein SW ((bestehen-aus *x *y) (ding *x)
(material *y))

allgemein SW ((hat-form *x *y) (ding *x)
(ding *y))

Wir erhalten damit ein Wissensnetz der Form:

Beispiel:



Im Fokus 'möbel' steht dann zur Konstitution der Bedeutung des Kompositums 'Metallschere' nur die komplexe Relation

(and (bestehen-aus *x *y)
(dienen *x (stabilisieren *x *z))
(möbelstück *z))
(beschlag *x) (metall *y))

zur Verfügung. Im Fokus 'bearbeiten' sind dagegen die beiden folgenden Lesarten möglich:

- a) (bestehen-aus *x *y) (schere *x) (metall *y)
b) (dienen *x (schneiden *x *y))
(schere *x) (metall *y)

6. Erweiterungen des Systems

Die 'depth first' Suchstrategie von Prolog ist sicher nicht optimal für ein solches Wissensnetz. Effektiver dürfte eine beschränkte Breitensuche sein. An einer solchen Implementation wird gearbeitet.

Eine naheliegende Erweiterung dieses Systems verwaltet mehrere Foki. Damit sind Phänomene der Dialogführung wie Turntaking, Themenwechsel, Zurückkommen auf vorher angesprochene Themen, Bearbeitung alternativer Suchräume etc. sowie vielfältige Kompositionsprobleme, bei denen ein gleichzeitiger Zugriff auf zunächst unabhängige Knoten des Wissensnetzes notwendig ist, modellierbar.

Ein weiterer wichtiger Komplex sind Netztransformationen, die aus zwei Gründen insbesondere für die Wissensakquisition in solchen Wissensnetzen notwendig sind. Zum einen kann die Netzorganisation ineffizient werden, wenn einfach nur neue Einträge an bestehende Knoten gehängt werden. Eine Restrukturierung kann daher effizientere Zugriffe ermöglichen. Zum anderen kann die Voraussetzung der schwachen Konsistenz durch neue Einträge verletzt werden. Durch eine Restrukturierung kann diese Voraussetzung wieder hergestellt werden. Netztransformationen setzen jedoch eine globale Übersicht über die Netzstruktur voraus. Hierfür existieren nur erst Ansätze.

Bislang wurde über die interne Struktur der Knoten eines Wissensnetzes nichts weiter gesagt. Tatsächlich sind die bisherigen Überlegungen auch vollkommen unabhängig davon. Auch die Kanten des Wissensnetzes haben bislang keine inhaltliche Interpretation erfahren. Damit stehen auch keine inhaltlich begründbaren Strukturierungsstrategien zur Verfügung. Es bleiben damit neben der schwachen Konsistenzforderung ausschließlich pragmatische Strukturierungsstrategien, wie Zugriffseffizienz.

Dieser Zustand ist sicher für viele Anwendungen unbefriedigend. Andererseits sind die bislang dargestellten Konzepte dadurch allgemeingültig (d.h. unabhängig von eventuellen inhaltlichen Interpretationen von Knoten und Kanten). In Analogie zu (Sneed 1971) können die einzelnen Knoten jedoch

noch weiter strukturiert werden. Dann werden auch die Kanten als Relationen zwischen Wissensmoduln interpretierbar. Es ist geplant, diesen Ansatz zur Repräsentation lexikalischer Strukturen in einem weiteren Forschungsprojekt zu verwenden.

Literatur

Wachsmuth, Ipke 1985: LAKOS - Ein Modell der Wissensrepräsentation zur Erklärung kognitiven Verhaltens. In: H. Mandl, D.M. Fischer (Hrsg.): Lernen im Dialog mit dem Computer. Urban und Schwarzenberg 1985, S. 24-39.

Belnap, N. 1977: A Useful Four-Valued Logic. In: J.M. Dunn, G. Epstein (Hrsg.): Modern Uses of Multiple-Valued Logic, Dordrecht 1977.

Dunn, J.M.; Epstein, G. (Hrsg.) 1977 Modern Uses of Multiple-Valued Logic. Dordrecht 1977.

Mandl, H.; Fischer, D.M. (Hrsg.) 1985: Lernen im Dialog mit dem Computer. Urban und Schwarzenberg 1985.

Kanngießler, Siegfried 1985 Strukturen der Wortbildung. In: Ch. Schwarze, D. Wunderlich: Handbuch der Lexikologie. Königstein: Athenäum 1985.

Rollinger, Claus-Rainer (Hrsg.) 1984: Probleme des (Text-)Verstehens. Tübingen: Niemeyer 1984.

Rollinger, Claus-Rainer 1984a: Evidenzbewertung für Fakten und Regeln. In: C.-R. Rollinger (Hrsg.): Probleme des (Text-)Verstehens. Tübingen: Niemeyer 1984, S. 209-219.

Sneed, J.D. 1971: The Logical Structure of Mathematical Physics. Dordrecht 1971.

Stegmüller, W. 1979: The Structuralist View of Theories. Heidelberg: Springer 1979.

Schwarze, Christoph; Wunderlich, Dieter 1985: Handbuch der Lexikologie. Königstein: Athenäum 1985.

Stellenanzeige

Zur Weiterentwicklung unseres PONS Wörterbuchprogramms suchen wir eine(n)

DV-Linguisten(in)

Wir denken dabei an eine(n) qualifizierte(n) Absolventen(in) des Studiums der linguistischen Datenverarbeitung mit sehr guten englischen und französischen Sprachkenntnissen.

Erwartet werden ausbaufähige Kenntnisse im Umgang mit PC, Großrechnern, den Programmiersprachen PL 1 und C sowie möglichst Erfahrungen mit relationalen Datenverwaltungssystemen. Kenntnisse in computerunterstützter Terminologiearbeit wären von Vorteil.

Vor einem persönlichen Gespräch, in dem wir mit Ihnen die Position und die vertraglichen Bedingungen besprechen möchten, bitten wir Sie, uns Ihre kompletten Bewerbungsunterlagen zuzuschicken. Bitte, geben Sie auch eine Telefonnummer an, unter der wir Sie zur Vereinbarung eines Vorstellungstermins erreichen können.



ERNST KLETT VERLAG
Referat Personal
Rotebühlstraße 77
7000 Stuttgart 1

Klett

SYSTEME ZUR GENERIERUNG VON TEXTEN EINE ÜBERSICHT

Dietmar Rösner

Projekt SEMSYN am Institut für Informatik
Universität Stuttgart
Herdweg 51
D-7000 Stuttgart 1

Die Generierung natürlicher Sprache und insbesondere die Generierung von Texten ist ein Gebiet, das weltweit erst in Entwicklung begriffen ist. Die meisten bisherigen Arbeiten hierzu stammen aus dem englischen Sprachraum, vornehmlich aus den USA. Die im Folgenden vorgestellten 12 Systeme unterscheiden sich in erster Linie in Bezug auf die Art der generierten Texte und der zugrundeliegenden (semantischen) Repräsentation sowie hinsichtlich Forschungsinteresse und Anspruch des gewählten Ansatzes.

1 Einleitung

Die Generierung natürlicher Sprache und insbesondere die Generierung von Texten ist ein Gebiet, das weltweit erst in Entwicklung begriffen ist. Die meisten bisherigen Arbeiten hierzu stammen aus dem englischen Sprachraum, in erster Linie aus den USA. In einer Übersicht (Mann et al. 1981) wurden noch 1981 lediglich sieben - ausschließlich US-amerikanische - Projekte aufgeführt, eine Zahl, die allerdings schon damals nicht mehr als repräsentativ gelten konnte. Im Anschluß an die IJCAI 1983 in Karlsruhe hat das Stuttgarter Projekt SEMSYN den ersten Internationalen Workshop zu Fragen der Sprachgenerierung veranstaltet (Stettenfels 1983) ([1]). Hier waren bereits knapp zwanzig Projekte vertreten. Im folgenden werden wichtige Systeme und Arbeiten zur Generierung natürlicher Sprache vorgestellt ([2]). Um Vergleiche wenigstens im Ansatz zu ermöglichen, wird sich die Besprechung an dem in Abb. 1 aufgeführten Raster orientieren.

- 1 Autor(en) und Name des/r Systems/e
- 2 Art der generierten/behandelten Texte
- 3 Gebiet(e) des Diskurses
- 4 behandelte Sprache(n)
- 5 Implementationssprache und Implementationsstatus
- 6 semantische Repräsentation:
Charakterisierung als Datenstruktur
Wie wird semantische Repräsentation gewonnen?
- 7 behandelte (text-)linguistische Phänomene
- 8 ggf. verwendete Grammatiktheorie
- 9 Simulation menschlicher Performanz?

Abb. 1: Ein Raster zum Vergleich von Generierungssystemen

2 Goldmans BABEL

Die auf Schanks Theorie der konzeptuellen Dependenz (engl.: conceptual dependency, kurz: CD) basierende semantische Repräsentation (Schank, Abelson 1977) ist Ausgangspunkt für eine ganze Reihe der zu besprechenden Generatoren. Das erste Programm, das aus CD-Strukturen Text generierte - und zwar zunächst einzelne englische Sätze -, war der von Goldman für Schanks MARGIE-Projekt entwickelte BABEL-Generator (Goldman 1975). Er wurde - nach einigen Erweiterungen - auch in vielen der späteren Projekte an der Universität Yale eingesetzt, um zum Beispiel Zusammenfassungen für verarbeitete Texte zu generieren. Versuchen mit Daten der entsprechenden Sprachen, soll BABEL auch Mandarin, Russisch, Holländisch und Spanisch produziert

haben (Schank, Abelson 1977). Eine wichtige Teilaufgabe für BABEL war es, für primitive Aktionen der CD-Theorie passende (englische) Verben zu finden. Diese Auswahl mußte kontextsensitiv sein. Sie wurde mithilfe von Diskriminationsnetzen implementiert. In diesen wurde in Abhängigkeit von den vorhandenen Tiefenkasus und den Eigenschaften der beteiligten "picture producers" solange verzweigt, bis eine eindeutige Wortwahl zustande kam. Ein aufs Deutsche übertragenes Beispiel (Wahlster 1982): die primitive Aktion INGEST wird als "saufen" lexikalisiert, wenn das OBJECT flüssig und der ACTOR ein Tier ist, sonst als "trinken"; ist das OBJECT andererseits gasförmig, wird "einatmen" gewählt. Bei der späteren Anwendung von BABEL in Systemen, die Scripts verwendeten - also Repräsentationen für stereotype Situationen wie den Besuch eines Restaurants (\$RESTAURANT-Script) -, war auch eine "script-sensitive" Lexikalisierung für primitive Aktionen und konzeptuelle Objekte möglich: Das konzeptuelle Objekt CHAIR wurde im \$BUS-Script als "seat" und im Kontext von \$RESTAURANT als "chair" lexikalisiert.

3 Incremental Procedural Grammar (IPG)

An der Universität Nijmegen entwickelten Hoenkamp und Kempen ein Programm, das stark psycholinguistisch orientiert ist und die inkrementelle Sprachproduktion simulieren soll. Das Programm ist so angelegt, daß es, einem Menschen vergleichbar, dem die auszusprechenden Gedanken auch erst während des Sprechens nach und nach ins Bewußtsein aufsteigen, seine Generierung beginnt, sobald eine erste konzeptuelle Einheit - kodiert als CD-Graph - vorliegt, und das die folgenden Äußerungen dann in den so entstehenden Rahmen einzupassen sucht. Die der Implementation zugrunde gelegte Theorie sollte gerade auch typische Fehler menschlicher Sprecher - unvollständige oder nicht beendbare Sätze, erneute Anläufe usw. - erklären können (Hoenkamp 1983). Neben Holländisch kann IPG auch ein Fragment des Deutschen produzieren.

4 Generierung von Japanisch aus CD

Ishizaki implementierte bei einem Aufenthalt an der Yale University einen Generator, der aus CD-Repräsentationen Japanisch produziert. Diese Arbeiten führte er in Japan fort (Ishizaki 1983). Die CD-Repräsentationen wurden bei der Analyse von Meldungen spanischsprachiger Zeitungen über internationalen Terrorismus gewonnen. Um der japanischen Textkonvention gerecht werden zu können, daß Sätze in

temporalen Ordnung aufeinander folgen, muß Ishizakis Programm zunächst die temporalen Beziehungen der berichteten Ereignisse erschließen. Dazu bedient es sich der von Schank vorgeschlagenen Datenstruktur der MOPs, der Memory Organization Packets (Schank 1980). Ferner sind, damit das generierte Japanisch akzeptabel wird, Konjunktionen entsprechend den kausalen Beziehungen aufeinanderfolgender Ereignisse zu erzeugen und Auslassungen von Konstituenten (vor allem Subjekt, ggf. auch Objekt) vorzusehen.

5 KAFKA, der Generator des XCALIBUR-Projekts

Mauldin stellt einen Generator für englische Sätze vor, der - wie er schreibt - auf Goldmans BABEL zurückgeht und mit seinem Vorläufer vor allem noch Gemeinsamkeiten bei der Verblektion hat (Mauldin 1984). Eingesetzt wird KAFKA im XCALIBUR-Projekt der Carnegie-Mellon University. In diesem Vorhaben wird eine natürlichsprachliche Schnittstelle für verschiedene Expertensysteme entwickelt. Zur Kommunikation zwischen den Moduln von XCALIBUR wird eine Frame-Notation als einheitliche Interlingua verwendet. Innerhalb dieses Formats können gleichzeitig semantische und syntaktische Informationen angegeben sein. KAFKA beschränkt seine Operationen - im Sinne von Chomskys generativer Grammatiktheorie - auf lediglich einen Formalismus: die Anwendung von Transformationsregeln auf Frames. Auf Frames mit semantischer Information werden solange Transformationsregeln angewandt, bis - sofern kein Fehlschlag eintritt - nur noch syntaktische Information vorliegt und die angestrebte Äußerung dann durch Traversieren dieses Baumes und Ausdrucken der lexikalischen Terminale möglich ist.

6 Jacobs' PHRED

Der von Paul Jacobs in Berkeley implementierte Generator PHRED (= PHRasal English Diction) wird vor allem im Rahmen des UNIX Consultant (UC) eingesetzt (Jacobs 1984). Natürlichsprachliche Fragen an den UNIX-Consultant werden von PHRAN (= PHRasal ANalyzer) analysiert; der UC-Planer produziert eine Antwort in Form einer erweiterten CD-Repräsentation (vgl. 2). PHRED generiert daraus die Hilfeinformation in Form englischer oder spanischer Einzelsätze.

```
Vom UC gelieferte konzeptuelle Struktur:
(planfor
  (result (state-change (actor file1)
    (state-name location)
    (from (inside-of (object current-directory)))
    (to (not (concept (inside-of (object current-directory))))))
  (method (mtrans (actor *users*)
    (object (command (name rm)
      (args (filename))
      (input *standard*)
      (output *standard*)
      (diagnostic *standard*)))
    (from *users*)
    (to *Unix*))))))
```

Von PHRED generierte Hilfeinformation:
 "Typing 'rm filename' causes the file filename to be removed from the current directory"

Abb. 2: Generierung im UNIX-Consultant

Der PHRED-Generator ist so entworfen worden, daß er mit dem Analyseprogramm PHRAN ein und dieselbe Wissensbasis teilen kann, in der

Wissen über die Sprache einheitlich in deklarativer Weise als sog. "pattern-concept pairs" (PC-Paare) abgelegt ist. "The pattern-concept pairs used in PHRAN/PHRED are associations between linguistic structures and their conceptual and functional representations" (Jacobs 1984). Unterschiedlich ist bei Parser und Generator lediglich die Art des Zugriffs auf die Wissensbasis. PHREDs Generierung besteht aus drei Phasen, die bei der rekursiven Verarbeitung der konzeptuellen Strukturen wiederholt durchlaufen werden:

- Fetching: mit einem speziellen Hashing-Mechanismus wird zum aktuellen Konzept aus der Wissensbasis ein geordneter Strom von PC-Paaren als mögliche Kandidaten erzeugt (theoretische Sicht: Predisposition).
- Restriction: die Kandidatenpaare werden solange nacheinander mit den Spezifikationen des Konzepts verglichen, bis hinreichende Übereinstimmung vorliegt; durch "Unifikation" können weitere Bedingungen zu einem Pattern hinzugefügt werden (theoretische Sicht: Selection).
- Interpretation: die Produktion der grammatisch korrekten sprachlichen Oberfläche zu einem gewählten und ausreichend spezifizierten Pattern. PHRED ist nur für die Generierung einzelner Sätze ausgelegt (".. it has no knowledge of coherence" (Jacobs 1984)); bei anaphorischen Konstituenten werden allerdings Heuristiken angewandt, die entweder zu Elision oder Pronominalisierung und nur in Ausnahmefällen zur erneuten Generierung führen.

7 Generierung mit "direkter Methode" in XPLAIN

Swartout hat sich bei XPLAIN ganz auf das Problem konzentriert, wie ein Expertensystem - hier ein Programm zur Unterstützung von Ärzten beim Einsatz von Digitalis - Erklärungen geben und Rechtfertigungen seines Handelns erzeugen kann. Der Generator diene, wie er selbst einräumt, lediglich dazu, diese Fähigkeiten dann in akzeptablem Englisch vorführen zu können. "The generator should really be viewed more as an engineering effort ... than as a generation system that encodes deep linguistic principles" (Swartout 1981). Wir besprechen den Generator dennoch als Beispiel für die sogenannte "direkte Methode". Zur Wissensrepräsentation verwendete XPLAIN sogenannte XLMS-concepts, d.h. ggf. verschachtelte Strukturen in der Notation

-(< ilk > * < tie > < cue >)-,

wobei < ilk > und < cue > selber wieder XLMS-concepts sein konnten und * < tie > die Beziehung zwischen < ilk > und < cue > bezeichnete. Für die verschiedenen Typen von * < tie > s gab es dann Techniken, um die verbundenen XLMS-concepts "direkt", also ohne jegliche "linguistische" Zwischenstufe, in Englisch umzusetzen. Für die Beziehungen *f für "function" und *o für "object" wurde aus -((pvcs*f dangerous)*f (induced*o (by*o digitalis))) - die Phrase "dangerous pvcs induced by digitalis" konkateniert (pvcs = premature ventricular contractions). Die Grenzen dieser "direkten Methode" sind offensichtlich. Ein - von Swartout selbst angegebenes - Beispiel mag zur Illustration genügen, welche Probleme der Verzicht auf eine linguistische Repräsentation mit sich bringen kann:

"If the concept -block- is flagged to indicate that an article should be inserted, the system will generate the phrase "the red the block" when asked to generate a phrase for the concept -(block*f red)-. This is because the system inserts an article because it inherits the flag from -block- when generating a phrase for -(block*f red)- then it inserts another article when generating a phrase for -block-" (Swartout 1981, p. 84).

8 McKeowns TEXT

Das System TEXT produziert mehrsätzliche Antworten auf (formalsprachliche) Anfragen zu Struktur und Inhalt einer Datenbank mit Informationen über Waffen und Schiffe. McKeown 1983 konzentriert sich vor allem darauf, wie Fokus-Information sowohl die Auswahl des "what to say", als auch die Entscheidung zwischen möglichen Oberflächenstrukturen beeinflussen kann. Ausgangspunkt für TEXTs Antwortgenerierung sind Propositionen - Datenstrukturen aus einem Prädikat und seinen Argumenten -, die nacheinander aus der Wissensbasis erschlossen werden. Propositionen sollen im englischen Text als Verben, ihre Argumente als Füller der zugehörigen Kasusrollen realisiert werden. Zu jeder ausgewählten Proposition wird als Fokus-Information das aktuell fokussierte Argument und die Liste der für eine Fokusverschiebung in Frage kommenden Kandidaten aufgezeichnet. Diese Information wird zunächst eingesetzt, um dann, wenn weitere Propositionen zur Debatte stehen, diejenige bevorzugt zu wählen, für die sich gemäß folgender Liste von "focus constraints" die höchste Präferenz ergibt:

- 1) Fokusverschiebung zu gerade erwähnter Grösse
- 2) Aufrechterhalten des Fokus
- 3) zurück zu Topik der früheren Diskussion
- 4) die Proposition mit der Maximalzahl von impliziten Beziehungen zur vorherigen Proposition.

Fokusinformation beeinflusst in TEXT aber auch Entscheidungen über die Oberflächenstruktur, z.B. die Wahl zwischen Aktiv, Passiv und There-Insertion oder die zwischen Pronominalisierung und Generierung einer kompletten NG. Ein Beispiel: Im Rahmen der Antwort auf die Frage nach dem Unterschied zwischen Missiles und Torpedos wurde die folgende Proposition gewonnen:

predicate = analogy-relation
 relation = ON
 carrier = (AIR-VEHICLE WATER-VEHICLE)
 weapon = MISSILE
 current focus = MISSILE
 potential focus list =
 ((AIR-VEHICLE WATER-VEHICLE) ON
 analogy-relation)

Nach Verbauswahl und Lexikalisierung ergibt sich die folgende funktionale Struktur:

verb === carry
 protagonist =
 conj === and
 head1 === aircraft
 head2 === water-going vehicle
 goal === missile
 voice === passive

Da der Füller der Rolle goal im Fokus steht, wurde hier Passiv gewählt und der erzeugte Teil der Antwort lautet damit:

"Missiles are carried by water-going vehicles and aircraft".

9 Gabriels YH: Wie schreibt man sorgfältig?

Gabriel beschäftigte sich bei der Arbeit an seinem Programm YH vornehmlich mit der Frage, wie sich "konzentriertes Schreiben" ("deliberate writing") simulieren lässt (Gabriel o.J.). YH erhält eine interne Repräsentation für den Code und die Kommentare eines LISP-Programms - erläutert wird YH mit dem Code für das "dutch national flag"-Problem, eine einfache Sortieraufgabe -, aus dieser wird dann eine Beschreibung des Programms als englischer Text gewonnen. YH ist objekt-orientiert implementiert als Ensemble von "Experten" für bestimmte Aufgaben. Manche Experten können sehr spezifisch sein, z.B. "weiß" der "array expert", auf welche Weise im Text über Felder gesprochen werden sollte, wenn sie als Datenstruktur in einem Programm verwendet werden. Andere sind zuständig für den Aufbau grammatischer Strukturen: "simple declarative sentence expert", "noun phrase expert", "adjective expert" usw.. Wieder andere - und dies ist kennzeichnend für Gabriels Vorgehen - sind dafür zuständig, die Syntaxbäume der Zwischenversionen des Textes kritisch zu "betrachten" und ggf. Verbesserungen vorzunehmen, z.B. aufeinanderfolgende Sätze mit gemeinsamem Subjekt oder gemeinsamem Prädikat zusammenzufassen. Dieser Prozeß des simulierten "Erneut lesen" und darauf folgender Modifikationen soll solange fortgeführt werden, bis keine wesentlichen Verbesserungen des Texts mehr zu erwarten sind. Eine von YH generierte Beschreibung der Datenstrukturen für das Sortierprogramm lautet (Stettenfels 1983):

The flag is represented by a 1-dimensional, 0-based array of n elements, FLAG1. There are three array markers, L, M, and R, standing for Left, Middle, and Right, resp. L and M are initialized to 0; R is initialized to n - 1.

10 Ein Ansatz zur Generierung mit PROLOG

Mellish beschreibt Ansätze und Vorarbeiten zu einem Programm, das englischsprachige Erklärungen von Plänen produzieren soll (Mellish o.J.). Er denkt langfristig an den Einsatz dieses Generators als Teil eines Hilfesystems für die POPLOG-Programmierungsumgebung. "Pläne" sind in diesem Zusammenhang Beschreibungen, mit welcher Sequenz von Aktionen von einem Anfangszustand aus ein gewünschter Endzustand erreichbar ist. Mellish exemplifiziert diesen Ansatz mit einer typischen Situation beim Programmieren: Ein Benutzer editiert einen File und will ihn auf die Platte retten. Da seine disk quota aber erschöpft ist, muß er, um dies tun zu können, vorher einige andere Files löschen. Die Schritte, die er dazu tun muß, soll ihm das Programm in Englisch erläutern. Mellish geht davon aus, daß ein Planungsprogramm eine interne Repräsentation des Plans konstruiert. Für diese Darstellung wird eine einfache "Sprache" für Zustände und Aktionen verwendet; Wissen des Programms über Vorbedingungen und Effekte von Aktionen wird in Form sog. STRIPS-Operatoren([3]) kodiert.

Für das jeweilige Anwendungsgebiet ist ein spezifisches Wörterbuch vorzusehen, in dem den primitiven Zuständen und Aktionen englische Verben und den Objekten Nominalphrasen zugeordnet werden. Für abgeschlossene oder aktuell laufende Aktionen

denkt Mellish an "rewrite rules", um deren zugeordnete Verben auf Normalformen zurückzuführen. Beispiele dafür:

```
rewrite(deleted(F),-done,delete(F)-).
rewrite(viewing(F),-doing,lookat(F)-).
```

Bei seinen Erklärungen soll der Generator das Wissen des Benutzers berücksichtigen. Als Ansatz hierzu schlägt Mellish vor, in "obviousness"-Klausen festzuhalten, was dem Benutzer offensichtlich sein dürfte: Aktionen, die nur für bestimmte Ziele nötig sind, und Resultate, die aus bestimmten Aktionen folgen. Dieses rudimentäre "Benutzermodell" wird dann verwendet, um aus der Repräsentation des Plans alle für den Hörer als bekannt voraussetzbaren Teile zu entfernen. Denn - so Mellish - :

"One of the central problems of natural language generation is deciding what to leave out".

Für das obige Beispielproblem des mangelnden Plattenplatzes gibt Mellish als Generierungsergebnis an ([4]):

"in order to write the first file you must have space on the disc. type ENTER_DIR and you will be looking at a directory listing. then mark the temporary file with a asterisk then type ENTER_DEL and the system will be displaying delete(temp). then type ENTER_DODCL and you will have deleteen the temporary file. then you will have space on the disk however in order to write the first file you must be looking at the first file. type ENTER_Q and you will not be looking at a directory listing. then you will be looking at the first file"

11 McDonalds MUMBLE

McDonalds MUMBLE ist das Generatorprogramm mit der längsten "Geschichte" und den unterschiedlichsten Anwendungen. Ausgangspunkt für die ersten Arbeiten an MUMBLE war ein praktisches Interesse: verschiedene KI-Programme mit jeweils spezifischen internen Strukturen sollten ihre Resultate in englischer Sprache präsentieren können. Die Inhalte, die vermittelt werden sollten, konnten diese Programme in ihrem jeweiligen konzeptuellen Vokabular an MUMBLE weitergeben, das dann für die "linguistischen" Aufgaben der Generierung zuständig war (Stettenfels 1983). Die erste lauffähige Version von MUMBLE soll Ende 1977 vorgelegen haben: bei dieser Anwendung wurden isolierte Formeln und komplette Beweise des Prädikatenkalküls als englische Texte wiedergegeben. In den darauffolgenden Jahren wurde mit weiteren Anwendungen experimentiert (McDonald 1983). Dazu gehörten:

- (1) Beschreiben der Komponenten eines semantischen Netzes, das als Menge binärer Assertionen gegeben war,
- (2) Generieren aus einer in Form von Frames vorliegenden Wissensbasis über Shakespeares "Macbeth",
- (3) Produzieren englischer Texte aufgrund der Ausgabe eines Programm zur Szenen-Analyse.

Dieser experimentellen Phase folgte eine Reimplementation: Es hatte sich gezeigt, daß nur ein Teil der in der ersten Implementierung angelegten Mechanismen tatsächlich benutzt worden war (Stettenfels 1983). Daraus zog McDonald die Konsequenz, die Definition einer minimalen "virtuellen Maschine" zu versuchen: *"The key to formalizing the virtual machine is to view Mumble as a transducer from a stream of specification expressions to a stream of*

text that is constrained to perform its operations in quasi-realtime, i.e. is not permitted to 'buffer-up' more than a bounded amount of planned but unspoken text nor to take more than linear time to arrive at its decisions" (Stettenfels 1983). Diese Randbedingungen waren durch psycholinguistische Ergebnisse über die menschliche Spracherzeugung motiviert; die "virtuelle Maschine" sollte auch als Simulationsmodell für diese Prozesse dienen. Wichtigste Konsequenz aus dem Anspruch der Generierung in Echtzeit: Die Produktion des Textes erfolgt inkrementell - Teile des Textes sind bereits geäußert, wenn weitere konzeptuelle Strukturen erst betrachtet werden - und einmal getroffene Entscheidungen sind auf allen Ebenen endgültig, denn: *"allowing backup could lead to polynomial runtimes"* (Stettenfels 1983). MUMBLE arbeitet mit zwei Repräsentationen: Als Eingabe erwartet wird eine "message", d.h. die Spezifikation der vom Text zu vermittelnden Information, ergänzt durch Angaben zur rhetorischen Organisation (z.B. Was soll hervorgehoben, was kontrastiert werden?). Bei der Realisierung wird daraus die linguistische Repräsentation der Oberflächenstruktur des Textes gewonnen: diese Baumstruktur wird in dem Maße ausgeweitet und ergänzt, in dem weitere Elemente der message verarbeitet werden. Tauchen in dieser Struktur englische Worte als Terminale auf, werden sie an eine morphologische Routine ("does not have to do much work in English", McDonald 1983) übergeben, deren Ergebnis dann in den Ausgabe-string eingefügt wird. (McDonald 1983) gibt u.a. das folgende Beispiel einer konstruierten message für die "Macbeth"-Anwendung:

```
(message1
(sequence
(macbeth (murder (duncan))) ; "murder-ma"
(macbeth (become (king))) ; "ma-become-king"
(lady-macbeth
(persuade (macbeth (action murder-ma)))) ; "persuade-ma"
(lady-macbeth (hq (ambitious)))) ; "ambitious-ma"
(time-frame (before-time-of-speech))
(focus (macbeth))
(ancillary-facts
((murder-ma (motive (ma-become-king))
(persuade-ma (purpose (cause (murder-ma))))))
```

Zur rhetorischen Strukturierung dienen die Angaben zur Abfolge der Informationen ("sequence"), zur Hauptperson ("focus") und die "ancillary-facts", mit denen Relationen zwischen den isolierten Informationen beschrieben werden. Der zugehörige Text lautet:

"Macbeth murdered Duncan in order to become king. He was persuaded to do it by Lady Macbeth who was ambitious."

Die jüngste Anwendung von MUMBLE erfolgt im COUNSELOR-Projekt, bei dem der Dialog mit einem Rechtsanwalt nachgebildet werden soll (McDonald, Pustejovsky 1985).

12 Generierung in HAM-ANS

Im Projekt HAM-RPM (Hamburger Redepartnermodell) wurde ein vielbeachtetes System implementiert, mit dem Dialoge über einen eingeschränkten Diskursbereich - z.B. über das Inventar eines Zimmers - in deutscher Sprache geführt werden konnten. Im Nachfolgeprojekt HAM-ANS (Hamburger Anwendungssystem) wurde dieses System auf verschiedene, bewußt unterschiedlich gewählte Dialogsituationen angewandt (Hoepfner et al. 1983):

- interessengesteuerter Dialog zwischen einem Hotelmanager und einem potentiellen Gast,

- Dialog über eine von einem Bildverarbeitungssystem analysierte Straßenverkehrsszene,
- Abfrage einer relationalen Datenbank mit Fischereidaten.

Die Generierungskomponente war in erster Linie für typische Dialogaufgaben ausgelegt (Wahlster 1982):

- Auf Objekte des Diskursbereichs mußte sprachlich verwiesen werden können. Mit Dialoghistorie und Gebietswissen wurde der Determinationstyp von Nominalphrasen bestimmt: Für bereits eingeführte oder aufgrund des Gebietswissens als existent voraussetzbare Objekte (z.B. das Bett, wenn von einem Hotelzimmer bereits gesprochen worden war) wurden definite NPs erzeugt; neue Objekte wurden meist als indefinite NPs eingeführt.
- Vor allem bei visuell präsenten Diskurswelten (z.B. Zimmerinventar) mußte versucht werden, die Objekte mit solchen Merkmalen zu benennen (z.B. der rote, gepolsterte Sessel), die dem Dialogpartner die Identifikation erleichterten und eine Verwechslung mit anderen Gegenständen möglichst ausschlossen.

Die Elliptifizierungskomponente von HAM-ANS versucht, möglichst knappe und dennoch verständliche Antworten zu generieren. Aus der semantischen Struktur für die Antwort werden zunächst alle Teile getilgt, die mit derjenigen der Eingabe übereinstimmen. Von dieser kürzestmöglichen Ellipse wird ausgegangen. Mit der für die Analyse entwickelten Komponente zur Ellipsenrekonstruktion wird die geplante Äußerung daraufhin überprüft, ob sie noch verständlich ist oder expliziter gemacht werden muß (feedback-loop).

13 SEMTEX - Generierung deutscher Zeitungstexte

Der an der Universität Stuttgart implementierte Textgenerator SEMTEX (Rösner 1986c) wird derzeit für verschiedene Anwendungen weiterentwickelt (z.B. zur Generierung von Texten, die die Schritte bei der Lösung einer geometrischen Konstruktionsaufgabe beschreiben). Seine erste Aufgabe war die Generierung deutscher Zeitungsmeldungen über den Arbeitsmarkt wie im folgenden Beispiel:

Geringfügige Reduzierung der Arbeitslosenzahl.

NURNBERG/BONN (cpa)

Die Zahl der Arbeitslosen in der Bundesrepublik Deutschland hat sich während des Oktober nur sehr wenig verringert. Sie ist von 2151600 auf 2148800 zurückgegangen. Die Arbeitslosenquote hatte Ende Oktober einen Wert von 8.6 Prozent. Sie hatte am Ende des Vergleichszeitraumes des Vorjahrs ebenfalls bei 8.6 Prozent gelegen. Regierungssprecher Ost bewertet die Verringerung der Arbeitslosenzahl positiv. Der stellvertretende DGB-Vorsitzende Muhr erklärt, daß der Rückgang der Zahl der Arbeitslosen nicht darüber hinwegtäuschen dürfe, daß sie jetzt unverändert unerträglich hoch sei.

Kern des Textgenerators SEMTEX ist ein für das SEMSYN-Projekt implementierter Generator (Rösner 1986a). Dieser erhielt als Eingabe semantische Strukturen - in Form einer erweiterten Kasusrahmennotation -, die aus Titeln japanischer

wissenschaftlicher Arbeiten durch einen Parser des ATLAS/II-Systems (FUJITSU) gewonnen wurden. Die Kombination von Analyse durch ATLAS/II und Generierung durch SEMSYN kann als erstes maschinelles Übersetzungssystem vom Japanischen ins Deutsche angesehen werden. Bei Design und Implementierung des Generators für SEMSYN wurde versucht, ein leicht erweiterbares und vielseitig einsetzbares System zu entwickeln. Rechnung getragen wurde diesen Vorgaben zum einen durch die bei der Implementierung gewählte objektorientierte Strategie, zum anderen dadurch, daß das Repertoire generierbarer sprachlicher Formen nicht auf Titelsprache (d.h. Nominalgruppen unterschiedlicher Komplexität) beschränkt blieb, sondern alternativ auch verschiedene Arten von Sätzen produziert werden konnten. Bei der Erweiterung zu SEMTEX (Rösner 1986b) werden nicht mehr lediglich isolierte Strukturen generiert, sondern mehrsätzliche zusammenhängende Texte. Ausgehend von Daten zum Arbeitsmarkt wird zunächst ein "Textplan" in Form einer Liste von semantischen Repräsentationen erstellt. Um solche Inhaltsangaben in akzeptable Zeitungstexte umwandeln zu können, mußte der Generator insbesondere um Mechanismen zur Einbeziehung kontextueller Information in seine Entscheidungen erweitert werden:

- der durch vorausgegangene Sätze gegebene Kontext steuert die Lexikalisierung semantischer Symbole, die Elision redundanter Information, die Entscheidung über Pronominalisierung und andere Arten der Referenz.
- der temporale Kontext (bestimmt durch Sprechzeitpunkt und Referenzzeitraum) dient in Verbindung mit temporaler Information in den semantischen Strukturen zur dynamischen Bestimmung der grammatischen Zeit sowie zur Generierung geeigneter Benennungen für Zeitangaben.

Die von SEMTEX generierten Zeitungsmeldungen können neben faktischen Aussagen auch simulierte Kommentare von Politikern enthalten. Dazu wurden verschiedene abstrakte rhetorische Schemata, die solchen Äußerungen häufig zugrunde liegen, formalisiert und Wissen über deren Realisierungsmöglichkeiten in den Generator einbezogen.

14 Nachtrag

Die Rohversion dieses Manuskripts wurde im Frühsommer 1986 erstellt. Mittlerweile wurde mir eine Fülle neuerer Arbeiten über Generierung zugänglich, vor allem durch den 3rd International Workshop on Language Generation in Nijmegen (19. bis 23. August 1986). Die hier vorgelegte Übersicht wird dadurch nicht entwertet, wird aber mehr und mehr zur "Retrospektive". Eine Übersicht zu den aktuellen Entwicklungen in der Sprachgenerierung sollte daher in einer der nächsten Ausgaben des LDV-Forums folgen.

- [1] Ihm folgten weitere: 1984 in Stanford, 1986 in Nijmegen. Der 4. Internationale Workshop über Generierung ist für 1988 wieder in den USA geplant.
- [2] Die Aufzählung kann keinen Anspruch auf Vollständigkeit erheben. Aufgenommen wurden vor allem Systeme, die für die Entwicklung des Gebiets besonders relevant waren und/oder von deren Leistungsumfang der Autor sich - bei Vorführungen oder eigenen Experimenten - überzeugen konnte.

- [3] STRIPS-Operatoren beschreiben mögliche Aktionen auf einer Wissensbasis aus Propositionen durch drei Listen: die als Vorbedingung benötigten Propositionen, die nach Durchführung zu löschenden und die neu hinzukommenden; vgl. z.B. (Winston 1977).
- [4] Das Textbeispiel wurde einschließlich der auf mangelhafte Morphologiebehandlung zurückzuführenden "Fehler" - z.B. "deleteen" - übernommen.

Literaturübersicht: Textgenerierung

Busemann, S. 1982: Probleme der automatischen Generierung deutscher Sprache. HAM-ANS Memo 8, Universität Hamburg, 1982

Emele, M.; Momma, S. 1985: SUTRA-S - Erweiterungen eines Generator- Front-End für das SEMSYN-Projekt. Studienarbeit, Institut für Informatik, Uni Stuttgart, 1985

Gabriel, R.P. o.J.: Deliberate Writing. Stanford University, Computer Science Department, o.J.

Goldman, N. 1975: Conceptual Generation. In: Schank, R.C. Conceptual Information Processing, North-Holland, 1975

Hoenkamp, E. 1983: Een Computermodel van de Spreker: Psychologische en Linguistische Aspecten. Dissertation (holl., einige Kapitel in Englisch), Universität Nijmegen, 1983

Hoepfner et al. 1983: (Hoepfner, Christaller, Marburger, Morik, Nebel, O'Leary, Wahlster): Beyond Domain-Independence: Experience with the Development of a German Language Access System to Highly Diverse Background Systems. In: COLING-83, Proceedings, Stanford, 1983

Horacek, H. 1983 Generierung im System VIE-LANG: Linguistischer Teil. Universität Wien, Bericht 83-04, 1983

Ishizaki, Sh. 1983: Generation of Japanese Sentences from Conceptual Representation. In: IJCAI 83, Proceedings, Karlsruhe, 1983

Jacobs, P.S. 1983: Generation in a Natural Language Interface. In: IJCAI 83, Proceedings, Karlsruhe, 1983

Jacobs, P.S. 1984: PHRED: A Generator for Natural Language Interfaces. Berkeley, Computer Science Division (EECS), Report No. UCB/CSD 84/189, Juni 1984

Mann, W.; Moore, J. 1980: Computer as Author - Results and Prospects. ISI Research Report 79-82, 1980

Mann, W. et al. 1981: Text Generation: The State of the Art and the Literature. ISI Research Report 81-101, 1981

Mann, W. 1982: Two Discourse Generators. ISI Research Report 82-102, 1982

Mann, W. 1983: An Overview of the Penman Text Generation System. ISI Research Report 83-114, 1983

Matthiessen, C. 1982: A Grammar and a Lexicon for a Text-Production System. ISI Research Report 82-102, 1982

Mauldin, M.L. 1984: Semantic Rule Based Text Generation. In: COLING 84, Proceedings, Stanford, 1984

McDonald, D.D. 1983: Natural Language Generation as a Computational Problem: an Introduction. In: Brady & Berwick (Eds.) Computational Models of Discourse, MIT-Press, 1983

McDonald, D.D.; Pustejovsky, J.D. 1985: Description-Directed Natural Language Generation. In: IJCAI 85, Proceedings, Los Angeles, 1985

McKeown, K.R. 1983: Focus constraints on language generation. In: IJCAI 83, Proceedings, Karlsruhe, 1983

Mellish, Ch. o.J.: Natural Language Generation from Plans. University of Sussex, Cognitive Science Research Paper, CSRP 031, o.J.

Ritchie, G. 1984: A Rational Reconstruction of the PROTEUS Sentence Planner. In: COLING 84, Proceedings, Stanford, 1984

Rösner, D. 1986a: SEMSYN - Wissensquellen und Strategien bei der Generierung von Deutsch aus einer semantischen Repräsentation. In: Batori & Weber (Eds.) Neue Ansätze in Maschinellem Sprachübersetzung: Wissensrepräsentation und Textbezug, Niemeyer Verlag, Tübingen, April 1986

Rösner, D. 1986b: Von Titeln zu Texten - Zur Entwicklung des Textgenerators SEMTEX. Erscheint in: GLDV-Jahrestagung 1986, Göttingen, (in Vorb.)

Rösner, D. 1986c: Ein System zur Generierung von deutschen Texten aus semantischen Repräsentationen. Dissertation, Institut für Informatik, Universität Stuttgart, 1986

Seligman, M. 1985: A Bibliography of Natural Language Generation. Berkeley, Manuskript, Nov. 1985

Stettenfels 1983: Rösner, D. (Ed.) "International Workshop on Language Generation - Documentation. Institut für Informatik, Universität Stuttgart, 1983

Swartout, W.R. 1981: Producing Explanations and Justifications of Expert Consulting Programs. MIT, Laboratory for Computer Science, MIT/LCS/TR-251, Januar 1981

Wahlster, W. 1982: Natürlichsprachliche Systeme: Eine Einführung in die sprachorientierte KI-Forschung. Universität Hamburg, Bericht GEN-10, 1982

Yazdani, M. 1982: How to write a story. In: European Conference on Artificial Intelligence (ECAI-82), Proceedings, Paris, 1982

Yazdani, M. 1983: Generating events in a fictional world of stories. Univ. of Exeter, Dep. of Computer Science, Research Report R-113, 1983

ZUR PROBLEMATIK LANG- UND KURZFRISTIGER PROJEKTE IM BEREICH DER COMPUTERLINGUISTIK

Manfred Thiel

Sonderforschungsbereich 100 /A2
Universität des Saarlandes
6600 Saarbrücken

Phänomene wie Theorien oder Systeme besitzen eine gewisse Lebensdauer und manche weisen auch in diesen Zeitgrenzen eine zyklische Entwicklung auf. Nicht nur auf dem Gebiet der LDV sind diese Zyklen eng mit dem Begriff des Redesign verbunden. Im folgenden soll auf die Problematik hingewiesen werden, die bei einer phasenverschobenen zyklischen Entwicklung von linguistischen Theorien, computerlinguistischen und softwaretechnologischen Systemkonzeptionen entstehen kann. Diese Überlegungen sollen also versuchen, die Probleme aufzuzeigen; es wird nicht die Absicht verfolgt, Lösungen anzubieten, höchstens eine: Erhaltung der Pluralität der Verhaltensalternativen von Projekten. Entsprechend abrupt endet der Beitrag.

Hat ein computerlinguistisches Projekt Zeit genug, um bestimmten Fragen auf den Grund zu gehen, entsteht zwangsläufig das Problem, daß es eigentlich alle 2-5 Jahre die linguistischen und/oder informatischen Grundlagen neu überdenken und Systeme neu konzipieren müßte. Tut man dies in einem langfristigen Projekt nicht, so ist es nach einem derartigen Zyklus

einerseits selten in der vordersten Reihe der Forschung, wenn es darum geht, z.B. neue linguistische oder informatische Konzepte zu realisieren (dies bedeutet nicht: nicht zu perzipieren!).

andererseits aber ist es ziemlich einfach, innerhalb von 2 oder 4 Jahren zu zeigen, daß z.B. eine linguistische Theorie in der Computerlinguistik anwendbar ist und zumindest einen Teil der Sprache beschreibt. Es bleiben aber immer Fälle offen, die man in dem relativ kurzen Zeitraum nicht erreicht. Z.B.: wieviele Jahre muß man an einem Parser arbeiten, bis er stabil und linguistisch derart mächtig ist, daß man damit einigermaßen sicher juristische Texte analysieren kann, in denen Sätze vorkommen, die länger als 100 Wörter sind? Wie verhält sich eine linguistische Theorie, wenn es darum geht, sog. "Ausnahmen" analysieren zu können, wobei "Ausnahmen" immer in Bezug auf das implementierte linguistische Modell zu verstehen ist? So können Regeln nicht bis zu einer filigranten Exaktheit im Detail restringiert werden, ohne daß ihre Allgemeingültigkeit verloren geht. Es hat sich außerdem gezeigt, daß die letzten Prozente der zu lösenden Probleme einen immer höheren Aufwand an Zeit und Arbeitseinsatz benötigen.

Daraus lassen sich folgende Thesen ableiten:

1. In unregelmäßigen Abständen entstehen neue theoretische Konzepte.
2. Ein System sollte nach einer gewissen Zeit einem Redesign oder einer Neukonzeption unterworfen werden, um den Anschluß an neue theoretische Konzepte nicht zu verlieren.
3. Nur wenn man den für ein System gewählten theoretischen Ansatz lange und intensiv genug verfolgt, können grundlegende Erkenntnisse, wie sie oben angedeutet wurden, gewonnen werden.

Der Widerspruch zwischen der 2. und 3. These könnte so mancher linguistischen und informatischen Konzeption Schwierigkeiten bereiten. Bei der Konzeption eines langfristigen Projektes muß also großer Wert auf die Planung der projektinternen Lebensdauer der gewählten theoretischen Grundlagen gelegt werden. Es ist dabei davon auszugehen, daß viele (d.h. thematisch anspruchsvolle) computerlinguistische Projekte eine verhältnismäßig langfristige Angelegenheit sein **müssen**. Andererseits darf man aber nicht vergessen, daß der Aufwand, die zu einem beliebigen Zeitpunkt eines Projektes übriggebliebenen Fragen zu lösen, überproportional steigt. So stellt sich für eine linguistische Theorie in einem sprachverarbeitenden System irgendwann das Problem, daß die Eleganz ihrer Konstruktion in Gefahr gerät, von Notlösungen und angebauten Balkönchen überwuchert zu werden. Lohnt es sich dann noch, weiter zu machen, oder soll man sich einer neuen Theorie zuwenden? Geschieht mit dieser dann das gleiche? Der Widerspruch liegt auch darin, daß diese Balkönchen gebaut bzw. ihre Ursachen erst einmal entdeckt werden müssen, um die Grenzen einer Konzeption abstecken und neue Ansätze aus diesen Erfahrungen aufbauen zu können. Ein sehr einfaches, aber doch typisches Beispiel für das oft beträchtliche und manchmal auch unübersehbare Wachstum eines Systems auch für den Fall, daß nur ein triviales Problem gelöst werden soll, ist folgendes: Zur Unterstützung der Wörterbuchkodierung wird eine Regel benutzt, die folgenden Inhalt hat:

wenn ein Substantiv auf das Suffix 'ung' endet, dann ist sein Genus feminin.

Sofort fallen uns die problematischen Fälle ein: "Dung", "Sprung" usw., also Beispiele, in denen 'ung' nicht Suffix ist. Über die eigentliche linguistische Regel hinaus muß sich der Computerlinguist also noch etwas einfallen lassen. Es bieten sich dabei verschiedene Alternativen an:

1. Die Graphemfolge 'ung' wird gegebenenfalls als Suffix präkodierte. Damit ist aber prinzipiell nichts gewonnen: ob man kodiert: "ist feminin" oder: "ist Suffix" läuft auf das gleiche hinaus.
2. Die Wörter, bei denen 'ung' kein Suffix ist, müssen vorher alle im Wörterbuch stehen. Dies ist eine Notlösung und verursacht ein wenig wissenschaftliche Bauchschmerzen.
3. Eine morphologische Analyse stellt dagegen eine adäquate Vorgehensweise dar.

Allerdings tritt hier der diskutierte Effekt ein: die morphologische Analyse schafft selbst wieder neue Probleme, die, genau wie das hier diskutierte, zu ihrer Lösung wieder anderen, möglicherweise

ähnlich großen oder noch größeren Aufwand erfordern. Sollte die Computerlinguistik als Vorlage für ein Bild von Escher herangezogen werden können? Dies scheint nicht der Fall zu sein: Es entstehen zwar neue Komponenten, die hochgradig voneinander abhängig sind, aber es ist zu erwarten, daß sich der Kreis irgendwann schließt und keine neuen Komponenten hinzukommen. Allerdings besteht immer die Gefahr, daß die Veränderung einer Komponente aufgrund der Vernetzung Auswirkungen auf andere haben kann.

Folgende Schlußfolgerungen sind daraus zu ziehen:

1. Die meisten Probleme in der Computerlinguistik sind nicht trivial.
2. Computerlinguistische Systeme sind von Natur aus sehr komplex; die Aussagekraft von Systemen, die sich auf einen Ausschnitt

der gesamten Problematik beschränken, ist mit äußerster Vorsicht zu bewerten. Es scheint oft so etwas wie ein "Alles-oder-nichts-Effekt" einzutreten.

3. Es ist notwendig, zur Unterstützung der Entwicklung von sprachverarbeitenden Systemen Werkzeuge zu besitzen.

Steht ein langfristiges Projekt diese Durststrecke des Beharrens auf einem gewählten Ansatz durch, um danach sagen zu können, welche Anforderungen eine linguistische Theorie oder eine Softwarekonzeption erfüllen muß, ist dies aufgrund der breiten Variation an Grammatikalität und Akzeptanz in der natürlichen Sprache zugegebenermaßen recht undankbar. Es soll aber auch nicht die Problematik vergessen werden: Derartige "praktische" Arbeiten müssen irgendwann einmal auch in wissenschaftliche Neuansätze münden.

Anzeige

Sprache und Information S+I

Thomas Herrmann Zur Gestaltung der Mensch-Computer-Interaktion: Systemerklärung als kommunikatives Problem

Ca. 280 Seiten. Kart. ca. DM 108,-. ISBN 3-484-31914-3 (Band 14)

Die Untersuchung orientiert sich an solchen Computerbenutzern, die mit dem Rechner nicht nur Routineaufgaben erledigen, sondern abwechselnden Problemstellungen gegenüberstehen, die einen kreativen Lösungsprozeß erfordern. Der Kenntnisstand solcher Benutzer ändert sich permanent während der Mensch-Computer-Interaktion. Für diese Benutzungsweise werden auch Erklärungen während des Dialogs notwendig. Unter diesem Blickwinkel wird untersucht, welche Leistung durch systemgebundene Selbsterklärung über EDV vermittelt werden kann. Das breite Spektrum möglicher Erklärungen wird hinsichtlich der Erklärungsinhalte sowie der Orientierung und der Form ihrer Darstellung differenziert. Diese Aspekte werden in einem Modell der Mensch-Computer-Interaktion zusammengefaßt, das sich am Computerbild des (Be-)Nutzers orientiert (CombiNo-Modell der M-C-I). Insbesondere enthalten Erklärungen zu EDV-Systemen auch konzeptuelles (im Gegensatz zu ereignis-orientiertem) Wissen, dessen kommunikativ-adäquate Vermittlung vom Kontext der Erklärungssituation relativ unabhängig ist und einen argumentativen Dialog benötigt – es findet sog. kruziale Kommunikation statt. Sie ist nach heutigem Stand der Sprechakt- und Argumentationstheorie nicht im Rahmen der Computerbenutzung technisch zu realisieren, wenn man ein unangemessenes Ausmaß unerkannter Mißverständnisse vermeiden will. »Kruzial« ist eine Eigenschaft, die als Abgrenzungskriterium bei der Gestaltung selbsterklärungsfähiger Systeme dienen kann.

Die Auseinandersetzung wird mit solchen Ansätzen geführt, die davon ausgehen, daß der Mensch kommunikative Erwartungen an die Mensch-Computer-Interaktion heranträgt. Dem wird ein Vorschlag gegenübergestellt, wie man Benutzer unterstützen kann, wenn sie sich bei fehlendem Wissen oder bei Störungen selbst helfen möchten: das Konzept intervenierender Benutzbarkeit.

Rainer Kuhlen

in Zusammenarbeit mit UDO HAHN und ULRICH REIMER

Informationslinguistik

Theoretische, experimentelle, curriculare und prognostische Aspekte einer informationswissenschaftlichen Teildisziplin

Ca. 250 Seiten. Kart. ca. DM 74,-. ISBN 3-484-31915-1 (Band 15)

Der Band informiert mit seinen 8 Beiträgen (4 davon sind in Englisch geschrieben) über die Entwicklung der Informationslinguistik, wie sie im Bereich Informationswissenschaft an der Universität Konstanz betrieben wird. Informationslinguistik ist keine neue wissenschaftliche Disziplin, sondern wird als zentrale Teildisziplin der Informationswissenschaft angesehen. Sofern diese experimentell arbeitet, befaßt sie sich weitgehend mit Fragen des Information Retrieval. Entsprechend der allgemeinen Forschungsentwicklung vom Referenz-/Dokument-Retrieval zu intelligenten Systemen hat auch die Informationslinguistik den Übergang von mehr quantitativen, strukturalistischen Verfahren zu wissensbasierten Verfahren der Künstlichen-Intelligenz-Forschung vollzogen.

Der Band enthält die folgenden Beiträge: R. KUHLEN, Informationslinguistik. – R. KUHLEN, Quantitative Probleme beim induktiven Wörterbuchaufbau. – R. KUHLEN, Linguistic aspects of information retrieval. – U. HAHN, Curriculum Informationslinguistik. – R. KUHLEN, Similarities and difference in intellectual and machine text understanding and condensation. – U. HAHN/U. REIMER, TOPIC – A text understanding and condensation system. – R. KUHLEN, Development of information linguistics. Results of an international Delphi poll. – U. HAHN/U. REIMER, »State of the art« der Volltextanalyse und Textkondensierung.

Niemeyer

THESEN ZUR KI-ORIENTIERTEN SPRACHFORSCHUNG aus der Sicht der Verarbeitung von Massentextdaten im Information Retrieval

These 1

Die KI-orientierte Sprachforschung könnte im Bereich des Information Retrieval in folgendes Dilemma geraten: Entweder sie beschäftigt sich praxisfern mit sehr begrenzten Bereichen und verschiebt großflächige Anwendungen in eine ferne Zukunft - oder sie wendet sich hauptsächlich der Grobanwendung zu und ist nicht mehr die reine Lehre, als die sie angetreten ist. Ergebnis einer solchen Praxisorientierung könnte dann z.B. ein trickreich-ingenieurhaftes Operieren mit teilweise durchaus traditionellen linguistischen Techniken in klassischen Programmierumgebungen sein - auf alle Fälle keine KI.

These 2

Im Bereich der Verarbeitung großer Textmengen bieten sich oft nur in begrenzten Bereichen Techniken der KI-orientierten Sprachverarbeitung an. Root-functions (zur Wortformen-Expandierung) in kommerziellen Information-Retrieval-Systemen sind ohne KI entstanden. Lexikonarme, robuste und sehr schnelle Wortreduktionsalgorithmen und partielles Parsing für Freitext (zu Indexierungs-, Abstracting- oder Klassifikationszwecken) sind keine Erfindungen der KI: Seit 3 Jahrzehnten wird an ihnen gearbeitet.

These 3

Unbeschadet ihres nicht KI-spezifischen Ansatzes gelten derartige Verfahren heute oft als "KI". Daß dies so ist, liegt auch an den Entwicklern dieser Verfahren, die nur zum Teil auf Distanz zur KI gehen. Es liegt generell an der Tendenz der DV-orientierten Linguistik, sich ihre Existenzberechtigung derzeit vor allem aus der KI zu holen, statt sich auf ihre Eigenständigkeit und, im industriellen Rahmen, auf die hohe Praxisrelevanz ihrer Verfahren im Bereich der Massentextdatenverarbeitung zu besinnen.

Christoph Schwarz, Siemens AG, ZTIINF 313, 8 München 83, Otto-Hahn-Ring 6

EXPERT- SYSTEM- SHELLS LISP PROLOG

Wir führen die professionellen Entwicklungswerkzeuge für Anwendungen in der Wissensverarbeitung.

Wir beraten Sie bei der Auswahl von geeigneten Expert-System-Shell, LISP oder PROLOG-Systemen, und bieten Ihnen einen qualifizierten Produktservice.

Wir unterstützen Sie bei der Entwicklung eigener Expertensysteme auf Ihrer vorhandenen konventionellen Hardware und schulen Sie im Umgang mit der Expertensystemtechnologie.

Die Aufwärtskompatibilität unserer Produkte - von PC bis zum MAINFRAME - garantiert Ihnen Flexibilität in der Zukunft.

Fordern Sie ausführliche Informationen an:



Brainware

Ihr Experte in Expertensystemen
Consulting · Schulung · Software

Brainware GmbH & Co. · Kirchgasse 24
6200 Wiesbaden · Tel. 061 21-37 20 11
Teletex 176121833

THESEN ZUR SPRACHORIENTIERTEN KI-FORSCHUNG

These 1

Sprachorientierte KI-Forschung bedeutet, daß die Entwicklung von künstlichen Sprachen sich am "Modell" der natürlichen Sprache orientiert. Das klingt zwar trivial, aber die Praxis zeigt, daß meistens eine entgegengesetzte Entwicklung betrieben wird.

These 2

Die Chance (!) der sprachorientierten KI-Forschung besteht darin, das instrumentelle Paradigma der Sprache an die Quelle seiner Flexibilität, nämlich den intersubjektiven und geschichtlichen Sprachgebrauch, zu binden.

These 3

Das Öffnen des instrumentellen Paradigmas der Sprache gegenüber der gesellschaftlichen und geschichtlichen Einrichtung der natürlichen Sprache erfordert Bescheidenheit und Kreativität zugleich. Nicht "titanische" KI-Projekte sind gefragt, sondern "sanfte" bzw. "schonende" (vgl. die Parallele: *Ausbeutung* der Natur durch Technik vs. *ökologisch-angepaßte* Technik).

These 4

Ein Extremfall der Instrumentalisierung der Sprache zuungunsten der sprachorientierten KI-Forschung dürfte die militärische Anwendung darstellen, wo man also von einem zu zerstörenden gesellschaftlichen Zusammenhang ausgeht. (vgl. H.-D. Lutz, LDV-Forum 3(1965)2: S. 77-79)

These 5

Sprachorientierte KI-Systeme müssen (wollen sie "benutzerfreundlich" sein) vielfältige "Sicherungen" einbauen. Sowohl die Sprachregelung als auch die Wissensbasis sind nämlich "tentativ". Man sollte die Grenzen nicht verschleiern sondern eher offenbaren (etwa nach dem Motto: Lieber den Benutzer frustrieren als ihn in die Irre führen...). (Vgl. A. Fauser, LDV-Forum 3(1985)2: S. 80-81, These 2).

These 6

Eine an Texten orientierte KI-Forschung darf die hermeneutische Einbettung des Verstehensvorgangs in einen (letztlich unerreichbaren) sich "horizontalartig" verschiebenden sozialen Kontext nicht aus den Augen verlieren, will sie sich an dem "höherwertigen" Phänomen der (möglichen) Produktivkraft eines Textes (als Quelle von Interpretationen) orientieren (vgl. U. Hahn, LDV-Forum 3(1985)2: S. 82-87).

These 7

Die geglückte Zukunft der sprachorientierten KI-Forschung hängt davon ab, ob sie sich in das Paradigma der Industriegesellschaft eingliedert, oder ob die Chance wahrgenommen wird, diese Technik "von Anfang an" im Sinne eines "schonenden" Umgangs mit ihrem "Gegenstand", nämlich der Sprache, zu gestalten (vgl. K. Morik, LDV-Forum 3(1985)2: 88-89).

Prof. Dr. phil. R. Capuro, Fachhochschule für Bibliothekswesen Stuttgart

**Aufnahmeantrag für persönliche Mitgliedschaft**

Ich beantrage, mich in die Gesellschaft für Linguistische Datenverarbeitung e. V. (GLDV) als Mitglied aufzunehmen.

Name: Vorname: Student: ja/nein

Akadem. Grad oder Berufsbezeichnung: beschäftigt bei:

Interessengebiete:

Arbeitsschwerpunkte:

Datum: Unterschrift:

PHONETISCHE BEITRÄGE ZUR AUTOMATISCHEN MASCHINELLEN SPRACHERKENNUNG

In Absprache mit dem Arbeitskreis Spracherkennung, Sprachsynthese, Phonetische Datenbanken wird im LDV-Forum eine Beitragsreihe eingerichtet, in der über aktuelle Forschungsergebnisse berichtet werden soll. Wissenschaftler, die unter dem Titel **Phonetische Beiträge zur automatischen maschinellen Spracherkennung** veröffentlichen wollen, sind herzlich aufgefordert, Manuskripte, nicht umfangreicher als ca. 6 Schreibmaschinenseiten, bei der Redaktion einzureichen. Die Beschränkung im Umfang soll sicherstellen, daß die Reihe ihre Aufgabe als aktuelle und offene Plattform erfüllen kann.

Bei dem folgenden Beitrag von Klaus Günther und Christopher Schweisthal handelt es sich um den ersten der neuen Reihe.

NACHWEIS VON KATEGORIALER IDENTIFIKATION DURCH ADAQUATE REPRODUKTION VON NATÜRLICHSPRACHLICHEN SIGNALEN IN PHONETISCHEN HÖRTESTS

Klaus Günther Schweisthal und Christopher Schweisthal

Institut für Phonetik und Sprachliche Kommunikation
an der Universität München
Schellingstr. 3
D-8000 München 40

Das für die bisherigen Verfahren zur automatischen Spracherkennung kaum lösbare Problem der notwendigen Zuordnung von physikalischen Meßdaten (akustische Produkte) zu einzelsprachlich geregelten Schriftformen (Buchstaben, Wörter, Sätze) wird experimentell durch einen phonetischen Ansatz behandelt. Als Lösung wird eine optische Zeichenerkennung auch fließender Rede vorgeschlagen, die auf Echtzeit-Analysen der Veränderung des Sprachsignals im Zeitbereich (Oszillogramm) beruht. Für die linguistische Datenverarbeitung erwächst die neue Aufgabe, die vom Computer ermittelten optischen Patterns der zeitlichen Veränderungen von relativen Luftdruckschwankungen einer sprachadäquaten symbolischen Repräsentation als Bindeglied zur heute geltenden einzelsprachlichen Orthographie zuzuordnen (Sprachdatenverarbeitung).

1. Einleitung

Die Verschriftlichung von Sprachen ist nicht immer sprecheradäquat gelungen. Trotzdem werden in phonetischen Reihenuntersuchungen häufig Urteile in Form von symbolischen Repräsentationen von Antworten auf die angebotenen Stimuli als Daten zugrundegelegt. Statistische Auswertungen der Tests ergeben oft wenig signifikante Ergebnisse. Auch im Rahmen der bisherigen Verfahren zur autonomen Spracherkennung stellen die notwendigen Zuordnungen von physikalischen Meßdaten (akustische Produkte) zu den einzelsprachigen Symbolen (Buchstaben = Wörter bzw. Sätze) ein schwer lösbares Problem dar.

Vielleicht liegt der Grund dafür in der Form der Daten: Aufbauend auf eine Anregung (H.G. Tillmanns 1980) und auf die Ergebnisse der Untersuchungen zum bayerischen Dialekt (Schweisthal 1984) wurden neben symbolisch repräsentierten Daten (Schrift, Sonderzeichen) erstmals Urteile in gesprochener Sprache als adäquate Reproduktionen der vorgegebenen Stimuli berücksichtigt. Die statistische Auswertung des Vergleichs der angewendeten Datenerhebungsmethoden ergab (Varianzanalyse) einen signifikanten ($p < 0.001$) Vorteil für die gesprochenen Daten.

2. Versuchreihe

Gängige phonetische Testverfahren untersuchen meist Produktionsweisen und Perception von Sprachsignalen getrennt. Bei vielen Perzeptionstests werden Hörurteile zur Identifikation linguistischer Kategorien in Form von schriftlichen Notatio-

nen als Daten erhoben. Die Ergebnisse sind oft unbefriedigend, d.h. die statistischen Auswertungen zeigen oft keine klaren Kategorientrennungen. Nach einem Vorschlag von H.G. Tillmann läßt sich kategoriale Identifikation im Rahmen seines Modells eines menschlichen Kommunikators durch adäquate Reproduktion vorgegebener sprachlicher Äußerungen ostensiv demonstrieren (Tillmann 1980). Eigene Erfahrungen im phonetischen Unterricht zeigen, daß beim Nachsprechen von vorgegebenen kategorialen Stimuli deutlich weniger Abweichungen vom Ausgangsstimuli zu verzeichnen sind als bei schriftlicher Notation. Zur Bestätigung dieser Beobachtungen wurden folgende drei Tests im Abstand von einer Woche mit 15 Studenten der Phonetik (12 ♀, 3 ♂) durchgeführt: 12 natürlichsprachlich produzierte Tonhöhenbewegungen vom Typus Level, Rise, Fall und Kombination davon wurden

1. Im Abstand von 2 Sekunden auf einem Testband präsentiert und waren vom Hörer unmittelbar nach der Präsentation im Zeitraum von 2 Sekunden nachzusprechen. Die Prozedur wurde auf Tonband zweikanalig aufgenommen.
2. Die gleichen Stimuli wurden den gleichen Studenten eine Woche später randomisiert dargeboten und waren in beliebiger Aufschreibeweise zu notieren.
3. Die gleichen Stimuli, erneut randomisiert waren wiederum eine Woche später von den gleichen Studenten in Form von folgenden Sonderzeichen: —, ↗, ↘, ↙, ↕, + Kombi aufzuzeichnen.

Die statistische Auswertung (VARAN-Paket) des Vergleichs der Häufigkeiten der Abweichungen vom vorgegebenen Stimulus aller 3 Versuchsreihen ergab ein hochsignifikantes Ergebnis: Die mündlichen Urteile wichen entschieden weniger häufig von den Tendenzen der vorgegebenen Tonhöhenbewegungen ab als die in schriftlicher Form notierten Urteile.

Im Fall der beiden Levelstimuli (—) waren selbst die Abweichungen des Sprechers der Ausgangsstimuli von einer gedachten Idealton-Kategorie gleicher Tonhöhe von vielen Testpersonen nachgeahmt worden.

3. Datenverarbeitung der akustischen Meßwerte

Die Tonbandaufnahmen der beiden Levelstimuli der Dauertonsprechsilbe [la:tS] (—) des vorsprechenden Autors des Transkriptionskurses, (Smalley 1968), und die entsprechenden Antwortstimuli dreier weiblicher Versuchspersonen (VP 1-3) wurden mit 20 KHz (sample rate) digitalisiert und mit dem institutseigenen Programmpaket TISYS (Autoren: Tillmann, Kotten) auf einer PDP 11/50 in einzelne aufeinanderfolgende Stimmtonperioden segmentiert. Ein Rechenprogramm ermittelte aus den digitalisierten Periodenlängen (in Millisekunden) die jeweiligen Werte in Hertz und druckten diese in Tabellenform aus.

Die Auswahl der digitalisierten Beispiele der Versuchspersonen geschah nach Gehör und Oszillomink-Diagrammen.

4. Anmerkungen zur Datenrepräsentation (Beispiele I und II)

Bei der Darstellungsweise der Hertz-Zeit-Diagramme muß berücksichtigt werden, daß zugunsten einer besseren Vergleichbarkeit der dargestellten Levelverläufe folgende Maßstabsangleichungen vorgenommen wurden:

1. Beim vorliegenden Ausgangstimulus (Smalley) repräsentiert jeder Kurvenpunkt den Meßwert in Hz nach der Formel:

Summe zweier aufeinanderfolgender Stimmtonperiodenwerte in Hertz

2

Hierdurch wird der Signalverlauf von I: 512 ms Dauer und 58 Stimmtonperioden und II: 483 ms Dauer und 54 Stimmtonperioden auf eine mit den für diese Datenrepräsentation ausgewählten weiblichen Versuchspersonen (VP 1-3) vergleichbare Länge gebracht.

2. Bei den vorliegenden Antwortstimuli (Frauenstimmen!) repräsentiert in der Regel jeder Kurvenpunkt den Meßwert in Hertz nach der Formel:

Summe dreier aufeinanderfolgender Stimmtonperioden in Hertz

3

Diese Maßnahme wird dadurch begründet, daß die höhere Stimmlage der Frau auch eine entsprechend höhere Anzahl von Stimmtonperioden beim Vergleich zum Ausgangsstimulus leicht kürzeren Signalverlauf der Antwortstimuli produziert.

3. Die extremere Ausprägung der Diagramme der weiblichen VP's ist in Abhängigkeit von der höheren Stimmlage durch die Hertz-Skalierung bedingt. Die relativen Tonhöhen-

schwankungen sind für den Hörer in etwa gleich groß. Daher wurden zusätzlich die der Hertz-Skalierung entsprechenden Tonstufen der Sprecher nach der temperierten Stimmung markiert.

5. Ergebnisse und Schlußfolgerung

1. Der optische Vergleich der adäquaten Reproduktionen eines vorgegebenen silbischen Levelstimulus bestätigt die vom Ohr längst getroffene Entscheidung: Selbst kleine Abweichungen der menschlichen Stimme bei der Realisierung einer gedachten Idealtonkategorie (Level-Aufgabenstellung durch Smalley) werden als registriert dokumentiert in adäquaten Abweichungen beim Nachsprechen dieser vorproduzierten Idealtonkategorie.
2. Diese optische Bestätigung von Vorgängen der Alltagskommunikation ist erst durch die verwendete Computerprogrammierung ermöglicht worden. Erst durch sie ist die optisch wahrnehmbare Dimension eines sehr rasch in der Zeit ablaufenden natürlichsprachigen Ereignisses als Quasi-Zeitlupe und sogar quasi-statische Bildaufzeichnung ermöglicht worden.
3. Mit den hochentwickelten Methoden der automatischen Bilderkennung wird in Zukunft in Verbindung mit den phonetischen Forschungsergebnissen zur optisch zugänglichen Realisierung von Sprachlautkategorien auch komplizierter Artikulationsvorgänge durch Analysen von Veränderungen im Zeitbereich eine automatische Spracherkennung auch fließender Rede auf unterster (physikalischer) Ebene ermöglicht.
4. Für die linguistische Datenverarbeitung erwächst die neue Aufgabe, die vom Computer in Realzeit ermittelten optischen Patterns der zeitlichen Veränderungen von relativen Luftdruckschwankungen einer sinnvollen sprachadäquaten Verschriftungsbrücke als Bindeglied zur heute geltenden einzelsprachlichen Schreibkonvention zuzuführen.

Bei dieser Untersuchung haben uns geholfen:

Stefan Forster, der die Periodensegmentierung gemacht hat; Lisa Schiefer, die die Oszillomink-Diagramme filtermäßig betreute; Bobby Pompino-Marschall, der die VARAN-Statistik erstellt und Till Schweisthal, der die Zeichnungen realisiert hat. Ganz besonders danken wir den Autoren des TISYS-Programmpakets, H.G. Tillmann und Kurt Kotten.

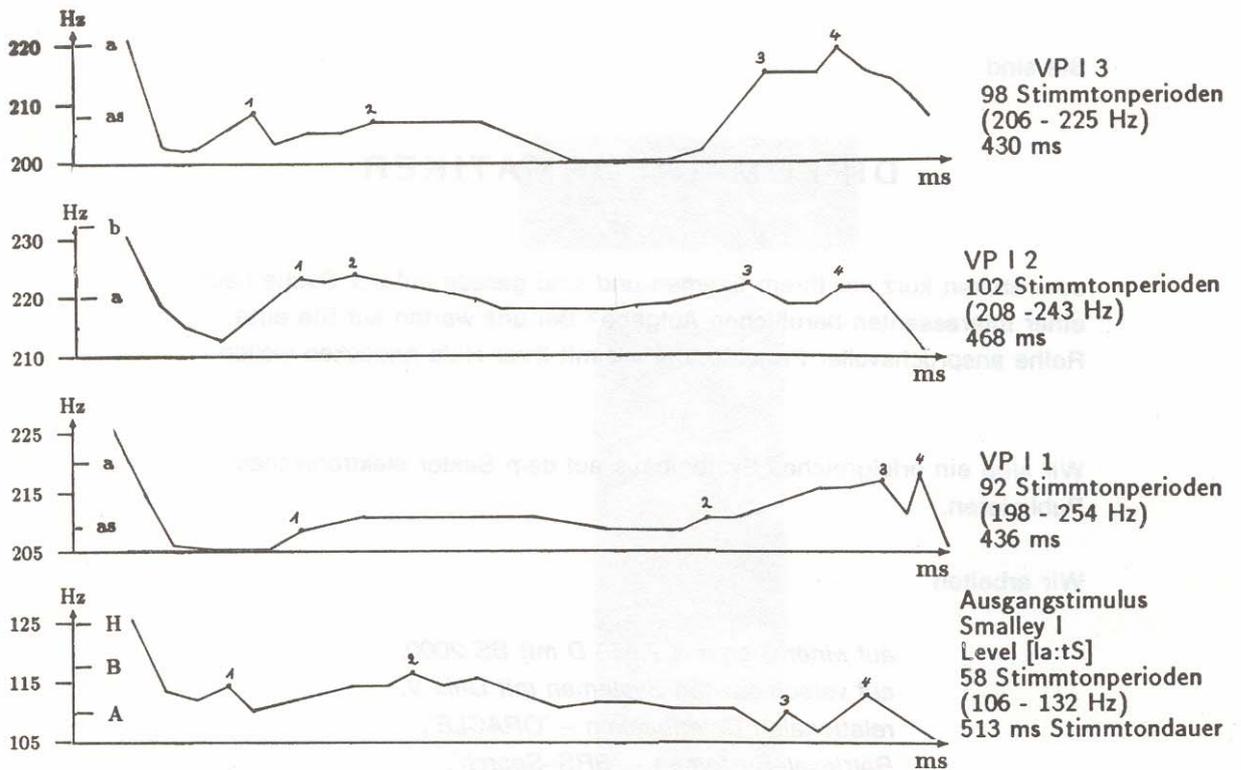
Literatur

Tillmann, H.G. 1980: Tillmann mit Mansell, *Phonetik*, Stuttgart 1980, 43-45.

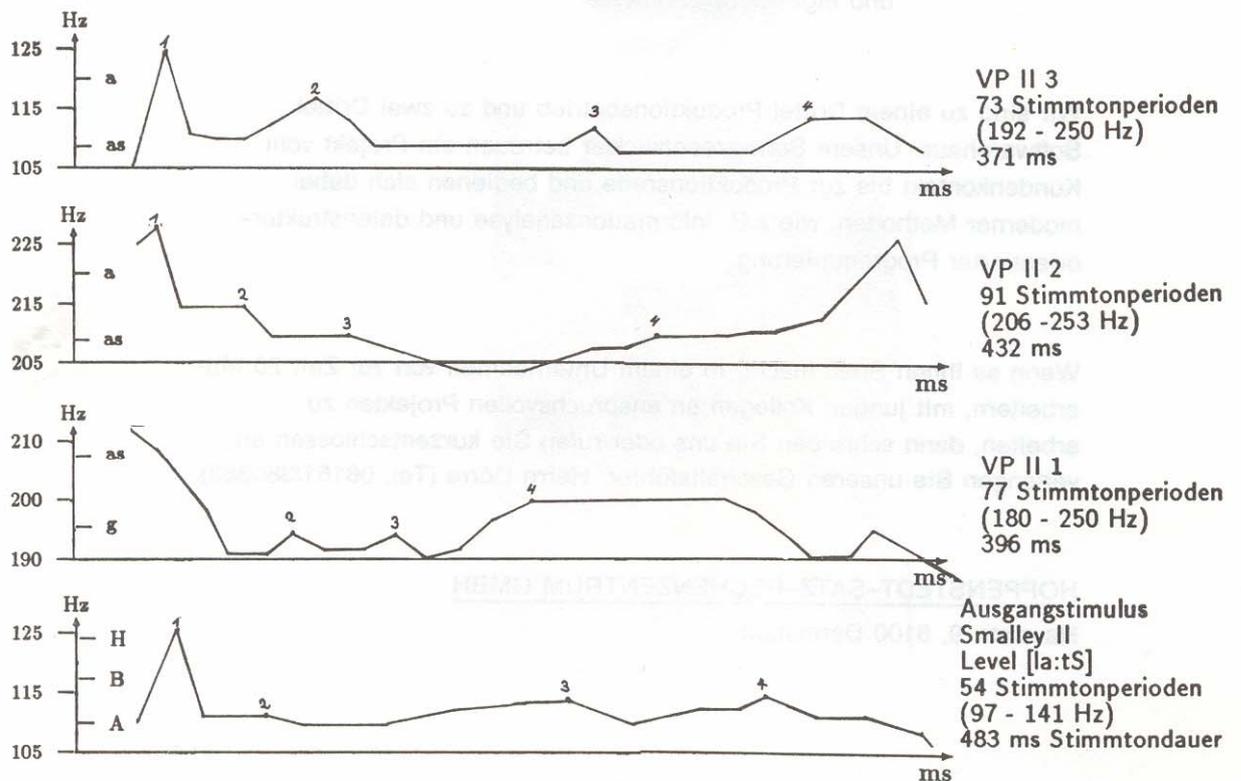
Schweisthal; Forstner; Kotten 1984: *Bavarian-a-Vowels*. In: Proc. 10th Int. Congr. Phon. Sci. Utecht 1984, 210-214.

Smalley, William 1968: *Manual of Articulatory Phonetics*, Tarrytown, N.Y. 1968.

Beispiele für die adäquate Reproduktion von charakteristischen Tonhöenschwankungen einer vorgesprochenen Dauertonsilbe durch nachsprechende Versuchspersonen (I)
 TISYS-Programm - PDP 11/50/ Institut für Phonetik, München



Beispiele für die adäquate Reproduktion von charakteristischen Tonhöenschwankungen einer vorgesprochenen Dauertonsilbe durch nachsprechende Versuchspersonen (II)
 TISYS-Programm - PDP 11/50/ Institut für Phonetik, München



Sie sind

DIPLOM-INFORMATIKER

oder stehen kurz vor Ihrem Examen und sind gerade auf der Suche nach einer interessanten beruflichen Aufgabe? Bei uns warten auf Sie eine Reihe anspruchsvoller Projekte, die wir mit Ihrer Hilfe anpacken wollen.

Wir sind ein erfolgreiches Systemhaus auf dem Sektor elektronisches Publizieren.

Wir arbeiten

*auf einer Siemens 7.550 D mit BS 2000,
auf verschiedenen Systemen mit Unix V,
relationalen Datenbanken - 'ORACLE',
Retrieval-Systemen - 'BRS-Search',
Computer-Graphik - 'GKS',
Sprache 'C'
und eigener Satzsoftware.*

Wir sind zu einem Drittel Produktionsbetrieb und zu zwei Drittel Softwarehaus. Unsere Softwareentwickler betreuen ein Projekt vom Kundenkontakt bis zur Produktionsreife und bedienen sich dabei moderner Methoden, wie z.B. Informationsanalyse und datenstrukturorientierter Programmierung.

Wenn es Ihnen Spaß macht, in einem Unternehmen von zur Zeit 20 Mitarbeitern, mit jungen Kollegen an anspruchsvollen Projekten zu arbeiten, dann schreiben Sie uns oder rufen Sie kurzentschlossen an; verlangen Sie unseren Geschäftsführer, Herrn Dörre (Tel. 06151/380383).

HOPPENSTEDT-SATZ-RECHENZENTRUM GMBH

Havelstr. 9, 6100 Darmstadt

IMPRESSIONEN VON DER COLING'86

11th International Conference on Computational Linguistics, Bonn

Vom 25. bis 29. August 86 fand die *11th International Conference on Computational Linguistics*, genannt *COLING*, in den altherwürdigen Räumen der Bonner Universität statt.

Um insgesamt 6 eingeladene Hauptvorträge, 53 Vorträge (40 Minuten) und 99 Kurzvorträge (20 Minuten) zu halten und zu hören, hatten sich rund 600 Teilnehmer eingefunden - so viele wie nie zuvor. Und auch von den Mitgliedern der Organizing Committees hatten sich, laut Martin Kay, noch nie zuvor so viele zu der Tagung selbst eingefunden. Etwa ein Drittel der Teilnehmer kam aus der Bundesrepublik, ein weiteres Drittel aus anderen europäischen Ländern und das letzte Drittel aus Übersee, vor allem aus den USA und Japan (für eine detailliertere Aufschlüsselung fordere man die auf dem abschließenden Banquet von Prof. Lenders gehaltene Rede an!).

Die Menge der Beiträge führte dazu, daß meistens fünf Sessions parallel stattfanden und sehr viele Teilnehmer in dem Bestreben, die interessantesten Vorträge mitzubekommen, ständig zwischen den verschiedenen Vortragsräumen hin- und herwechselten (was aber glücklicherweise dank der meist recht straffen Führung möglich war).

So gab es während der vier Sitzungstage insgesamt

- 3 halbtägige Sessions zum Thema **Semantik**
- 1 Session zum Thema **Dialog**
- 4 Sessions zum Thema **Grammatik**
- 3 Sessions zum Thema **Software**
- 5 Sessions zum Thema **Maschinelle Übersetzung** (die Besetzung des Program Committee Chairman mit Makoto Nagao hinterließ ihre Spuren...)
- 4 Sessions zum Thema **Parsing**
- 3 Sessions zum Thema **Diskurs**
- 2 Sessions zum Thema **Knowledge**
- 2 Sessions zum Thema **Morphologie**
- 2 Sessions zum Thema **Lexikon**
- 2 Sessions zum Thema **Generierung**
- 2 Sessions zum Thema **Speech**

und last but not least

4 Podiumsdiskussionen.

Wem die Qual der Wahl zu schwer fiel, der konnte sich auch im an das Tagungsgebäude angrenzenden Hofgarten auf einer Bank oder in der nahen Bonner Innenstadt auf einem Kaffeehausstuhl niederlassen und den 675 Seiten dicken und 1800 Gramm schweren Tagungsband in einer für seine individuelle "Knowledge Acquisition Komponente" angemessenen Geschwindigkeit studieren.

Wie auf vielen Großveranstaltungen ließen sich auch auf der COLING '86 die berühmten Tagungsphänomene verzeichnen. So passierte es, daß Vortragende

- ... kaum lesbare Folien haben
- ... so weit neben dem Mikrofon stehen, daß man neben dem fehlenden Opernglas auch noch ein Hörrohr vermißt
- ... einen vorher aufgeschriebenen Vortrag Wort für Wort vom Blatt ablesen (was bei Redewendungen wie "and later in my paper I will show that..." besonders komisch klingt)

oder

- ... das Ganze in einer Geschwindigkeit abspulen, daß einem schwindelig wird.

Und manchmal taten die Sitzungsleiter hierzu noch ein übriges.

Ein Beispiel: Der Sitzungsleiter kommt drei Minuten zu spät hereingestürmt. "Who is the first speaker? — What's your name? —" und dann noch schnell zum Publikum "The first speaker in the afternoon session is ..." bevor er sich auf eine Sitzgelegenheit fallen läßt und tief durchatmet.

Oder eine andere typische Situation:

Vortragende. Müdes Klatschen. Der Chairman: "Questions?" Eine Minute Schweigen. "Thank you. Now we come to ..."

Ob die Sitzungsleiter alle schon vergessen haben, wie wichtig für einen leicht nervösen Speaker ein paar freundliche Worte über seine Person und seine Arbeit zu Beginn und wenigstens eine sinnvolle Frage nach Beendigung seines Vortrages sind?

Aber es gibt auch viel Positives zu berichten. Und da seien als erstes genannt die gute Gestaltung (ein gut aufeinander abgestimmtes Tagungs- und Rahmenprogramm ergaben ein abgerundetes Bild) und der reibungslose Ablauf der Tagung - von der Hotelreservierung, über den bereitliegenden Tagungsband, die strikte Einhaltung der Vortragszeit bis hin zu den vielen Bechern Pausen-Kaffee.

Weitere Eindrücke nun in chronologischer Reihenfolge:

Montag, 25.8.:

Nach einer angenehm kurzen Eröffnungssitzung der erste Hauptvortrag von Maurice Gross über *Lexicon-Grammar. The Representation of Compound Words*. Er hat keine Folien vorbereitet (der Zuhörer möge doch bitte die entsprechende Seite im Tagungsband aufschlagen) und kann über den Inhalt des nach compound words durchforsteten Text nichts aussagen (das ginge uns Linguisten-Kollegen auf der Suche nach bestimmten sprachlichen Strukturen doch nicht anders - erster Lacher auf seiner Seite). Dafür kann er umso mehr als Idiome lexikalisierte Komposita bieten - oder können Sie etwa zwischen "the" und "United States" noch ein Adjektiv wie z.B. "powerful" einfügen? Also, in Frankreich sei

dies "out" (zweiter Lacher...). Ein somit amüsanter, ansonsten aber recht müder Auftakt der Veranstaltung.

Nachmittags der erste Slalomlauf zwischen Dialog, Grammatik und Software. Ein paar interessante Mosaiksteinchen, aber keine Highlights darunter.

Zwischendrin bleibt ein bißchen Zeit, um die diversen Stände derjenigen Firmen zu besuchen, die irgendwelche Workstations oder Software-Pakete für den Computerlinguisten anbieten. Die Maschinen sind eigentlich überall die gleichen, aber es ist schon eindrucksvoll zu sehen, wie z.B. die Arbeit mit Grammatiken durch eine graphische Benutzeroberfläche unterstützt werden kann.

Abends gibt es einen Stehempfang in der Aula der Universität mit Wein und Brötchen und vielen alten und neuen Gesichtern. Solche Veranstaltungen zeigen, daß es sinnvoll ist, auf Tagungen zu fahren.

Dienstag, 26.8.:

Vormittags eine Portion Knowledge. Viele framartige Strukturen und exemplarische Beispiele. Es ist leider fast unmöglich, in der Kürze der Zeit die wesentlichen Unterschiede, Möglichkeiten und Grenzen der verschiedenen Ansätze herauszufinden.

Nachmittags in einer Grammatik-Session, die ganz im Zeichen der sich immer mehr durchsetzenden "Unification Based Approaches" steht. Sie schließt mit *Stuart Shiebers* spannungsvoll erwarteten Vortrag "A Simple Reconstruction of GPSG", in dem die Reduktion eines GPSG-Formalismus auf einen einfacheren wie z.B. PATR (einschließlich der verbleibenden Schwierigkeiten) gut nachvollziehbar ist, der Beweis dessen aber schon gar nicht mehr so "simple" erscheint.

Ein Wörtchen sollte man auch verlieren zu der parallel stattfindenden Podiumsdiskussion "Natural Language Interfaces - Ready for Commercial Success?", organisiert und geleitet von Wolfgang Wahlster.

Als erster stellt *Jaime Carbonell* das Language Craft System vor. Es soll seinen Käufer bei der Entwicklung diskursweltabhängiger natürlichsprachlicher Schnittstellen unterstützen. Sein Vortrag zeigt insbesondere Anforderungen an kommerzielle Systeme auf, die zum Teil durch empirische Untersuchungen motiviert sind.

Danach präsentiert *Gary Hendrix* das Q&A-System als kommerziellen Erfolg. Er schränkt jedoch ein, daß es nicht die natürlichsprachliche Schnittstelle sei, die diesen Erfolg ausmache, sondern die Tatsache, daß Q&A ein vollständiges System für den PC ist - mit File Management, Report Generator, Textprocessor, Tippfehlerkorrektur und eben zusätzlich noch mit einer NL-Schnittstelle. Und mit oder ohne letztere werde die Produktivität eines Q&A-Benutzers erhöht...

Harry Tennant argumentiert anschließend, daß herkömmliche NL-Schnittstellen es dem Benutzer nicht gestatten, ihre Leistungsfähigkeit einzuschätzen, was zu allerlei Frustration führe. Als Alternative dazu stellt er NL-Menu vor, bei dem der Benutzer, durch ein Menü geführt, seine Frage eingibt. Vorteil: Der Benutzer weiß von Anfang an, was das System beantworten kann. Nachteil: Eingaben können nur in dem vom System vorgegebenen Rahmen gemacht werden.

Gesamteindruck: In puncto Professionalität ist ein Vortrag besser als der andere und man hat eher das Gefühl, in einer Verkaufsveranstaltung denn in einer Panel Discussion zu sitzen. Dennoch (oder gerade deshalb?) ein interessanter Nachmittag!

Auf dem am Abend vom Barockorchester der Universität Bonn veranstalteten Konzert kann man sich dann von den vielen Wortbeiträgen erholen und zur Abwechslung musikalischen Beiträgen lauschen.

Mittwoch, 27.8.:

Der Mittwoch ist mit einer Exkursion nach St. Goar per Bahn und Schiff den vielen persönlichen Gesprächen am Rande einer Tagung gewidmet. Der Wettergott spielt mit und vom "welcoming gunshot" (das Bild der Kanone auf dem Rasenmäher wird wohl so manchem in Erinnerung bleiben) bis hin zum kalten Abendessen auf dem Schiff (für Engländer allerdings verwirrend, da es Brot und Aufschnitt bei Ihnen nur zum Frühstück gibt) eine angenehme Tour.

Donnerstag, 28.8.:

Morgens in der Diskurs-Session ein Vortrag von *Jürgen Allgayer* über 'Combining Deictic Gestures and Natural Language for Referent Identification'. Schön zu sehen, wie seine Saarbrücker Kollegen alle in der ersten Reihe sitzen und wie Wolfgang Wahlster bei der Beantwortung einer schwierigen Frage sofort Hilfestellung leistet. Die Folien sind - wie immer - absolut professionell.

Eine Working-Session ganz nach dem Geschmack eines COLING-Teilnehmers gibt es am Nachmittag nach der Kaffee-Pause zum Thema "Parsing". Geboten werden vier Vorträge zu Problematiken im Rahmen der Lexical-Functional Grammar (LFG).

Ulli Block präsentiert seinen zusammen mit *Hans Haugeneder* erarbeiteten Ansatz zur Behandlung von "long-distance movement phenomena", basierend auf Grundprinzipien sowohl der LFG als auch der Government & Binding-Theorie.

Jürgen Wedekind zeigt, wie sich LFG reduzieren ließe auf eine einzige Repräsentationsebene, indem die F(unktionale)-Struktur - angereichert mit Informationen über die C(onstituenten)-Struktur - direkt abgeleitet wird.

Rudolf Hunze spricht darüber, wie die C- und F-Strukturen beim Parsen inkrementell aufgebaut werden können mit einem auf dem EARLEY-Algorithmus basierenden Parser.

Klaus Netter hat sich mit dem bekannten, aber weniger geliebten Phänomen der flexiblen Wortstellung des Deutschen beschäftigt und stellt hierzu Regeln vor, die das LFG-Konzept der "Functional Uncertainty" inkorporieren.

Die Vorträge sind gut gegliedert und werden spannend und deutlich vorgetragen, so daß der übliche spätnachmittägliche Müdigkeitsanfall diesmal ausbleibt. Das Publikum, durchsetzt mit LFG-Größen wie Ron Kaplan, Annie Zaenen oder Mark Johnson, reagiert entsprechend: mit Lob und Bestätigung, mit ad hoc entwickelten Vorschlägen zu noch nicht gelösten Problemen, mit Reaktionen auf Reaktionen - kurz mit einem richtigen und intensiven Arbeitsgruppen-Klima!

Freitag, 29.8.:

Zum Auftakt ein interessanter und gut präsentierter Vortrag von Paul Jacobs über 'Knowledge Structures for Natural Language Generation'. Seine ACE-Repräsentationen behandeln konzeptuelles und linguistisches Wissen innerhalb dergleichen Rahmenkonzeption und nutzen dieses Wissen für die Generierung.

Nach dem Mittagessen Walter von Hahn über 'Pragmatic Considerations in Man-Machine Discourse'. Seine Statements zum Thema Integration und ganzheitliche Betrachtungsweise hätten es wirklich verdient gehabt, etwas weniger vorgelesen und mehr vorgetragen zu werden!

Die letzte Sitzung zum Thema Generierung beginnt mit einem sehr schönen Vortrag von Hajo Novak über 'Generating a Coherent Text describing a Traffic Scene'. Hier spürt man den Unterschied zwischen Leuten, die über Konzepte und Ideen reden und jemandem wie Hajo Novak, der in aller Bescheidenheit über ein fertig implementiertes System berichten und Ergebnisse vorweisen kann.

Zu den bemerkenswerten Vorträgen dieses Nachmittags gehört noch der von Graeme Ritchie, der es als allerletzter Vortragender schafft, aus einem relativ trockenen Thema wie 'The Computational Complexity of Sentence Derivation in Functional Unification Grammar' aufgrund seines expressiven Vortragsstils und seiner guten Didaktik richtig spannende zwanzig Minuten zu machen.

Die Closing Session mit den üblichen Danksagungen wie auch das abendliche Banquet mit den vielen Schlußreden zwischen Haupt- und Nachspeiße erhalten ihre Farbe durch Martin Kay, der es wie immer fertig bringt, in Kürze alle Aufmerksamkeit auf sich gerichtet, alles Wichtige gesagt und den Teilnehmern ein Schmunzeln ins Gesicht gezaubert zu haben.

P.S.: Und falls es jemandem nun leid tut, nicht dabeigewesen zu sein: Die nächste COLING findet 1988 aller Voraussicht nach in Budapest und die übernächste 1990 in Helsinki statt.

Brigitte Bartsch-Spörl, Inter Face Consilium München; Birgit Wesche, IBM Stuttgart, Projekt LILOG

DAS LEXIKON ALS BINDEGLIED ZWISCHEN KI UND LDV? Eindrücke von der COLING '86

In der Zeit vom 25. bis zum 29. August 1986 fand in Bonn die "11th International Conference on Computational Linguistics (COLING)" statt, zu der alle zwei Jahre von der Association for Computational Linguistics (ACL) - diesmal gemeinsam mit der DGfS (Deutsche Gesellschaft für Sprachwissenschaft), der GI (Gesellschaft für Informatik) und der GLDV eingeladen wurde.

Dem Ruf ins Rheinland folgten nach einer im Verlauf der Tagung vorgelegten Liste mindestens 650 internationale Teilnehmer, wobei der recht hohe Anteil derer aus den U.S.A. und Japan auffiel. Die umsichtige Wahl des Tagungortes verhielt sicher so manchem ausländischen - teils mit Familie angereisten - Tagungsteilnehmer ein paar stimulierende Tage in Loreleyens Nähe; der Kongreßtourismus blieb offenbar von der Tschernobyl-/Ghaddafi-Flaute verschont.

Welches sind die bleibenden Eindrücke nach einer Tagung dieser Größenordnung? Eine Antwort darauf kann und soll nur subjektiven Charakter haben. Neidlos muß der Bericht etwa, im Rheinischen nicht fremd und mit nur knappem zeitlichen Budget ausgestattet, einen ganzen Tagungstag ausklammern, der anderen für Exkursionen nach St. Goar, Burg Rheinfeld, Loreley etc. zur Verfügung stand. Ein anderer potentiell bleibender Eindruck war die hiesige LDV-Verhältnisse bei weitem übertreffende Größe der Tagung. Es wurden insgesamt etwa 160 Vorträge von 256 Autoren präsentiert, denen auf der Organisationsseite nicht weniger als 84 Reviewer gegenüberstanden.

Zur Bewältigung dieser Informationsmenge in den gegebenen vier Tagen fanden die Vorträge größtenteils in jeweils fünf parallel ablaufenden Bändern (Parsing, Discourse, Dialogue, Know-

ledge, Morphology, Machine Translation, Generation, Grammar, Software, Dictionary und Speech) statt. Dies und das allgegenwärtige Wegweisersystem der Tagungsorganisation sorgten anfangs für ausgedehntere Spaziergänge durch die endlosen Flure der Bonner Universität. Unangenehmer Nebeneffekt: ständige Störung der Referenten durch Nachzügler und/oder Abwanderer. Bei einem zwanzigminütigen Wechsel der Vorträge (etwa zwei Drittel der Vorträge waren auf 20 Minuten begrenzt; einem Drittel standen 40 Minuten zur Verfügung) ist hier kaum noch konzentrierte Aufmerksamkeit möglich.

An dieser Stelle auch der fast schon obligatorische Verweis auf die zu 90 Prozent unleserlichen Folien der Referenten. Gepaart mit dem in etlichen Fällen kaum dechiffrierbaren gesprochenen Tagungsenglisch blieb als einzigen Möglichkeit der Informationsbeschaffung oft nur ein Blick in den 675-seitigen Tagungsband.

Angesichts der Größe der Tagung war das fachliche Interesse des Berichters auf Vorträge zur (maschinellen) Lexikologie im weitesten Sinne konzentriert. Erwartungsgemäß fanden sich interessante einschlägige Themen in den teilweise parallel verlaufenden Vortragsbänden Semantik, Software, Machine Translation, Knowledge, Morphologie und selbstverständlich Dictionary. Angesichts der Anzahl der Beiträge zu dem Thema kann man von einem wiedererwachten Interesse an der Lexikologie und der wortorientierten Semantik sprechen, die ja auch tatsächlich eine zentrale Rolle in allen Bereichen der maschinellen Sprachverarbeitung spielen. Infolge der leistungsfähiger werdenden Systeme bzw. der Verbreiterung deren Einsatzgebiete gerät das Lexikon sowohl im Bereich der LDV (z.B. maschinelle Übersetzung und automatische Indexierungsverfahren) als auch im Bereich der KI (z.B. Expertensysteme) zu einem zentralen Forschungsgegenstand.

Einen Schwerpunkt bilden hier zweifellos die Probleme, die bei der Bearbeitung größerer Mengen an Wortgut durch mehrere Bearbeiter entstehen. - Zur Konsistenzsicherung der semantischen Beschreibung lexikalischer Einheiten durch mehrere beteiligte Lexikographen schlägt C. Zielinski-Wibbelt ein spezielles semantisches Merkmalsystem vor ("An Empirically Based Approach Towards A System of Semantic Features"). Dem gleichen Ziel dienen auf der anderen Seite Überlegungen zum Design konsistenzsichernder lexikalischer DBMS (Domenig/Shann: "Towards a Dedicated Database Management System for Dictionaries") und komfortablen nutzerfreundlichen Dialogsystemen auf konventioneller Basis eines RDBMS (Barnett/Lehmann/Zoepritz: "A Word Database for Natural Language Processing").

Die Probleme sehr umfangreicher und mehrsprachiger lexikalischer Datenbasen behandelte C. Boitet ("Towards Integrated Dictionaries for M(a)T: Motivation and Linguistic Organization"). Zentral

für ihn sind die Ergebnisse des ESOPE-Projekts zur zur Kompatibilität existierender konventioneller Lexika und das Fehlen adäquater DBMS, die sowohl die strukturierte Speicherung arbiträrer Informationen erlauben, wie es lexikalische Einheiten sein können, als auch die angemessene Repräsentation internationaler Zeichensätze (z.B. Diakritika).

Einen weiteren Schwerpunkt der Vorträge bildete die Untersuchung der Möglichkeit des automatischen Aufbaus maschineller Lexika. Als "Wissensquellen" kommen dabei unterschiedliche bereits vorhandene Wörterbücher infrage (Boitet) oder textuelle Dokumente (Tanaka/Yoshida: "Acquisition of Knowledge Data by Analyzing Natural Language" und Trost/Buchberger: "Towards the Automatic Acquisition of Lexical Data"), letzteres mit dem Ziel der automatischen Gewinnung eines morphologischen Analyselexikons.

Die Grenze zur KI-orientierten Wissensrepräsentation vollends verwischten Ansätze zur automatischen Erkennung semantischer Relationen zwischen Lexikoneinheiten (Tsurumaru/Hitaka/Yoshida: "An Attempt to Automatic Thesaurus Construction from an Ordinary Japanese Language Dictionary" und Yokoyama/Hanakata: "Conceptual Lexicon Using an Object-Oriented Language").

Zusammenfassend kann man die in diesem Bereich derzeit stattfindenden Aktivitäten zwei Hauptgruppen zuordnen:

- Aufbau großer lexikalischer Datenbasen für dedizierte computerlinguistische Systeme z.B. im Bereich der maschinellen Übersetzung;
- Analyse und Repräsentation komplexer Begriffssysteme und deren Relationen.

In der ersten - mehr der traditionellen LDV zuzurechnenden Gruppe geht man in der Regel von relativ einfachen und statischen Datenstrukturen innerhalb eines Lexikons aus, benutzt jedoch bisweilen zur Konsistenz- und Integritätssicherung relationale Daten(bank)modelle. In der zweiten Gruppe setzt man in der Regel Wissensrepräsentationssprachen neuerer Generation ein und ist in erster Linie an dynamischen Repräsentationsformen und deren Verhalten, z.B. hinsichtlich Erweiterungen der lexikalischen (Wissens-)Basis interessiert. Der Einsatz solcher lexikalischen Basen ist denkbar im Umfeld von Textverstehens- (Digester- und Informantensysteme) und/oder Expertensystemen. Dabei ist schon jetzt erkennbar, daß sich beide Gruppen aufeinander zu bewegen. Die zunehmende Verfügbarkeit objektorientierter Repräsentationssprachen wird diesen Prozeß noch beschleunigen.

Günter F. Schulz, Gesellschaft für Information und Dokumentation (GID), Frankfurt

LEXIKALISCHE WISSENSQUELLEN IN DER MASCHINELLEN SPRACHDATENVERARBEITUNG

Ein lexikonbasierter Streifzug durch COLING '86

Die gegenwärtige Renaissance des Lexikons in der Linguistik ist insbesondere auch in der Computerlinguistik deutlich spürbar. Dies ist eines der Resümees, die sich nach Abschluß von COLING '86 ziehen lassen. Denn auch wenn Probleme der maschinellen Übersetzung eindeutig im Mittelpunkt dieser 11th International Conference on Computational Linguistics standen, die vom 25. bis 29. Aug. 1986 in Bonn stattfand, so war doch - vielleicht nicht zuletzt auch gerade wegen dieser Schwerpunktsetzung - in vielen Beiträgen aus ganz verschiedenen Sektionen ein starkes Interesse an Fragen des Aufbaus, der Organisation und des funktionalen Stellenwertes von lexikalischen Wissensquellen im Rahmen natürlichsprachlicher Systeme deutlich festzustellen.

Eine Sichtung dessen, was im Verlauf von COLING '86 zum Thema Wörterbuch gesagt wurde, kann sich daher nicht auf eine Zusammenfassung der Vorträge aus den Sektionen *Dictionary I und II* beschränken. Vielmehr - und darum wollen wir uns im folgenden bemühen - gilt es aufzuzeigen, welche sektionsübergreifenden Themenschwerpunkte bei der Erörterung von Fragen des Wörterbuchs während COLING '86 erkennbar wurden.

Da bei COLING '86 fast während der gesamten Konferenzdauer fünf Sektionen parallel tagten, haben wir uns bei unserem nachstehenden Überblick im wesentlichen auf die in den Proceedings veröffentlichten schriftlichen Fassungen der Vorträge gestützt. Dieser voluminöse Tagungsband, der auf 675 zweispaltig und eng bedruckten DIN-A4-Seiten 154 Beiträge aus 23 Ländern vereinigt und damit einen hervorragenden Überblick über den weltweiten Stand der Forschung auf dem Gebiet der Computerlinguistik bietet, ist übrigens erhältlich beim IKS, Institut für angewandte Kommunikations- und Sprachforschung e. V., Poppelsdorfer Allee 47, D-5300 Bonn 1.

Vom maschinenlesbaren Wörterbuch zur allgemeinen Wissensbasis

Oben haben wir die Ausdrücke *Wörterbuch*, *Lexikon* und *lexikalische Wissensquelle* quasi synonym verwendet. Dieser schwankende Sprachgebrauch entspricht einer vorfindlichen terminologischen Unsicherheit: zwischen (maschinenlesbaren) Wörterbüchern, Lexika, lexikalischen Datenbanken und -basen, lexikalischen Wissensbanken, -basen und -quellen sowie allgemeinen Wissensbasen, -banken und -quellen wird selten präzise begrifflich unterschieden, auch wenn sich eine gewisse Tendenz feststellen läßt, eher von Wörterbüchern und Lexika zu reden, wenn diese im wesentlichen morphologische und syntaktische Angaben enthalten, und die mit *Wissen* zusammengesetzten Ausdrücke dann

zu verwenden, wenn semantische Repräsentationen ins Spiel kommen. Mehr oder weniger implizite Andeutungen, die sich in Richtung einer terminologischen Differenzierung interpretieren lassen, finden sich u.a. bei ISODA u.a. (*Model for Lexical Knowledge Base*), NEUHAUS (*Lexical Database Design: The Shakespeare Dictionary Model*), TANAKA (*DCKR - Knowledge Representation in Prolog and its Application to Natural Language Processing*), TANAKA/YOSHIDA (*Acquisition of Knowledge Data by Analyzing Natural Language*), TROST/BUCHBERGER (*Towards the Automatic Acquisition of Lexical Data*) und ZERNIK/DYER (*Disambiguation and Language Acquisition through the Phrasal Lexicon*).

Morphologie im Wörterbuch

Einen ersten Schwerpunkt der Ausführungen stellte die Auseinandersetzung mit morphologischen Phänomenen dar. So verdeutlichen DOMENIG/SHANN (*Towards a Dedicated Database Management System for Dictionaries*) den Charakter ihres dedizierten Datenbankverwaltungssystems für Wörterbücher am Beispiel der Flexionsmorphologie des Deutschen; GOESER/MERGENTHALER (*TBMS: Domain Specific Text Management and Lexicon Development*) heben auf die Bedeutung lemmatisierter Korpuswörterbücher im Rahmen von Textbankverwaltungssystemen ab; NEUHAUS arbeitet am Beispiel der Morphologie die Vorzüge eines an linguistischen Strukturen orientierten Datenbankdesigns heraus; RUSSEL u.a. (*A Dictionary and Morphological Analyser for English*) entwickeln Vorschläge zur effizienten Handhabung großer Mengen morphologischer Daten im Rahmen eines an der Generalized Phrase Structure Grammar orientierten Systems zur Verarbeitung natürlicher Sprache; TROST/BUCHBERGER schließlich beschäftigen sich mit dem semiautomatischen, regelbasierten Erwerb morphologischer Daten beim schrittweisen Aufbau eines Lexikons als Bestandteil eines natürlichsprachlichen Systems. Dieses große Interesse an morphologischen Fragen ist wohl nicht zuletzt damit zu erklären, daß die (Flexions-)Morphologie ein einerseits für natürlichsprachliche Systeme wichtiger, andererseits durch die traditionelle Linguistik hinreichend beschriebener Teilbereich der Sprache ist.

Repräsentationsformalismen für lexikalische Wissensquellen

Das zentrale Thema im Zusammenhang mit der Erörterung lexikalischer Wissensquellen war sicherlich das Bemühen um die Bereitstellung geeigneter Repräsentationsformalismen. Dieses Problem gewinnt vor allem dann an Bedeutung, wenn in lexi-

kalischen Wissensquellen nicht nur morphologisches und/oder syntaktisches, sondern auch semantisches und/oder allgemeines Weltwissen abgelegt werden soll.

Als erstes gilt es bei der Erarbeitung solcher Repräsentationsformalismen natürlich die Anwendbarkeit kommerziell verfügbarer Software zu überprüfen. So argumentiert NEUHAUS - u.a. aus Gründen der Portabilität - für den Einsatz kommerzieller Datenbanksysteme beim Aufbau lexikalischer Wissensquellen, während ansonsten - besonders deutlich z.B. bei DOMENIG/SHANN - eher eine Tendenz zu maßgeschneiderten, selbstgestrickten Lösungen erkennbar ist.

Im Mittelpunkt des Interesses stand aber die Frage danach, wie sich die Einsicht in die Vernetztheit lexikalischer Bedeutungen in entsprechende Repräsentationsformalismen umsetzen läßt, d.h., wie sich die Vielfalt derjenigen einzellexmübergreifenden lexikalischen Relationen, die in traditionellen Wörterbüchern durch die Wörterbuchmakrostruktur in der Regel nur unzureichend erfaßt werden, in einer lexikalischen Wissensquelle zugreifbar darstellen lassen.

DAHLGREEN/MCDOWELL (*Kind Types in Knowledge Representation*) versuchen in diesem Zusammenhang, Wortbedeutungen als commonsense knowledge aufzufassen und durch eine Kombination von kognitiven Modellen der Sprachteilhaber mit prototypensemantischen Beschreibungsansätzen darzustellen; FASS (*Collative Semantics*) arbeitet in seiner 'kollationalen Semantik' mit semantischen Netzwerken mit frameartigen Wortbedeutungsknoten; HORI/TODA/YASUNAGA (*Learning the Space of Word Meanings for Information Retrieval Systems*) versuchen sich von den Beschränkungen symbolischer Bedeutungsrepräsentationen dadurch zu befreien, daß sie einen n-dimensionalen semantischen Raum einführen, dessen Elemente sie als bedeutungsfähige Entitäten interpretieren, denen als ihre Bedeutung die Konfiguration ihrer Umgebung im semantischen Raum zugordnet wird; ISODA u.a. realisieren diejenigen Relationen, die einen variablen und intelligenten Zugriff auf eine lexikalische Wissensquelle sicherstellen sollen, durch dynamische *association interpreters*; NEUHAUS bemüht sich um eine Repräsentation linguistischer Zusammenhänge mit Hilfe des den CODASYL-Datenbanksystemen zugrunde liegenden Netzwerk-Datenmodells; SHIRAI/HAMADA (*Linguistic Knowledge Extraction from Real Language Behavior*) verwenden auf der Analyse von Abhängigkeitsrelationen beruhende Clusterungsverfahren zum Aufbau semantisch relevanter Wortschatzklassifikationen; TANAKA stellt einen DCKR-Formalismus (DCKR = Definite Clause Knowledge Representation) zur Diskussion, in dem die Bedeutungen lexikalischer Einheiten in PROLOG objektförmig durch Mengen von Hornklauseln beschrie-

ben werden, wobei der Formalismus so aufgebaut ist, daß die 'semantische Verarbeitung' dann durch die PROLOG-Systemprädikate und den eingebauten Unifikationsmechanismus erfolgen kann; TANAKA/YOSHIDA betonen ebenfalls die Notwendigkeit, Relationen zwischen lexikalischen Einheiten herauszuarbeiten; TSURUMARU/HITAKA/YOSHIDA (*An Attempt to Automatic Thesaurus Construction from an Ordinary Japanese Language Dictionary*) beschäftigen sich mit der Gewinnung von hierarchischen Relationen, die für den Aufbau eines Thesaurus benötigt werden; YOKOYAMA/HANAKATA (*Conceptual Lexicon Using an Object-Oriented Language*) erörtern die Systematisierung und Kategorisierung abstrakter Begriffe, d.h. den Aufbau eines Lexikons, das abstrakte Begriffe repräsentiert, wobei sie (genauso wie TANAKA) auf einem objektorientierten Formalismus aufsetzen; ZERNIK/DYER schließlich stützen sich bei der Repräsentation von Einträgen in ihr *phrasal lexicon* auf den Schankschen Script-Formalismus.

Semiautomatischer Wissenserwerb beim Aufbau lexikalischer Wissensquellen

Aufgrund des Umfangs und der komplexen Vernetzung natürlichsprachlicher Wortschätze ist der Aufbau lexikalischer Wissensquellen bekanntlich äußerst zeit- und kostenintensiv, insbesondere dann, wenn eine lexikalische Wissensquelle echte, 'semantikbedürftige' Anwendungen unterstützen soll. Es verwundert daher nicht, daß Überlegungen zum semiautomatischen Wissenserwerb beim Aufbau lexikalischer Wissensquellen während COLING '86 einen breiten Raum einnahmen. Im Hinblick auf die dem Wissenserwerb zugrundegelegten Ausgangsdaten lassen sich dabei drei Vorgehensweisen unterscheiden:

Ein Ansatz versucht, lexikalisches Wissen aus Texten bzw. Abfolgen von Sätzen zu extrahieren. So beschreiben SHIRAI/HAMADA die semiautomatische Erzeugung von Wortschatzklassifikationen auf der Basis dependenzanalysierter isolierter Sätze, und TANAKA/YOSHIDA unterziehen die in isolierten Sätzen vorkommenden Relationen zwischen Wörtern einer semiautomatischen Analyse.

Ein anderer Ansatz besteht darin, lexikalisches Wissen nicht aus beliebigen Texten zu gewinnen, sondern gezielt aus den Quellen, in denen man lexikalisches Wissen in konzentrierter Form vermutet: aus Wörterbüchern. TSURUMARU/HITAKA/YOSHIDA etwa entwickeln ein auf der Analyse der Strukturen von Bedeutungserläuterungen in Wörterbüchern basierendes Verfahren zur semiautomatischen Extraktion von Thesaurusrelationen, und auch YOKOYAMA/HANAKATA stützen sich beim Aufbau ihres Begriffslexikons auf Bedeutungserläuterungen aus Wörterbüchern.

Bei einem dritten Ansatz erfolgt der Wissenserwerb im wesentlichen durch einen regelbasier-

ten, auf systeminternes Vorwissen gestützten Dialog mit dem Benutzer. HORI/TODA/YASUNAGA geben einen derartigen Lernmechanismus zum Aufbau der von ihnen eingeführten semantischen Räume an; TROST/BUCHBERGER beschreiben den interaktiven Erwerb morphologischen Wissens; ZERNIK/DYER schließlich erläutern die Prozedur zum schrittweisen Aufbau ihres *phrasal lexicon*.

Syntax im Wörterbuch

Sowohl aus praktischer als auch aus theoretischer Sicht wurden Fragen der Darstellung von syntaktischer Information im Wörterbuch bei der COLING '86 erörtert. BARNETT/LEHMANN/ZOEP-PRITZ (*A Word Database for Natural Language Processing*) stellen sehr detailliert die Kodierung deutscher Verben, Adjektive und Substantive im Heidelberger LEX-Projekt dar. Die Beschreibung beruht auf der Valenztheorie, wie sie auch in den Wörterbüchern von Engel/Schumacher, Helbig/Schenkel, Sommerfeld/Schreiber verwendet ist.

SCHMIDT (*Valency Theory in a Stratificational MT-System*) stellt aus theoretischer Sicht und mit Blick auf die Einbettung in den Formalismus von EUROTRA ein System von Komplementklassen dar, dem EUROTRA-D bei der Kodierung von Wörterbucheinträgen folgt. SOMERS (*The Need for MT-Oriented Versions of Case and Valency in MT*) untersucht die Modelle von Valenz und Tie-

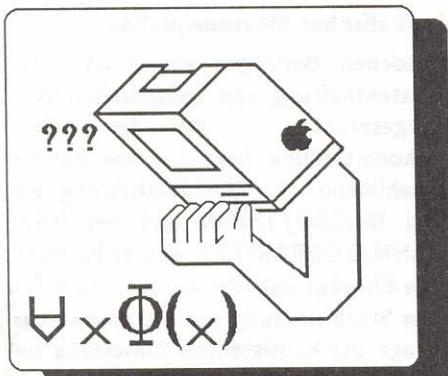
fenkasus im Hinblick auf ihre Anwendbarkeit in der maschinellen Übersetzung. Zur Trennung von Verblesarten schlägt er statt der unscharfen Trennung von Komplementen und Adjunkten ein sechsstufiges System vor, das von idiomatischen Wendungen (*to keep pace*) und Kollokationen (*to take a risk*) über obligatorische und fakultative Komplemente, sogenannte Middles (z.B. Instrumental- und Maßangaben), bis zu Adjuncts und Satzadverbien (*personally*) reicht.

Zwischen der Beschreibung von Syntax und Semantik steht der Vortrag von SCHENK (*Idioms in the Rosetta Machine Translation System*), der zeigt, wie idiomatische Wendungen (aber auch Kollokationen und Mehrwortlexeme, z.B. *to lend someone a hand*) im montagegrammatischen Ansatz seines Systems als komplexe *basic expressions* dargestellt werden können. Auf dieselbe Weise lassen sich einwortige und mehrwortige Äquivalente einander zuordnen.

Semantische Merkmale und semantische Tiefenkasus im Wörterbuch

Für das japanische Regierungsprojekt zur maschinellen Übersetzung (MU-Projekt) stellen SAKAMOTO/ISHIKAWA/SATOH (*Concept and Structure of Semantic Markers for Machine Translation*) ein System von semantischen Merkmalen für 12 Nominalklassen und eine Liste von 33 seman-

Anzeige



PROLOG II

bringt Ihren Computer
ins Gräbeln

PROLOG II ist die ideale Programmiersprache für :

Expertensysteme
Datenbank-Design

Künstliche Intelligenz
Programmentwicklung

natürlich-sprachliche Systeme
Computergestützten Unterricht

PROLOG II ist u.a. verfügbar auf:

Macintosh IBM-PC/-XT/-AT MicroVAX SUN

Die Lieferung umfaßt ein deutschsprachiges Handbuch und ein deutschsprachiges Lehrbuch.

Vertrieb: PhiloLogos Software Wolfgang Kreitmayr & Manfred Krifka, Rauschenerstr. 1-B, D-8060 Dachau

tischen Tiefenkasus vor. Diese Tiefenkasus entsprechen weitgehend den in den *Eurotra Linguistic Specifications* vorgeschlagenen semantischen Rollen. Darüber hinaus gehen die Vorschläge von RUUS/SPANG-HANSEN (*A Theory of Semantic Relations for Large Scale Natural Language Processing*) und STEINER (*Generating Semantic Structures in EUROTRA-D*) zur Beschreibung semantischer Rollen in Wörterbüchern für die maschinelle Ersetzung. STEINERs Rollen werden durch syntaktische Tests zugewiesen. Das System wird bei EUROTRA-D bei der Verbkodierung angewandt.

Wörterbücher für interlinguabasierte maschinelle Übersetzung und Textgenerierung

PAPEGAAIJ/SADLER/WITKAM

(*Experiments with an MT-Directed Lexical Knowledge Bank*) stellen ein lexikonbasiertes Wortexpertensystem vor, das Teil eines englisch-französischen maschinellen Übersetzungsprojekts mit einer dem Esperanto nahen Interlingua ist. Alle Informationen sind in Paaren von Interlinguaausdrücken abgelegt, die hierarchisch geordnet sind (Oberbegriffe). In der lexikalischen Wissensquelle sind auf diese Weise auch Kontexterwartungen ausgedrückt, die dem entsprechen, was Mel'cuk *lexical universe* genannt hat. Dieser Ansatz würde sich auch zur Beschreibung von Kollokationen eignen. Beim Übersetzungsvorgang werden Vererbungsmechanismen im Wörterbuch ausgenutzt.

Die lexikalische Wissensbasis des von JACOBS (*Knowledge Structures for Natural Language Generation*) beschriebenen Formalismus, der einem Textgenerator zugrundeliegt, baut in noch stärkerem Maße auf der Vererbung von Merkmalen auf. Konzeptwissen und linguistisches Wissen werden hierarchisch organisiert, und die beiden Bereiche werden durch *strukturierte Assoziationen* miteinander verbunden. So wird *selling* als eine Unterkategorie von *action* verstanden, wobei *seller* die für *selling* spezifische Ausprägung von *actor* ist. Mit dem System lassen sich linguistische Generalisierungen beschreiben und für die Generierung ausnutzen.

Wörterbücher für transferbasierte maschinelle Übersetzung

Einen kurzen Überblick über Probleme des lexikalischen Transfers und einen Vorschlag zur Überprüfung der Qualität von maschinellen Übersetzungssystemen im Hinblick auf diese Frage gibt MELBY (*Lexical Transfer: A Missing Element in Linguistics Theories*).

Während die in Japan entwickelten Transferwörterbücher sprachpaarbezogen sind und den Auf- oder Abbau von Struktur nicht ausschließen, wird in Europa auch an mehrsprachigen Transferwörterbüchern und an einfachem Transfer ohne Strukturveränderung gearbeitet. NAGAO/TSUJII (*The Transfer Phase of the MU Machine Transla-*

tion System) zeigen, wie Strukturunterschiede zwischen dem Japanischen und dem Englischen in einem dreifach gestuften Transferprozeß behandelt werden können. Das von BOITET/NEDOBEJKINE (*Toward Integrated Dictionaries for M(a)T: Motivations and Linguistic Organisation*) vorgestellte *Fork Integrated Dictionary* enthält sehr detaillierte syntaktische Information sowie stilistische und semantische Markierungen für die jeweilige Quellsprache. Der Übersetzung dienen lediglich Pointer auf Äquivalente in anderen Sprachen. Die Merkmale dieser Äquivalente sind in den jeweiligen Wörterbüchern für diese Sprachen zu finden. Details zur Kodierung der von BOITET/NEDOBEJKINE beschriebenen Wörterbücher finden sich in BUKOWSKI (*Indexage Lexical au Geta*).

Während sowohl die Wörterbücher des japanischen nationalen Forschungsprojekts zur maschinellen Übersetzung (NAGAO/TSUJII) als auch der von BOITET/NEDOBEJKINE beschriebene Typ nur für eine Sprachrichtung konzipiert sind, ist das Wörterbuch von KUDO/NOMURA (*Lexical Functional Transfer: A Transfer Framework in a Machine Translation System Based on LFG*) umkehrbar. Im Rahmen der Lexical Functional Grammar enthält es sprachpaarbezogene Gleichungen (englisch-japanisch), deren beide Seiten sowohl Eingabe als auch Ausgabe des Transfers sein können; die Teilstrukturen, die aufeinander abgebildet werden, erfüllen jeweils Grammatikregeln der betreffenden Sprachen und eignen sich auch für Äquivalente mit unterschiedlicher Syntagmenstruktur.

Konsistenz lexikalischer Wissensquellen

In verschiedenen Beiträgen wurde das Problem der Konsistenzhaltung von lexikalischen Wissensquellen angesprochen. Bei der Kodierung und Lexikonerstellung liegt oft ein Formular oder Auswahlmenu zur Benutzerführung zugrunde (so bei NAGAO/TSUJII und bei BARNETT/LEHMANN/ZOEPPRITZ), das sicherstellt, daß nur sinnvolle Einträge gebildet werden. Zum Teil wird auch bei der Strukturierung von Merkmalssystemen auf die Frage der konsistenten Zuweisung besonders geachtet; beim multilingualen EUROTRA-Projekt hat diese Frage einen besonders hohen Stellenwert (vgl. ZELINSKY-WIBBELT (*An Empirically Based Approach Towards a System of Semantic Features*) und STEINER). DOMENIG/SHANN weisen darauf hin, daß eine lexikalische Wissensquelle dann leichter in einem konsistenten Zustand zu halten ist, wenn sie nicht nur über deklarative, sondern auch über prozedurale Komponenten verfügt.

Anmerkungen

- 1) Unsere Zusammenschau kann - da wir neben den Sektionen *Dictionary I-II* vor allem noch die Sektionen *Knowledge I-II*, *Semantics I-III*, *Machine Translation I-V* und *Generation I* berücksichtigt, die anderen Sektionen aber

kaum beachtet haben, in dieser Hinsicht allerdings keinen Anspruch auf Vollständigkeit erheben.

- 2) Weit über zwei Drittel der Beiträge stammen aus vier Ländern: der Bundesrepublik Deutschland, den USA, Japan und Frankreich. Mit deutlichem Abstand - was die Anzahl der Beiträge anbelangt - folgen Großbritannien, die Niederlande, Kanada, Italien und die UdSSR. Die restlichen 14 Länder sind mit maximal zwei Beiträgen vertreten.

Literatur (alle Angaben beziehen sich auf die Proceedings von COLING '86)

BARNETT, Brigitte / LEHMANN, Hubert / ZOEPPRITZ, Magdalena: A Word Database for Natural Language Processing. S. 435-440

BOITET, Christian / NEDOBEJKINE, N.: Toward Integrated Dictionaries for M(a)T: Motivations and Linguistic Organisation. S. 423-428

BUKOWSKI, Jędrzej: Indexage Lexical au Geta. S. 429-431

DAHLGREEN, Kathleen / MCDOWELL, Joyce P.: Kind Types in Knowledge Representation. S. 216-221

DOMENIG, Marc / SHANN, Patrick: Towards a Dedicated Database Management System for Dictionaries. S. 91-96

FASS, Dan: Collative Semantics. S. 341-343

GOESER, Sebastian / MERGENTHALER, E.: TBMS: Domain Specific Text Management and Lexicon Development. S. 235-240

HORI, Koichi / TODA, Seinosuke / YASUNAGA, Hisashi: Learning the Space of Word Meanings for Information Retrieval Systems. S. 374-379

ISODA, Michio u.a.: Model for Lexical Knowledge Base. S. 451-453

JACOBS, Paul S.: Knowledge Structures for Natural Language Generation. S. 554-559

KUDO, Ikuo / NOMURA, Hirosato: Lexical Functional Transfer: A Transfer Framework in a Machine Translation System Based on LFG. S. 112-114

MELBY, Alan K.: Lexical Transfer: A Missing Element in Linguistics Theories. S. 104-106

NAGAO, Makoto / TSUJII, Jun-ichi: The Transfer Phase of the MU Machine Translation System. S. 97-103

NEUHAUS, H. Joachim: Lexical Database Design: The Shakespeare Dictionary Model. S. 441-444

PAPEGAAIJ, B.C. / SADLER, V. / WITKAM, A.P.M.: Experiments with an MT-Directed Lexical Knowledge Bank. S. 432-434

RUSSEL, G.J. u.a.: A Dictionary and Morphological Analyser for English. S. 277-279

RUUS, Hanne / SPANG-HANSEN, Ebbe: A Theory of Semantic Relations for Large Scale Natural Language Processing. S. 20-22

SAKAMOTO, Yoshiyuki / ISHIKAWA, Tetsuya / SATOH, Masayuki: Concept and Structure of Semantic Markers for Machine Translation. S. 13-19

SCHENK, A.: Idioms in the Rosetta Machine Translation System. S. 319-324

SCHMIDT, Paul: Valency Theory in a Stratificational MT-System. S. 307-312

SHIRAI, K. / HAMADA, T.: Linguistic Knowledge Extraction from Real Language Behavior. S. 253-255

SOMERS, Harold L.: The Need for MT-Oriented Versions of Case and Valency in MT. S. 118-123

STEINER, Erich: Generating Semantic Structures in EUROTRA-D. S. 304-306

TANAKA, Hozumi: DCKR - Knowledge Representation in Prolog and its Application to Natural Language Processing. S. 222-225

TANAKA, Yasuhito / YOSHIDA, Sho: Acquisition of Knowledge Data by Analyzing Natural Language. S. 448-450

TROST, H. / BUCHBERGER, Ernst: Towards the Automatic Acquisition of Lexical Data. S. 387-389

TSURUMARU, Hiroaki / HITAKA, Toru / YOSHIDA, Sho: An Attempt to Automatic Thesaurus Construction from an Ordinary Japanese Language Dictionary. S. 445-447

YOKOYAMA, Shoichi / HANAKATA, Kenji: Conceptual Lexicon Using an Object-Oriented Language. S. 226-228

ZELINSKY-WIBBELT, Cornelia: An Empirically Based Approach Towards a System of Semantic Features. S. 7-12

ZERNIK, Uri / DYER, Michael G.: Disambiguation and Language Acquisition through the Phrasal Lexicon. S. 247-252

Andreas Blumenthal, Projekt COLEX, Germanistisches Seminar der Universität Heidelberg, Hauptstraße 207-209, D-6900 Heidelberg.

Ulrich Heid, Institut für Linguistik, Romanistik, Universität Stuttgart, Keplerstraße 17, D-7000 Stuttgart 1.



Wissenschaftliche Fachkraft auf dem Gebiet der Linguistischen Datenverarbeitung

Das **Bundessprachenamt** in Hürth bei Köln, eine Bundesoberbehörde mit ressortübergreifenden Aufgaben, sucht eine(n) Mitarbeiter(in) mit abgeschlossenem Hochschulstudium (Linguistik, Informatik, Linguistische Datenverarbeitung, Informationswissenschaft), der/die einschlägige praktische Erfahrung hat und mindestens gute Kenntnisse in Englisch, vorzugsweise auch in Französisch besitzt.

Das Bundessprachenamt verfügt über einen bedeutenden Übersetzerdienst, der vorwiegend technische und wissenschaftliche Texte übersetzt. Dem Übersetzer steht eine große terminologische Datenbank (LEXIS) als Hilfe zur Verfügung. Auch auf anderen Gebieten wird der Rechner eingesetzt, so in der Textverarbeitung und in der Information und Dokumentation.

Das Referat "*Rechnerunterstützte Verfahren*" innerhalb der Gruppe "*Terminologie und Lexikographie*" benötigt eine(n) wissenschaftliche(n) Mitarbeiter(in), der/die in Zusammenarbeit mit dem Referatsleiter *Lösungen für linguistische Probleme* in Form von Programmiervorlagen erarbeitet, die von einer Programmiergruppe umgesetzt werden. So sind im Zuge der Weiterentwicklung von LEXIS *maschinelle Hilfen für die Lexikographie* zu entwickeln. Bei der Einführung eines neuen Informations- und Dokumentationssystems zur Vermeidung von Doppelübersetzungen und zum Auffinden von ähnlichen Übersetzungen als Übersetzungshilfen sind *Untersuchungen zu der Dokumentationssprache* erforderlich. Weiterhin sollte der/die Mitarbeiter(in) die *wissenschaftliche Recherche* im Rahmen des Informations- und Dokumentationssystems DOKIS durchführen bzw. koordinieren.

Bezahlung und soziale Leistungen richten sich nach dem Bundesangestelltentarif (BAT) für den öffentlichen Dienst.

Ihre Bewerbung senden Sie bitte mit Lichtbild, Lebenslauf und Zeugnisabschriften über den schulischen und beruflichen Werdegang an das

Bundessprachenamt
Personalstelle
Horbeller Str. 52
5030 Hürth

Tagungsberichte

ON AUTOMATING THE DICTIONARY

Workshop in Marina di Grosseto

Vom 19. bis 23. 6. 1986 fand auf Einladung von Antonio Zampolli, der gemeinsam mit N. Calzolari, L. Rolling (Luxemburg), J. Sager (Manchester) und Don Walker (Morristown), das Organisationskomitee bildete, in Marina di Grosseto ein Workshop zum Thema "On Automating the Dictionary" statt. Mit dieser Veranstaltung wurde eine 1983 in Stanford begonnene Reihe von Meetings zur automatischen Lexikographie fortgesetzt. Besonderes Merkmal der diesjährigen Tagung war, daß Vertreter mehrerer großer Wörterbuchverlage (Longman, Oxford University Press, English Today, Houghton Mifflin, Bibliographisches Institut, Langenscheid) eingeladen und, mit Ausnahme der eingeladenen deutschen Verlagsvertreter, auch erschienen waren. Die Veranstaltung wurde mit Hilfe der EG, der Universität Pisa, der Stadt Grosseto und der ACL finanziert.

In dem um diese Zeit (Mitte Mai) noch relativ ruhigen Küstenbadeort Marina di Grosseto wurden die Teilnehmer von den örtlichen Veranstaltern hervorragend betreut. Die Atmosphäre ermöglichte einen ungestörten und häufigen Austausch von Informationen und Gedanken, trotz des relativ dichten Tagungsprogramms, das die ca. 50 Teilnehmer von 9 Uhr beinahe durchgehend bis 19 Uhr an den Konferenzort band.

Wie aus dem nachfolgenden Programm ersichtlich ist, waren die USA besonders stark vertreten, gefolgt von Großbritannien, Italien, und, mit jeweils nur wenigen Vertretern, Belgien, Bundesrepublik Deutschland, Dänemark, Frankreich, Japan, Kanada, Niederlande, Schweiz und Spanien.

Bemerkenswert war, daß von dem sonst bei Tagungen üblichen Vortragsschema abgewichen wurde: Die vorbereiteten Beiträge, die zum größten Teil vor der Tagung allen Teilnehmern vorlagen und die in einem Sammelband veröffentlicht werden sollen, wurden nicht von ihren Autoren vorgetragen, sondern von einer anderen Person ("discussant") vorgestellt und in die Diskussion eingebracht. Die Folge dieses Verfahrens war, daß sich an jeden Beitrag bzw. an die Stellungnahme des jeweiligen "discussant" eine lebhaft diskussion anschloß, die den anfangs vorgenommenen Zeitrahmen sprengte. Da der Schwerpunkt der Veranstaltung auf der gemeinsamen Diskussion lag, war glücklicherweise von vorneherein genügend Spielraum eingeplant worden.

Themen, Referenten und "discussants":

Identifying the linguistic foundations for lexical research and dictionary design (Richard Hudson/Roland Kaplan)

A survey and comparative evaluation of work on the role of semantics in the lexicon (Beth Levin/Leonard Talmy)

Lexical Information for Parsing Systems: Points of Convergence and Divergence (Robert Ingria / Graeme Ritchie)

The Lexicon in Text Generation (Susanna Cumming/Igor Mel'cuk)

Strategies for dealing with idioms, collocations and routine formulae in Dictionaries (Anthony Cowie/Yorik Wilks)

Lexicographic practice as a source of information for designing dictionaries intended for electronic access (Laurence Urdang/Reinhard H.H. Hartmann)

Adapting existing information sources to create lexical knowledge bases (Robert Amsler/Winfried Lenders)

The use of machine-readable dictionaries for the general public (Timothy Bendow, Edmund Weirer/Howard Webber)

Dictionary Systems for Office Practice (Roy Byrd/Harald Zimmermann)

The potential relevance of new technology for storing and accessing lexical information (Alan Melby/Pilippe Amiel)

The role and the lexicon and machine-readable dictionaries in translation systems (Jonathan Slocum/Benoit Thouin)

Problems and practice in the design and use of bilingual dictionaries (Frank Knowles/Makoto Nagao)

The use of existing machine-readable dictionaries in education and the design of more appropriate ones for the future (McArthur)

The structure of the lexicon as it reflects psychological and social factors (Robert Schreuder/Martha Evens)

Why use words to label ideas: the use of dictionaries and thesauri in information retrieval (Michael Lesk/John Sowa)

Establishing the relevance of dictionary entries for computational linguistic research (Bran Bugaev / Archibald Michiels)

Structure and access in an automated lexicon and related issues (Nicoletta Calzolari/Martin Kay)

Ergänzt wurde dieser Reigen von Vorträgen und Diskussionen durch die Vorstellung einiger nationaler Projekte.

Ein Ergebnis der Tagung läßt sich kaum in wenigen Worten zusammenfassen. Vielleicht waren die folgenden Punkte besonders bedeutsam:

a) Es besteht ein deutlicher Bedarf an PC-Programmen und Workstations

- für den Übersetzer, der die Benutzung von Wörterbüchern und gleichzeitig einen Rückgriff auf Textkorpora ermöglicht;
- für normale Textverarbeitung, die Wörterbücher und Texte (Korpora) im Hintergrund führen, zum Beispiel zum Formulieren in einer fremden Sprache;
- für Autoren von Wörterbüchern, mit der Möglichkeit, auf große Datenbestände im Hintergrund zurückzugreifen;
- für die Lehre, zur Kombination von Textbeispielen mit Wörterbuchinformation; zur Präsentation von Aussprachebeispielen etwa bei Wörtern mit schwieriger Aussprache.
- Bei Lexika, die für Parser erstellt werden, stellen sich die Fragen nach der Vertretbarkeit von Redundanzen, nach der Art des Zugriffs, nach der Orientierung des speziellen Parsers (z.B. semantische oder syntaktische Orientierung).

b) Zur Produktion von Wörterbüchern

- Es ist zu unterscheiden zwischen "dictionaries for people" und "dictionaries for machine". Beide sind - ihrem jeweiligen Zweck entsprechend - unterschiedlich strukturiert und weisen unterschiedliche Inhalte auf; sie können aber auch als verschiedene technische Realisationen einer gemeinsamen lexikalischen Wissensbasis aufgefaßt werden.
- Für verschiedene Zwecke, z.B. in der Lehre und für psycho-linguistische Untersuchungen, wäre es nützlich, auf der Grundlage von Korpora entwickelte Wörterbücher mit Frequenzangaben zu Verfügung zu haben. Benötigt werden nicht nur Frequenzangaben der Grundformen, sondern solche der Wortformen und der Bedeutungsstellen.
- Als wichtige automatisch erschließbare Quellen für Wörterbücher und lexikalische Wissensbanken erweisen sich mehr und mehr Texte und Wörterbücher, die, etwa aufgrund von computerisiertem "typesetting", maschinenlesbar vorliegen, ferner Thesauri und Enzyklopädien. Um aus diesen Quellen lexikalische Angaben, z.B. Definitionen, abzuleiten, genügen in vielen Fällen schon relativ einfache Prozeduren. Für die Ableitung komplexer semantischer Relationen aus solchem Material sind jedoch aufwendige Analysetechniken erforderlich.
- Verlage gehen - neben dem Einsatz der EDV bei der Korpusammlung für Wörterbücher (z.B. New Oxford English Dictionary) - in zunehmendem Maße auch zur Produktion maschinenlesbarer Wörterbücher z.B. auf CD über.
- Besonders wichtig ist die Produktion bilingualer Wörterbücher, in denen der traditionelle Zugang über die alphabetische Liste mit einem wissensbasierten Zugang verknüpft wird.

c) Zum Gebrauch von Wörterbüchern

- Man kommt in der Praxis des Übersetzens nicht mit bilingualen Wörterbüchern aus. Oft muß man in monolingualen Wörterbüchern nachschlagen, um die exakte Wortbedeutung zu finden. Wie kann der Computer bei diesem heuristischen Prozeß helfen?
- Allgemeine Benutzerstudien im Bereich der Wörterbücher wären erforderlich, neben den speziellen, wie sie von Verlagen (etwa OUP) erarbeitet werden.

d) Weitere Aktionen

- Für die Zukunft wird eine engere Zusammenarbeit zwischen Verlagen und Computerlinguisten angestrebt, um sowohl Wörterbücher für den allgemeinen Gebrauch, als auch für die Verwendung in Computerprogrammen zu erstellen.
- Auf Veranlassung der Europäischen Gemeinschaft sollen nationale Übersichten über den Stand der Wörterbucherstellung mit Hilfe des Computers angefertigt werden.
- Als unmittelbares Ergebnis der Beratungen und Diskussionen in Grosseto haben sich schon kleinere Gruppen etabliert, die in Zukunft zusammenarbeiten wollen.
- Es sollen "user studies" erstellt werden, um Anwenderprofile zu erhalten.
- Mehrere Tagungen, Workshops und Tutorials in naher Zukunft befassen sich mit den offenbar hochaktuellen Themen der Automatisierung des Wörterbuchs.

Winfried Lenders, Institut für Kommunikation und Phonetik (IKP), Bonn

DIE KLASSIFIKATION UND IHR UMFELD

10. Jahrestagung der Gesellschaft für Klassifikation in Münster

The Gesellschaft für Klassifikation had a successful 10th Annual Conference from 18-21 June 1986 at Münster with some 130 participants including 15 representatives from 7 other countries.

In the opening session Dr. Ingetraud Dahlberg, chairperson of this Society and of the FID Committee for Classification tried to explain the idea behind the topic of the conference (Classification and Its Environment), explaining that at this point of development of the Society the entire field of classification was to be represented as well as those environmental fields in which classification is somehow embedded, like logic, systems theory, science of science, philosophy, and those fields which are very close to its aims, as linguistics, artificial intelligence and knowledge engineering, economics of thinking and productivity in knowledge handling. The moderator of this session was Dr. Hans-Joachim Hermes, University Library of Münster and responsible for the local organisation of the conference. The inaugural lecture was given by Prof. Dr. Heinrich Schepers, Director of the Leibniz-Forschungsstelle in Münster. He spoke on "The Concept of the Universal Characteristic of Leibniz", a most interesting topic for all those who are - like Leibniz - looking for a universal language of conceptual elements which can form a basis for the expression of any thought. Leibniz had wanted to construct such a language during his lifetime, however, it turned out that he was too far ahead with his ideas as against his possibilities, he did not succeed with this particular project, although he designed the theoretical basis but could not develop a single example of such a language. But his ideas and theories were taken up already by mathematicians and logicians of the last century and can also be taken up today by us!

In the first session (chair: O. Nacke), Wolfgang Dahlberg, Frankfurt, tried to summarize the contents of his voluminous book of 1981/1984 (Order, Being and Consciousness). Hans Sträter, Hamburg, followed; he meant to base his topical statement "A partitive relationship is an abstraction relationship" on empirical studies of (mis)interpreted relationship indications in existing thesauri. He was, however, heavily criticized in the discussion to the effect that the statement of his topic cannot be kept up. - Walter Krumholz, Berlin, reported in "Use and Mis-Use of Thesauri" critically on the results of the recent conference held in Brussels, March 12-13, 1986. - In the next session, (Chair: W. Gödert) two papers were devoted to subject cataloguing: Heiner Schnellling, Berlin, ("The actual function of the subject catalogue in a scientific general library: necessary precision and supplements to existing rules") deplored the lack of theoretical foundation in the new German subject cataloguing rules and made some distinctions to be used in the attempt for later improvements; Hans-Ullrich Weidemüller, Frankfurt, ("RSWK application in the Deutsche Bibliothek (German Library): machine permutation of index-term chains"), showed how the German National Library deals (since 1986) with the particular computerization problems connected to the new subject cataloguing rules (RWSK). - In the last session of this day (Chair: H. Sträter) most interesting papers from three Dutch colleagues showed the pioneering work done in our field in the Netherlands: Gerhard Riesthuis, Amsterdam, ("Alphabetical Subject Index and Classification") explained how hierarchical chains and also the context method improves access to indexes for periodicals. He had used these devices in establishing the alphabetic subject index of a journal for the History of Books and Printing ("Het Boek"). - Anatol Vasiliev, Delft, described (in "Online subject access to library holdings") the online catalogue development and the subject access potency. He extended in particular on the measures to be taken towards putting the Delft University Library catalogues online. - Anne Marie Colenbrander-Dijkman and Gerhard J.A. Riesthuis, Amsterdam, ("Subject Access to online central catalogs: incompatibility of library classifications and subject-heading") dealt with a further burning problem, namely how to overcome the lack of standardization and compatibility to traditional subject cataloguing when trying to access a central catalogue by subject in the online mode. They start from indexed works and use an expert system in order to solve problems such as "How to increase recall?", "How to cope with zero-hit responses?", "How to eliminate irrelevant data from the search request?" etc.. Mrs. Colenbrander-Dijkman presented the results of a feasibility study working with an experimental catalogue of some 60000 bibliographic records.

The first session of the next day (Chair: I. Dahlberg) was devoted to questions of order and classification in economy. Helge Gasthuber, Vienna, outlined some methodological possibilities of "Application of ordering principles in industrial innovation" using many examples from actual cases. It was questioned in the discussion whether one could by this replace the inventive spirit. But Gasthuber did not mean to offer an alternative to replace a genius inventor. Innovation plays a role in many activities of a company and can be assisted very effectively by using the wealth of principles offered by a systematic organization of knowledge. - Another urgent problem was treated by Joseph Hätzl, Vienna, in "Recommendations for commodity descriptions" providing ideas on what is necessary when trying to list those features of a commodity which are requested for a specialised understanding especially for informative purposes in the distributional area. This paper is in line with work of the SIG-WK, a Special Interest Group (of German Classification Society) which held also a meeting on this topic on that day. - A third paper by J. Gesell, Munich, treated the "New Edition of the International Patent Classification (IPC)" and informed on the introduction of so-called hybrid systems and other new features in IPC-4, officially introduced since Jan. 1, 1985.

The section on "System and Taxonomy of Knowledge" (Chair: B. Endres-Niggemeyer) included (1) the paper by Helmut Löckenhoff, Backnang, on the "Didactics of the systems approach: outline of knowledge", which demonstrated the helpful effect of using systems principles in the organization of knowledge for the teaching of business ma-

agement; (2) the attempt of Jiri Panyr, Munich, ("Knowledge and an approach to its taxonomy in the area of Artificial Intelligence: knowledge extraction and -reception") to find definitions for kinds of knowledge, looking at knowledge as an aim of learning ("Wissen als Lernziel, oder Sequenz von Lernzielen"); and (3) the presentation of Norbert Meder, Cologne, ("Automatic learning, concept formation and concept movement") summarizing the most important forms of automatic learning, reducing them to cognition theoretic phenomena of human learning and characterizing kinds of concept formation, organization and movement.

A further section (Chair: G. Greiner) dealt (1) with Robert Fugmann's "Function of semantic categories in indexing languages and in indexing" showing how an indexing language with "multicategorical" concepts can be improved by an indexing grammar which, however, is not to be equalled with the categories of a faceted classification, and (2) with an entirely new approach in handling a faceted classification system in the paper by Lothar Menner and Wiebke Rescheleit, Frankfurt/Offenbach ("Comparison of UDC-faceted classification with the knowledge representation language FRL by selected examples"). They reported on experiences in introduction sections of the UDC and the 2nd ed. of the Bliss Bibliographic Classification into the FRL (Frame Representation Language). As FRL knowledge bases need a faceted structure in order to utilize all FRL functions in an optimal way a faceted classification system seems well prepared for use in such base. They found that facets of the systems studied were not well enough defined. Thus such an application might well have a positive feedback on the construction of adequate faceted classification systems.

The last session dealt with classification in some application fields such as the arts, linguistics and applied statistics. Kim Veltman, Toronto, showed a new framework for the classification of fine arts and visual images in terms of connecting, matching, ordering, systematizing, mixing, exploring and integrating. He had again a great number of slides at his hand for a survey of 19th and 20th century developments. - Klaus D. Dutz, Münster, presented "Classification and systematics in the linguistic area, problems and perspectives" and demonstrated the classification system used in his institute. In general the discussions were somewhat shocked about the lack of theory in this system. However, somebody also said, a bad system is better than no system; and with bad system a big cardfile of document cards is organized, helping teachers and students to find their literature. - Last but not least, Maria Domokos, Budapest, spoke about the "Metaorder in statistical data documentation" which is a special feature of this kind of documentation, as it has to handle very big masses of data. Here one has to work on two levels, a factographic question may then well be answered by a bibliographic reference. For such an approach, meta-information systems have to be established.

The last morning (Chair: N. Meder) was devoted to a section on "Algebraic methods and concept analysis". The four papers - all from the Technical University of Darmstadt - dealt with the following problems: Bernhard Ganter: "The problem of clustering in view of formal concept analysis", who showed that only those partitions make sense which can be separated on the basis of characteristics. From the point of view of concept analysis it appears that only those partitions are of interest for which all classes are extensions. - Uwe Kipke ("Concept lattice schemes for the determination of objects") presented a program which classifies objects through those characteristics which the user has recognized. The searcher is assisted during the program run by the graphic representation of the actual position in the Hasse-diagram of a concept lattice, as well as simultaneously by the next menu of the rest of characteristics. - Jürgen Stahl ("Measurability theorems in formal concept analysis") explained how the formal concept analysis method, developed at the Technical University of Darmstadt can be used for a measurement theory based on concept structures. - Rudolf Wille treated the problem of "Dependencies of attributes" (characteristics) which occurs between mono-valued characteristics and multi-valued ones and which can - for limited quantities of characteristics - equally be solved by formal concept analysis methods. -

In her concluding paper on "Future perspectives of classification and indexing" I. Dahlberg gave a short survey on the topics and quantity of interest in classification, expressed through the annual conferences and their proceedings of German Society for Classification throughout the last ten years, and presented ten current problems in classification which she finally summarized for a general discussion as well as for the possible future work of the Society in the form of ten theses.

The proceedings of this conference will be published by INDEKS Verlag, Frankfurt in the Fall of 1986.

Ingetraut Dahlberg, INDEKS-Verlag Frankfurt

SEMIOTIK UND WISSENSCHAFTSTHEORIE

Kolloquium an der Ruhr-Universität Bochum

Semiotik und Wissenschaftstheorie war das Thema eines Kolloquiums, das von W. A. Koch in Verbindung mit dem Bochumer Semiotischen Kolloquium (BSC) und von R. Posner in Verbindung mit dem Bereich "Sprache" der DGS am 5. und 6. Juni 1986 an der Ruhr-Universität Bochum veranstaltet wurde. Ziel des Kolloquiums war es, die Trennung zwischen semiotischer Forschung und den wissenschaftstheoretischen Bemühungen in der Analytischen Philosophie und den Einzelwissenschaften überwinden zu helfen und zu einem engeren Verhältnis dieser beiden interdisziplinären, fächerübergreifenden Bestrebungen beizutragen. Zu diesem Unternehmen trugen die einzelnen Referenten bei, indem sie in ihren Vorträgen den wissenschaftstheoretischen Status der Semiotik untersuchten, semiotische Aspekte wissenschaftstheoretischer Probleme analysierten und einzelwissenschaftliche Fragestellungen von einem gleichermaßen semiotischen wie wissenschaftstheoretischen "Metastandpunkt" behandelten.

H. Schnelle (Bochum) sprach in seinem Vortrag über "Wissenschaftstheoretische Voraussetzungen für eine theoretische Semiotik" und über die Probleme, die sich für eine Grundlegung der Semiotik auf der Basis einer naturalisierten Wissenschafts- und Erkenntnistheorie ergeben, die die Ergebnisse der KI-Forschung integrieren. Unterschiedliche Auffassungen des Gegenstandsbereichs und der Methoden der Semiotik behandelte Ch. Hubig (Berlin) in seinem Vortrag "Semiotik als Objekt- und als Metawissenschaft". Hubig untersuchte die Auffassung von Semiotik als Einzelwissenschaft neben anderen, die die verschiedenen Zeichensysteme zu ihrem Gegenstandsbereich hat, und die Konzeption, nach der Semiotik ein fächerübergreifendes, rekonstruktives und analysierendes Unternehmen ist, das diejenigen Aspekte der Einzelwissenschaften untersucht und vereinheitlicht, die mit Zeichenprozessen verbunden sind.

Eine wissenschaftstheoretisch relevante, wichtige und einflussreiche Sprachphilosophie analysierte Ch. Stetter (Aachen) in seinem Beitrag "Die logisch-semiotische Problematik in Wittgensteins Tractatus". Er untersuchte die Beziehungen von Wittgensteins Bedeutungstheorie im Tractatus zur Philosophie Kants sowie die Implikationen, die sich aus einer solchen Interpretation Wittgensteins für das Verhältnis von Logik und Ästhetik ergeben. Dabei konfrontierte er Wittgensteins Bedeutungstheorie mit der Semantik Freges. - "Semiotische Aspekte der Simulation von Zeichenprozessen, speziell von Sprachprozessen" behandelte S. Kanngießer (Osnabrück). Er stellte Versuche vor, Ansätze der formalen Semantik (Montague) in die KI-Forschung zu integrieren, und behandelte in diesem Zusammenhang die Fragestellung nach Gemeinsamkeiten und Unterschieden menschlicher und maschineller Zeichenprozesse. Kanngießer zeigte, welche Rolle die "Churchsche These" für die Beantwortung der Fragen nach dem Verhältnis von Mensch/Geist und Maschine/ Programm spielt. - G. Heinzmann (Saarbrücken) stellte in seinem Vortrag Überlegungen "Zur Ausbildung der semiotischen Aspekte in pragmatischen Ansätzen der Mathematik" an. Er untersuchte exemplarisch

zwei solche Ansätze: Ferdinand Gonseths Versuch einer Grundlegung der Geometrie und Hugo Dinglers Ansatz zur Begründung der Arithmetik. Heinzmann zeigte, welche Problemstellungen semiotischer Art sich in diesen (und vergleichbaren) Ansätzen der konstruktiven Mathematik ergeben und welche Lösungsversuche vorgeschlagen worden sind. - "Basistheorien der Semiotik" hatte A. Eschbachs (Esen) Vortrag zum Thema. Eschbach untersucht die entwickelteren Theorieansätze, die in diesem Jahrhundert in der Semiotik vorgeschlagen worden sind (insbesondere den Ansatz Bühlers) und ihre Beziehungen hinsichtlich ihres wissenschaftstheoretischen Status zu älteren, "klassischen" Ansätzen (z.B. dem von Locke). Eschbach warnte davor, die unreflektierte Übertragung von Ausdrucksweisen schon für eine theoretisch fruchtbare Semiotik zu halten, und er sprach sich (mit Bühler) für eine genauere und methodisch strengere Abgrenzung des Gegenstandsbereichs der Semiotik aus. - Das Verhältnis von Wissenschaftstheorie und Semiotik behandelte K. Robering (Berlin) in seinem Vortrag "Semiotik und die historisch vorliegenden Programme zur Zeichentheorie". - "Der wissenschaftstheoretische Status der angewandten Semiotik in Architektur und Design mit Rücksicht auf die Kunstwissenschaft" war der Titel des Vortrags von C. Dreyer (Detmold). Innerhalb dieses Themenrahmens besprach Dreyer auch die Konzeption von Wissenschaft, Wissenschaftstheorie und vom wissenschaftstheoretischen Status der Semiotik, wie sie in der Postmoderne-Diskussion der Architektursemiotik vertreten werden. - K. Eimmermacher (Bochum) behandelte in seinem Vortrag "Semiotik und Literaturwissenschaft" die wissenschaftstheoretischen und semiotischen Probleme der Interpretation von Zeichen, speziell von literarischen Texten. - Semiotische Aspekte der KI-Forschung, Konsequenzen, die sich aus den Ergebnissen der "Kognitiven Wissenschaften" für die Semiotik ergeben und die Semiotik der Wissenschaftstheorie besprach E. Hohenstein (Bochum) in seinem Beitrag "Kognitive und semiotische Wissenschaft". - Eine großangelegte Systematik der Gegenstandsbereiche der Einzelwissenschaften und eine Analyse der Stellung der Semiotik innerhalb dieser Systematik lieferte W.A. Koch (Bochum) mit seinem Beitrag "Semiotik, Kultursemiotik und Evolutionäre Kultursemiotik als Prototypen für die Aufteilung von Wissenschaft und Wirklichkeit". Vor dem Hintergrund der Evolutionären Erkenntnistheorie entwickelte Koch eine Theorie der Gegenstandsbereiche der Einzelwissenschaften, wobei der evolutionären Herausbildung verschiedener Typen von Beziehungen zwischen Strukturen/Ganzheiten eine besondere Rolle sowohl für die Erweiterung eines solchen Gegenstandsbereichs als auch für die Stellung der Semiotik im System der Wissenschaften zukommt. - Im letzten Vortrag behandelte R. Posner (Berlin) "Die systematisch denkbaren und heute de facto vertretenen Konzeptionen zum wissenschaftlichen Status der Semiotik". Posner unterschied zunächst verschiedene Abgrenzungskriterien für Arten von Wissen (z.B. eigener Gegenstandsbereich, eigener Beschreibungsapparat, eigene Darstellungsmethoden usw.) und entwickelte dann in einer Art Kombinatorik dieser Kriterien eine Systematik verschiedener Typen (mehr oder weniger) systematisierten Wissens (Beispiele solcher Typen: Doktrin, Analytik, Theorie usw.). In einem zweiten Schritt untersuchte Posner dann anhand einer umfangreichen Autorenliste, die von Poincaré 1637 bis Ivanov 1985 reichte, wie einzelne Autoren in der Geschichte der Semiotik den Status der Semiotik gesehen haben.

Ich glaube, daß man auf dem Hintergrund des Bochumer Kolloquiums drei Problembereiche ausmachen kann, denen momentan das Hauptinteresse der wissenschaftstheoretischen Diskussion innerhalb der Semiotik gilt: (1) Welchen Status hat die Semiotik selbst im System der Wissenschaften (ist sie eine Theorie, ein Theorienbündel, ist sie überhaupt "systematisiertes Wissen", sollte sie das sein)? (2) Welche Rolle spielen naturalisierte und evolutionäre Erkenntnis- und Wissenschaftstheorien (findet die Semiotik ihren Platz im Gebäude der Wissenschaften nur dann, wenn dieses Gebäude selbst auf solchen Theorien gründet, oder sind diese Theorien ein Rückfall in Soziologismus, Psychologismus und Biologismus)? (3) Welche Implikationen haben die "Kognitiven Wissenschaften" einschließlich neuer Ansätze der "philosophy of mind" (z.B. Sear-

les Intentionalitätstheorie) für die künftige Entwicklung der Semiotik? Ein differenzierteres Bild über diese und andere in Bochum behandelte Fragen wird sich der interessierte Leser machen können, wenn, wie geplant, die Vorträge und Ausschnitte der jeweils anschließenden Diskussionen veröffentlicht sind. - Es steht zu hoffen, daß die Bochumer Tagung, die von den Veranstaltern vorzüglich organisiert worden ist, dazu beiträgt, daß die Diskussion zwischen Semiotik und Wissenschaftstheorie in Zukunft intensiver verläuft, als dies in den letzten Jahren der Fall war. Hier muß jedoch auch der m.E. einzige berechtigte Kritikpunkt angebracht werden: Zwar waren alle Teilnehmer der Tagung wissenschaftstheoretisch interessiert (was ja bereits der Umstand ihrer Teilnahme beweist), wäre es aber dem Zwecke des Kolloquiums nicht noch förderlicher gewesen, wenn auch ein "hauptberuflicher Wissenschaftstheoretiker" teilgenommen hätte?

Klaus Robering, Berlin (überlassen vom Arbeitskreis für Semiotik, Berlin)

SPRACHE UND INFORMATION IN WIRTSCHAFT UND GESELLSCHAFT Internationaler Kongress in Hamburg

Zum Thema Sprache und Information in Wirtschaft und Gesellschaft fand vom 30.9. bis 3.10.1985 in Hamburg ein internationaler Kongress statt. "Ist die Gesellschaft, in der wir leben, bereits das, was wir eine Informationsgesellschaft nennen, und wenn ja, was ist zu tun, damit wir ihre Krisen und Herausforderungen besser bestehen als bisher, d.h. konkret: was muß geschehen, damit aus der Informationsgesellschaft eine informierte Gesellschaft wird?" Das war sicher die wichtigste Frage, die im Zentrum der Diskussionen des Kongresses stand, der unter der Leitung von Theo Bungarten (Hamburg) von der Internationalen Vereinigung "Sprache und Wirtschaft" organisiert worden war. Wie es sich für einen Kongress dieser Art gehört: Internationalität war gefragt, und entsprechend attraktiv war das Angebot der Referenten. So sprachen immerhin Experten aus 10 europäischen Ländern, darunter aus Ungarn, Jugoslawien, Italien, Frankreich, Dänemark, Finnland und der DDR, wobei es sich als außerordentlich fruchtbar erwies, daß die Referenten aus Wissenschaft und Praxis kamen - die Wissenschaftler also nicht unter sich waren. Das bewirkte, daß sich ein lebhafter Dialog zwischen den Theoretikern und denen entwickelte, die die Ergebnisse der Theoretiker in Betrieb, Rechenzentrum, Verwaltung und Übersetzerbüros in ihrer täglichen Arbeit anzuwenden haben, ein Dialog, der vor allem folgenden Sachverhalt zu Tage förderte: daß es so etwas wie einen "doppelten Vorwurf" gibt, der dafür verantwortlich ist, daß man so schwer zueinander findet. So warfen die Praktiker den Wissenschaftlern vor, daß ihre Ergebnisse zu wenig praxisnah seien, und die Wissenschaftler konterten, indem sie darauf hinwiesen, daß die Praktiker sich grundsätzlich weigerten, ihre Ergebnisse zu rezipieren, geschweige denn anzuwenden. Dies war eigentlich in allen Sektionen so, insgesamt 10 übrigens, in denen es um folgende Schwerpunkte ging: Information als Rohstoff, Kommunikationstechnologien und gesellschaftliche Zukunft, Übersetzungsprobleme in der Wirtschaft, Sprache zwischen Produzent und Konsument, Interkulturelle Kommunikation im Außenhandel, Kommunikation am Arbeitsplatz, Fremdsprachen in Industrie und Handel, Datenbanken und Informationssysteme, Wirtschaftsterminologien und Fachsprachen und schließlich Gutachtertätigkeiten in der Wirtschaft.

Eröffnet wurde der Kongress von Theo Bungarten, der in seiner Begrüßungsrede auch gleich einen der entscheidenden Akzente setzte, und zwar indem er auf die unübersehbaren gesellschaftlichen Probleme hinwies, die die "informationelle Revolution" uns beschert habe und die zu lösen uns noch nicht gelungen sei. Dabei komme es darauf an, so

Bungarten weiter, sich nicht damit zufrieden zu geben, daß es sich bei den gemeinten gesellschaftlichen Problemen um Krisenerscheinungen handle, die nun einmal unvermeidbar seien, sondern vielmehr darauf, sie auch als Chance zu begreifen, die Welt, in der wir leben, mit mehr Gewinn zu erfahren, denn, so sein Fazit: Wo eine Krise ist, da ist auch eine Chance. Natürlich ist es nicht möglich, an dieser Stelle auf die einzelnen Sektionen bzw. auf die Referate - immerhin nahezu 50 an der Zahl - näher einzugehen bzw. sie unter dem Aspekt zu werten, wie sie einzuschätzen sind hinsichtlich ihrer Bedeutung für die Bewältigung aktueller und zukünftiger Probleme. Dafür waren die einzelnen Sektionen in ihrer Thematik auch zu unterschiedlich, und das ist durchaus auch als Kritik gemeint. Aber dennoch ist es angebracht, auf einige Referate besonders hinzuweisen, z.B. auf das von Justa Holz-Manttari von der Universität Tampere in Finnland, die sich mit der Thematik "Textbau-Experten: Fachkompetenzen, Berufsprofile, Arbeitsplatzkompetenzen" auseinandersetzte, sowie das von Agota Borgulya von der Universität Pecs (Ungarn), die sich "Merkmale der deutschsprachigen Texte der Dokumentation in der Wirtschaft" vorgenommen hatte, und zwar aus der Perspektive eines (ungarischen) Übersetzers, der Begriffe einer wirtschaftlichen Erfahrungswirklichkeit zu übersetzen hat, die mit der eigenen wirtschaftlichen Wirklichkeit nicht in Einklang zu bringen sind, und nicht zuletzt auch das Referat von Rosemarie Gläser von der Karl-Marx-Universität in Leipzig, die die "Neologismen im englischen Wortschatz der Ökonomie und Probleme der Übersetzung" behandelte. Keine Frage, daß ihr Vortrag besonderen Eindruck hinterließ, wobei vor allem anzumerken ist, daß sie die Problematik von einer Perspektive aus anging, die bei Wissenschaftlern nicht selbstverständlich, sondern eher die Ausnahme ist: von der Perspektive des Praktikers. - Erwähnenswert dürfte allerdings noch ein anderer Aspekt sein, und zwar die Einrichtung der Sektion "Kommunikation am Arbeitsplatz", einer Sektion, die auf den Tagungen dieser Vereinigungen zuvor de facto keine Rolle gespielt hatten. In den Vorträgen dieser Sektion wurden vor allem jene Probleme reflektiert, die neuerdings allgemein unter der Überschrift "Wertewandel der Arbeitsgesellschaft" abgehandelt werden; also solche Probleme, die dadurch entstehen, daß mit zunehmender Anwendung immer neuer Produktionstechnologien in vielen Produktionsbereichen der Kommunikationspartner Mensch durch den Informationspartner Computer ersetzt wird, und zwar mit der nachweisbaren Konsequenz, daß die so hergestellte Sprachlosigkeit der Arbeitswelt bzw. die so verursachte Sprachverkümmung am Arbeitsplatz grundsätzlich zur menschlichen Verkümmung der berufstätigen Individuen führt. Dabei wurde deutlich, daß in diesem Zusammenhang vor allem die berufsbildenden bzw. berufsbegleitenden Schulen gefordert sind, dieser Tendenz entgegenzusteuern, und zwar dadurch, daß sie die neuen Probleme am Arbeitsplatz, die die berufstätigen Individuen zunehmend zu bestehen haben, zu ihren Unterrichtsgegenständen machen. Nur auf diese Weise, so die übereinstimmende Auffassung, könne sichergestellt werden, daß die Heranwachsenden in ihrer späteren Rolle als berufstätige Individuen nicht zum Mündel degradiert werden, sondern ihre berufliche Handlungsfähigkeit behalten. Ein sicher wichtiger Aspekt, wie man sieht, der hier behandelt wurde, aber eben nur einer neben ebenso wichtigen anderen. Wie man überhaupt sagen kann, daß die in den einzelnen Sektionen erarbeiteten Ergebnisse ganz unstrittig dazu beigetragen haben anzuzeigen, wohin die modernen Arbeitsgesellschaften treiben werden und daß es an uns liegt, d.h. an unserer Fähigkeit, die Verhältnisse angemessen zu versprachlichen und die Fülle der Informationen zu nutzen, ob wir nun nur die Alternative haben, uns der zukünftigen Entwicklung passiv anzupassen oder ob wir in der Lage sind, sie aktiv mitzugestalten.

Hilmar Grundmann, Universität Hamburg (überlassen vom Arbeitskreis für Semiotik, Berlin)

MEDIENWIRKLICHKEIT

Semiotisches Seminar am Institut für Wissenschaft und Kunst, Wien

Im Zusammenhang mit seiner Differenzierung der Kulturkritik in eine integrierte, die Aspekte der Massenkultur euphemisierende, und in eine apokalyptische, negativistische der Theorie des Zerfalls verhaftete Kritik der Kulturindustrie hat Umberto Eco den Ausblick auf eine semiotische Guerilla gegeben, um den Gedanken auszudrücken, daß es nicht allein um den von ihm ebenfalls postulierten tätigen Eingriff der Kulturschaffenden in die "Medienlandschaft" gehe, sondern darüber hinaus um eine massive Veränderung der Rezeptionsgewohnheiten. Er stellte sich das so vor, daß "kulturell gut vorbereitete Gruppen den anderen helfen könnten, das Fernsehen, die Zeitungen, die Unterhaltungsliteratur und die Werbung zu lesen." Genau dieser Aspekt der Sensibilisierung für den Wahrnehmungsprozeß der unterschiedlichen Botschaften in Fernsehdokumentationen und Reportagen stand im Mittelpunkt eines im Oktober/November 1985 am Institut für Wissenschaft und Kunst (Wien, Berggasse 17/1) durchgeführten Seminars zur MEDIENWIRKLICHKEIT (Koordination: Haimo L. Handl und Frank Hartmann). In vier Sitzungen wurde das bereits im Zusammenhang mit einem "Newspeak"-Seminar vorgestellte Modell einer Kritik der Kulturindustrie (vgl. z.B. auch die Analyse zum Fernsehen von Th. W. Adorno) mit Anschauungsmaterial ausgestattet: nach einer theoretischen Einführung in das Thema - das untergliedert war in 1) die Frage von Objektivität und Authentizität von Nachrichtsbildern, 2) das Problem der ontologischen Differenz von realem und medialem Weltbild, 3) das Verhältnis von Medienauftrag und Rezeptionsgewohnheiten besonders im Hinblick auf das Umschlagen von Information in Unterhaltung, Voyeurismus und Sensationsgier, und schließlich 4) die Schaffung von Faktizität durch Selektion am Beispiel von Gewaltdarstellung in den Medien - und nach einer Einführung in Filmanalysetechniken wurden in gemeinsamer Diskussion Analysen und Interpretationen der auf Video gezeigten Dokumentationen und Reportagen erarbeitet. Dabei wurde die Perspektive auf einen möglichen emanzipativen Mediengebrauch nicht außer acht gelassen; so zeigt allein schon die Arbeit mit dem Videogerät, daß der medien-spezifische Code nur gebrochen werden kann, wenn die Chance besteht, die Mechanismen des Mediums gegen dieses selbst zu wenden. Erst die Verallgemeinerung eines Zugangs zur Medienproduktion könnte deren Homogenität brüchig werden lassen, erst Nachbearbeitung, Montage oder Isolierung einzelner Sequenzen deren Prinzipien durchsichtig machen. Der veränderte Impuls ist ja keineswegs allein aus der Technik herzuleiten; gerade in der Elektronisierung ist diese vorerst nur ein weiteres Mittel, die Rezipienten ganz im Sinne der Rationalisierung von Arbeitsprozessen zur Übernahme von Dienstleistungen zu zwingen. Die Diskussion über die Veränderung durch "neue Medien" wäre stärker an die Frage der Mediennutzung zu binden - einer Analyse von Andre Gorz zufolge ist diese nicht so festgeschrieben, wie es bei den herkömmlichen Großtechnologien der Fall war. Die potentielle Offenheit der mikroelektronischen Revolution im allgemeinen weise nämlich in beide Richtungen, in die der Hyperzentralisierung wie in die der Selbstverwaltung. Eine endgültige Orientierung würde demnach davon abhängen, wie weit eine transversale Kommunikation, also die horizontale Nutzung der Informationsaustauschmöglichkeiten zwischen den Individuen gegenüber der die Rezipienten atomisierenden Vermittlungshierarchie gelingen wird, wie weit also eine autonome Organisation von Kommunikationsprozessen mit den neuen Techniken gelingen könnte. Neben der Untersuchung von Bedingungen der Produktion und der Rezeption bedarf es hier einer intensiveren Strukturanalyse der massenmedialen Kommunikation als es bislang von Kommunikationstheoretikern vorstellig gemacht wurde. So wäre im Zusammenhang mit den neuen Medien, wie immer man diesen Begriff jetzt faßt, vor allem die Stellung des Individuums zum öffentlichen Leben kommunikationstheoretisch zu begreifen, das heißt die Tragweite ihrer Reduktion auf bloße Adressaten einer ununterbrochenen Kette serieller Botschaften zu analysieren. Wie in diesem Seminar zu zeigen versucht wurde, kann die semiotisch inspirierte Analyse einiges dazu beitragen, derartige Untersuchungen aus dem Bereich rein quantitativ vorgehender Sozialforschungsmethoden zu lösen, wenn es etwa, um nur

ein Beispiel zu nennen, darum geht, die Parallelisierung von Realitätsabbildung und Fiktion aufzuzeigen (etwa: der Westenheld und Politiker, oder die Suggestion von Leben nach dem Atomkrieg in täuschend realer Bilderwelt). Die Möglichkeit zur Verständigung über Rezipiertes wird jedoch vom Medium Fernsehen systematisch abgeschnitten, so daß Absurditäten konsumiert werden, ohne daß sie Widerspruch auslösen. War die methodische Problematik in einem doch auch didaktisch angelegten Seminar nur andeutbar, so konnte die allgemeine Diskussion vor allem zwei Punkte deutlich herausstellen:

1) Die Medien konstituieren unser Weltbild zu einem nicht geringen Teil, determinieren es aber nicht. Ein Zurück zur Immanenz der rein personalen Weltbildvermittlung wäre eine regressive Utopie. Diese bricht sich schon daran, daß die Massenmedien, die sich mit standardisierten Produkten an eine undifferenzierte Öffentlichkeit wenden, selbst einem Wandel unterworfen sind, der nicht nur technischer Natur ist - es zeichnet sich zur Re-Individualisierung der Medien ab, da immer speziellere Adressaten, eine Art Vielzahl von Mini-Öffentlichkeiten angesprochen werden. Die Orientierung an spezifischen Bedürfnissen könnte dazu beitragen, das Verhältnis des einzelnen zum Medienprodukt grundlegend zu ändern, und sei es, indem er sich vorerst wieder konkret als Adressat wissen kann.

2) Ein Erfahrbarmachen der Medienbotschaft ist einer schlecht pädagogisierenden Vermittlung vorzuziehen. Konkrete Kritikpunkte waren hier die Fettschierung von Information (aus Nachrichtensendungen etc.) zu einem Wert an sich; sodann der eigentümliche Zwang zur Teilhabe am Mediengeschehen, die Anlage von Voyeurismus, der nur noch mittels individueller Strategien zu durchbrechen wäre (der klassische Voyeur schöpft bekanntlich seinen Genuß aus dem selbstgeschaffenen Zwang des Nicht-eingreifenkönnens, was sich immer mit einer Billigung des Geschehens verbindet). Wenn es gelingen würde, diesen Zwangszusammenhang erfahrbar und durchschaubar zu machen, indem die Abstraktion von potentiellen Handlungen als der wesentliche Faktor allen Mediengeschehens herausgestellt wird, dann wäre es vielleicht möglich, jene Vereinzelung aufzuheben, die im Zentrum des kulturindustriellen Betriebs steht. Erst dann ließe sich die Hoffnung einlösen, die Walter Benjamin in den 30er Jahren bezüglich der reproduzierenden Medien zu formulieren wagte: daß an die Stelle der hierarchischen Vermittlung mit ihrer Konsequenz der Vereinzelung die Kollektivrezeption der zerstreuten Examinatoren tritt, aus der hieraus die (wissenschaftlich-) kritische und die (künstlerisch-) genießende Haltung wieder ineinander fließen.

Frank Hartmann, Wien (überlassen vom Arbeitskreis für Semiotik, Berlin)

"SPRACHE UND HANDELN"

DGS-Jahrestagung 1986 in Heidelberg

Die 8. Jahrestagung der Deutschen Gesellschaft für Sprachwissenschaft in Heidelberg vom 26. bis 28. Februar 1986 stand unter dem Schwerpunkt-Thema Sprache und Handeln. Den Eröffnungsvortrag hielt G. Meggle (Münster) über "Handlungstheoretische Semantik und linguistische Pragmatik". Ein weiterer Vormittag war der Vorstellung von Teilprojekten des DFG-Forschungsschwerpunktes "Verbale Interaktion" gewidmet, ein Nachmittag wurde mit der Mitgliederversammlung verbracht. Damit drängte sich das fachliche Programm der Tagung auf über 100 Vorträge in den verbleibenden 12 Stunden. Es gliederte sich in acht parallel tagende Arbeitsgruppen zu den Themen "Syntax außereuropäischer Sprachen" (Leitung: P.K. Andersen, Bielefeld), "Universalien der Kindersprache" (Leitung: H. Clahsen, Düsseldorf, A. Mills, Tübingen), "Temporalsemantik" (Leitung: V. Ehdorf, Nijmegen, H. Vater, Köln), "Therapeutische Kommunikation" (Leitung: D. Flader, Berlin, A. Koerfer, Essen), "Diskurse bei Kindern: Ihre Struktur und Entwicklung" (Leitung: G. Kann-Delius, Berlin, U. Quasthoff-Hartmann, Bielefeld), "Kooperatives Handeln" (Leitung: F. Liedtke,

Düsseldorf), "Satzmodus zwischen Grammatik und Pragmatik" (Leitung: J. Meibauer, Köln), "Text und Diskursdynamik" (Leitung: W. Wildgen, Bremen, W. Kindt, Bielefeld).

Ein großer Teil der Vorträge hatte mit dem Rahmenthema der Tagung nichts zu tun. Ihm am nächsten kam vielleicht noch das Programm der sechsten Sektion, in der der Begriff der "Kooperativität" im sprachlichen Handeln intensiv diskutiert wurde. Insbesondere die Vorträge von A. Burkhardt (Braunschweig), K. Ehlich (Wuppertal), K. Gloy (Oldenburg), H.J. Heringer (Augsburg), W. Holly (Trier), R. Keller (Düsseldorf), F. Liedtke (Düsseldorf) und K. Mutersbach (Heidelberg) ließen eine Tendenz zur Skepsis gegenüber zu stark idealisierenden Positionen in der Handlungstheorie erkennen und zeugten von einem neu erwachenden Interesse an der "realistischen" Erforschung dialogischen Handelns.

Ernest W.B. Hess-Lüttich, Berlin (überlassen vom Arbeitskreis für Semiotik, Berlin)

AUFDECKUNG UND VERDECKUNG:
AFFIRMATION UND SUBVERSION
IN WORT-BILD-BEZÜGEN

Seminar am Institut für Wissenschaft und Kunst, Wien

In den Monaten April, Mai und Juni 1985 fand am Institut für Wissenschaft und Kunst (Wien, Berggasse 17/1) ein Seminar über "Aufdeckung und Verdeckung: Affirmation und Subversion in Wort-Bild-Bezügen" statt. Es stand unter der Leitung des ÖGS-Vorstandsmitglieds H.L. Handl, der die Problemstellung folgendermaßen umriß: "Wann kommentiert das Wort das Bild, wann illustriert die visuelle Darstellung die Verbalaussage? Welche Zusammenhänge bestehen zwischen diesen beiden Repräsentationen? Was geschieht dort, wo keine Entsprechung, Übersetzung (Metapher), sondern widersprüchliche Willkür oder Irreführung vorliegen? (Die Bildunterschriften der Zeitschrift *Der Spiegel* sind bekannt als Verwirrlied, und Magrittes Bildtitel sind nicht logisch erklärbar...). Kann Wissenschaft ohne Metapher auskommen? Wie wird die anschauliche, bilderreiche Sprache benutzt? Wie wird das Verhältnis von Bild und Text in seiner Wechselwirkung verstanden? "

Den Einführungsabend gestaltete Martin Stegu mit dem Vortrag "Text und Bild in Presse und Werbung - Konventionen ihres Zusammenhanges und deren Durchbrechung". In zeichentheoretischer Kurzeinleitung behandelte er zunächst die Interdependenz- bzw. Prioritätsproblematik ikonischer und sprachlich-arbiträrer Zeichen, um sich einerseits auf die genetische Dominanz sprachlicher Zeichensysteme festzulegen, und andererseits zu betonen, daß es nicht möglich ist, das Phänomen Bild als Problematik vom sprachlichen Kontext abzukoppeln. (Dies führte ihn zu einer grundsätzlichen wissenschaftstheoretischen Problematisierung des Verhältnisses von Linguistik vs. Semiotik). Er konzentrierte sich in der Folge auf den "Wirklichkeitsersatz"-Charakter des Bildes und seiner anschaulich-unterstützenden Funktion in Sprache/Bild-Kombinationen, um daraus die Berechtigung eines Instrumentariums abzuleiten, in welchem die textlinguistische Analyse von semiotischen Ansätzen flankiert wird. Er thematisierte weiters Kategorien wie Kohäsion, Kohärenz, "Weltwissen", Code, "Vertrauensgrundsatz der Kommunikation", erläuterte dies und voriges an einigen Beispielen aus dem werblichen Bereich, und widmete sich schließlich sehr ausführlich der Analyse eines Exemplars des österreichischen Boulevard-Tagblattes *Neue Kronen-Zeitung* (Sams-tagsausgabe, 6. April 1985). Dabei kam auch die Ikonizität des Gesamttextes zur Sprache, weiters die Denotations/Konnotation-Problematik, und schließlich der instrumentell-idiologiekritische Wert zeichentheoretischer Überlegungen und Analysen, die in diesem Sinne an der Gesamtbotschaft des Textes ansetzen müßten. Der Referent betonte

jedoch, daß gerade die Ambiguität von Sprache-Bild-Kombinationen die Rückführung auf klar umreißbare Ideologeme zu einem komplexen und schwierigen Unterfangen mache. Auch gebe es noch kritikwürdigere mediale Sachverhalte als die in üblichen Tageszeitungen qua Zeichensysteme selbst inkorporierten ideologisch-manipulativen Bestandteilen der Botschaften. (In der Diskussion stand sodann die erkenntnistheoretische und genetisch-prioritätsbezogene Problematik von Bild- vs. Sprach-Rezeption im Vordergrund, da die diesbezügliche Festlegung des Vortragenden nicht unwidersprochen blieb).

Der für die Reihe Verantwortliche, *Haimo L. Handl*, folgte mit dem Vortrag "Ceci n'est pas Magritte. Anmerkungen zum Wort-Bild-Bezug" nach. Er behandelte vor allem jene Übergangszonen, in welchen die (schrift-)sprachlichen Zeichen übergehen, sowie die damit verbundenen Probleme der Veränderung der Bedeutung und des Entstehens vielschichtiger Aussage-Komplexe. Einer seiner veranschaulichenden Bezugspunkte war hierbei René Magrittes Text "Die Wörter und die Bilder", dem er eine sehr kritische Reflexion angedeihen ließ, ein weiterer der mit Zeichnungen Magrittes versehene Text "Dies ist keine Pfeife" von Michel Foucault. Im Anschluß an letzteren ging es dann um das Kalligramm, um die "Zweideutigkeit von sprachlichem und visuellem Spiel", um die Konkurrenz (aber auch "Überspielung" derselben) von Diskurs und Bild. Weiters traten Probleme des Abstraktionsprozesses, der Mittelbarkeit der Anschauung, der "Differenz/difference" in den Vordergrund, zum Teil mit Beispielen aus der modernen Kunst (Klee, Kandinsky) bzw. aus "Grenzbereichen" (Aphasie) verbunden. (In der Diskussion gesellte sich auch das Thema des dreidimensionalen Ikonos und der Bedeutungsverschiebung qua Kontextverschiebung hinzu, etwa am Beispiel der "Musealisierung" von Objekten).

Frank Hartmann setzte mit einem sehr ausführlichen Beitrag zum Thema "Bilderverbot und Theorie dialektischer Darstellung" fort, in dem es um die sprachliche Präsentation und Repräsentation von philologisch-wissenschaftlicher Theorie ging, vor allem hinsichtlich der aus dem Adäquanzpostulat abzuleitenden Wahl optimaler Vermittlung. Dies bedingt eine eingehendere Beschäftigung mit der - an sich unumgänglichen - Metaphorizität der Sprache, aber auch eine Unterscheidung von Bild und Abbild. Er stellte in der Folge das seitens der Kritischen Theorie gegenüber dem Purismus des logischen Positivismus eingeklagte "Recht des Bildes" zur insofern immanenterweise zwiespältigen Diskussion, als ja andererseits gerade von Horkheimer und Adorno im Zusammenhang mit deren Kritik der Kulturindustrie *Herzblatt*, eine Persiflage auf die Stilisierung der Gefühle im Genrefilm, vor, und ein weiterer Streifen bildexperimentellen Charakters mit dem Titel *Valium Hitze* (von *R. Müller*) gab Anlaß, die Bedingungen kreativer Semiosen zu untersuchen. *Jeff Bernard* präsentierte eine "Eintagsausstellung" von pikto-diagrammatischen Blättern mit dem Titel "Ikonizität, Relationalität, Arbitrarität" und hielt hierzu ein Impulsreferat, das die Dialektizität der Repräsentationen und der Zeichenproduktion zum Inhalt hatte. Der amerikanische Künstler *Terry Bendt* zeigte eine Dokumentation von neueren Collagen und Assemblagen des Titels "A Novelty - An Altered Novel", in welchen eine spannungsreiche Konfrontation fragmentierter Objekte und verfremdeter schriftlicher Texte intendiert war, die auch de facto neue bislang verborgene Bedeutungsebenen erschloß. Einem ähnlichen Thema war auch die von *Haimo L. Handl* vorgelegte Dia-Dokumentation "Bilderschriften/Schriftbilder" gewidmet, in welcher er im Sinne des exemplarischen Belegs an die Ausführungen seines eingangs erwähnten Vortrags anknüpfte.

Die materialreichen Darbietungen dieses Abschlußtreffens veranlaßten die Seminarteilnehmer, den Gedankenaustausch auch nach Beendigung der Veranstaltung informell, doch exzessiv fortzusetzen, wobei sowohl formale, als auch soziogenetische Fragen des Ausgangsthemas, der Enthüllungs- bzw. Verschleierungsfunktion im Wort-Bild-Bezug, von praktischer Anschauung wieder auf die Metaebene zurückgeführt wurden.

Jeff Bernard, Wien; (überlassen vom Arbeitskreis Semiotik, Berlin)

VOM ZEICHEN ZUM TEXT

Internationale Kolloquium in Israel

Am 20. und 21. Mai 1985 kamen über 40 Wissenschaftler aus 9 Ländern in der Ben-Gurion-Universität des Negev in Beer Sheva (Israel) zusammen, um das Thema *Vom Zeichen zum Text* zu diskutieren. Die Zeichenbezogenheit der Fragestellungen ermöglichte es den Teilnehmern, sowohl die Mauern zwischen den traditionellen Disziplinen Linguistik, Literaturwissenschaft und Literaturkritik einzureißen, als auch die zeitgenössische Etikette "Pragmatik", "Sondersprache", "Diskursanalyse", "Textlinguistik", usw. zu unterlaufen. Das Kolloquium war in englisch- und französischsprachige Gruppen eingeteilt, die ihrerseits eine Reihe von Schwerpunkten setzten. Die englischsprachige Gruppe diskutierte über 40 Vorträge, die in folgenden 4 Schwerpunkten zusammengefaßt waren:

- Zeichen, Kontext, Text in der Sprachanalyse,
- Der Text als Zeichen in der Literatur,
- Vom Zeichen zum Text in besonderen Textsorten,
- Zeichen der Abweichung in Texten,

Die kleinere französischsprachige Gruppe behandelte zwei Schwerpunkte:

- Das sprachliche Zeichen
- Text als Kommunikationsmittel

Jeder Vortrag des Kolloquiums folgte dem gleichen Muster: Er formulierte zunächst ein spezifisches theoretisches oder methodologisches Problem, verankerte dann dieses Problem in der Triade Zeichen - Kontext - Text und ging schließlich zur Analyse eines bestimmten Textes oder Textgenres über. Die gemeinsame Definition von Sprache als Zeichensystem erwies sich als guter Ausgangspunkt für die kontextbezogene Untersuchung von zeichenartigen Elementen in Texten.

Das Kolloquium wurde von *Yishai Tobin* im Institut für Fremde Sprachen und Literaturen der Ben-Gurion-Universität organisiert.

Überlassen vom Arbeitskreis für Semiotik, Berlin

SPRACHTHEORIE: DER SPRACHBEGRIFF IN WISSENSCHAFT UND ALLTAG

IDS-Jahrestagung 1986 in Mannheim

Die Jahrestagung des Instituts für Deutsche Sprache vom 11. bis 14. März 1986 in Mannheim war dem Thema *Sprachtheorie: der Sprachbegriff in Wissenschaft und Alltag* gewidmet. Im theoretisch-systematischen Teil der Tagung (11.3.) sprach *H.H. Lieb* (Berlin) über den "Sprachbegriff in der gegenwärtigen Sprachwissenschaft", *G.P. Baker* (Oxford) über "Moderne Sprachtheorien aus philosophischer Sicht", *R. Keller* (Düsseldorf) plädierte für einen "Evolutionären Sprachbegriff" der "dritten Art" und *P. Hellwig* (Heidelberg) gab einen Überblick über "Sprachtheoretische Auffassungen in der Künstliche-Intelligenz-Forschung und in der Computerlinguistik".

Im zweiten historisch orientierten Teil des Programms (12.3.) informierte *W. Neumann* (Berlin, DDR) über "Handlungstheoretische Auffassungen in der Sprachwissenschaft des frühen 19. Jahrhunderts", *H. Birus* (Göttingen) sprach über das Problem "Grammatische Interpretation: zum Verhältnis von Hermeneutik und Sprachtheorie im 18. Jahrhundert", *L. Jäger* (Aachen) suchte "Die Sprachtheorie Wilhelm von Humboldts" aus dessen umfangreichem Werk herauszufiltern. Fünf Experten widmeten sich in einer lebendigen Podiumsdiskussion unter der Leitung von *D. Cherubim* (Göttingen) den sprachtheoretischen Auffassungen von F. de Saussure, K. Bühler und N. Chomsky: *A. Eschbach* (Essen) informierte dabei u.a. über den Stand der Bühler-Edition, *J. Heringer* (Augsburg) würdigte die Bedeutung Saussures für die Entwicklung der Sprachwissenschaft, *H. Leuninger* (Frankfurt, Main) setzte sich für die Generative Transformationsgrammatik ein. *K. Heger* (Heidelberg) und *W. Oesterreicher* (Freiburg) skizzierten vor dem so entfalteten Hintergrund eigene sprachtheoretische Überlegungen.

Der Duden-Preis der Stadt Mannheim wurde 1986 an *H. Weinrich* (München) verliehen, der sich für die lebenswürdige Laudatio des IdS-Präsidenten *H. Rupp* (Basel) mit einem brillanten Festvortrag revanchierte, in dem er in Anspielung auf Goethes Faust fragte: "Lügt man im Deutschen, wenn man höflich ist?"

Der dritte Teil der Tagung (13.3.) fragte nach dem sprachtheoretischen Wissen von Laien in Literatur und Alltag. *H.P. Althaus* (Trier) nahm sich unter dem allgemeinen Titel "Sprachauffassungen in der belletristischen Literatur" insbesondere die "Etymologie" Arno Schmidts vor, *H. Glinz* resümierte auf dem Boden seiner reichhaltigen Erfahrung, was er in didaktischer Absicht an Sprachtheorien "in und hinter den Lehrern" vermutet und was daraus für "die Entwicklung der Sprachfähigkeiten in den Schülern" folgt. *S. Wichter* (Münster) hatte unter dem Titel "Sprache, Sprachen, Zeichenbegriff" erkundet, was Laien eigentlich mit linguistischen Begriffen verbinden, und *J. Wirrer* (Bielefeld) hatte Urteile von Sprechern des Plattdeutschen über den Gebrauch des Dialektes gesammelt: "Alltagswissen über Dialekte". Zum Abschluß des öffentlichen Teils der Tagung berichtete *G. Stieckel* (IdS, Mannheim) über das Ergebnis einer Mannheimer Zeitungsumfrage: "Was halten Sie vom heutigen Deutsch?" Der letzte Tag (14.3.) war den Mitgliedern des Wissenschaftlichen Rates des Institutes für deutsche Sprache vorbehalten.

Ernest W.B. Hess-Lüttich, Berlin (überlassen vom Arbeitskreis für Semiotik, Berlin)

DIALOGANALYSE

Arbeitsstagung in Münster

Vom 5. bis 7. März 1986 fand im Germanistischen Institut der Westfälischen Wilhelms-Universität Münster eine Arbeitstagung über Dialoganalyse statt. Der Veranstalter dieser Tagung, *F. Hundsnurscher* (Münster) und *E. Weigand* (Bochum), hatten deren Schwerpunkt auf die Aufdeckung dialog-"grammatischer" Strukturen, d.h. vor allem auf die Analyse von Dialogmustern gelegt und dementsprechend den "Musterbeschreibungen" zwei der insgesamt neun Sektionen gewidmet. Um dem Vordringen der Computertechnik auch in Bereiche der linguistischen Datenerhebung, -auswahl und -analyse Rechnung zu tragen, waren zwei weitere Sektionen auf die "automatische Dialoganalyse" zugeschnitten. Die übrigen Sektionen befaßten sich mit "Methodologie der Dialoganalyse", "Dialogtypologie", "Strategien", "grammatischen Aspekten sprachlicher Ausdrucksmittel", "Kommunikationstherapie" sowie "Dialogmaximen".

Für die Verbindung von Sprachanalyse und Interaktionsanalyse sowie dafür, daß neben natürlichsprachlichen auch fiktive (literarische) "Gesprächstexte" im Hinblick auf eine "ideale Dialoganalyse" zu berücksichtigen seien, plädierte *A. Betten* (Regensburg). *T. Bliessner* (Köln) befaßte sich mit methodologischen Problemen bei der "Entdeckung" von Dialogmustern. Anhand von Ergebnissen einer Untersuchung aufgabenorientierter Dialoge, illustrierte *P. Bosch* (Nijmegen), in welcher Weise formale Merkmale sprachlicher Äußerungen mit den kognitiven Modellierungen des Äußerungskontextes von Sprecher und Hörer interagieren. Wie die Argumentationsstrategien von Teilnehmern an einem "öffentlichen" Gespräch beschrieben werden können, zeigte *K. Brinker* (Hamburg) an einem Ausschnitt aus einer Fernsehshow auf. Mit Fragen des Verursachens sprachlichen Verhaltens und der sich daraus ergebenden These von der relativen Unwichtigkeit des genauen Wortlauts von Formulierungen für den Hörer befaßte sich *B. Geurts* (Nijmegen). *J. Graffe* (Münster) zeigte auf, wie im Fremdsprachenunterricht durch die "Konfrontation" mit verschiedenen Dialogzusammenhängen die Fähigkeit zu sprachlichem Handeln erweitert werden könnte. Als ein Beispiel für die Sprechaktsequenz "Informieren/auf Informationen reagieren" untersuchte *G. Hindelang*

(Münster) das medizinische Aufklärungsgespräch. Als Programm zur Dialogforschung erläuterte *F. Hundsnurscher* (Münster) den Zusammenhang einer allgemeinen Dialogtypologie mit (Analysen von) als Realisierungen von Gesprächsmustern begriffenen authentischen Gesprächstexten. *E.W.B. Hess-Lüttich* (Berlin) erläuterte Grundbegriffe einer Dialoglinguistik. Mit welchen Techniken Interviewer in englischen Nachrichteninterviews auf die durch deren institutionellen Rahmen vorgegebenen Beschränkungen für das Widersprechen zu reagieren versuchen, illustrierte *A.H. Jucker* (Zürich). *G. Kalivoda* (Kassel) befaßte sich mit den dialogstilistischen Aspekten der parlamentarischen Debatte, und zwar insbesondere mit Fragen zur Rekonstruktion und Beschreibung des für Dialogtypen dieser Art charakteristischen Redestils. Mit dem Einfluß der Ziele, die Gesprächsteilnehmer (allgemeinen handlungs- und dialoganalytischen Überlegungen zufolge) zu erreichen beabsichtigen, auf die internen Vorgänge in und den Ablauf von Gesprächen beschäftigte sich *M. Kohl* (Münster). *M. Kohrt* (Münster) erläuterte die wissenschaftstheoretischen und methodologischen Konsequenzen der Konzeption "Dialoggrammatik". *H.W. Kramski* und *W. Zillig* (beide Münster) stellten Überlegungen im Hinblick auf ein Computerprogramm zur Simulation von Planungsgesprächen vor. *H. Kreye* (Bremen) führte vor, wie sich mit Hilfe einer EDV-Auswertung ein Textprofil für einen Text von H. v. Kleist darstellen läßt. *J. van Kuppevelt* (Nijmegen) präsentierte einen Entwurf hinsichtlich der Art, in der "sentence topics" zum Diskurs in Beziehung gesetzt werden könnten. *F.W. Liedtke* (Düsseldorf) setzte die Überlegungen von *A. Schütz* zum Problem der Relevanz in Beziehung zu der Griceschen Konversationsmaxime der Relation. Einen Einblick in die Struktur kommunikativer Handlungen sowie in das Programm einer handlungstheoretischen Semantik verschaffte *G. Meggle* (Münster). *N. Montulet* (Nijmegen) stellte Überlegungen zu Fokus, Kontext und definiten Kennzeichnungen in Dialogsituationen vor. Anhand von Tonbandaufzeichnungen mündlicher Prüfungen (zum Staatsexamen, zum Abitur) analysierten *B. Naumann* (Erlangen) und *H.-M. Dederding* (Nürnberg) initiiierende Gesprächssteuerungsmittel. *W. Niehüser* (Münster) zeigte auf, mit welcher Art von sprachlichen Ausdrucksmitteln Sprecher der Maxime "Kritisiere dich selbst, bevor du von anderen kritisiert wirst" zu folgen versuchen. Daß die Wahl zwischen alternativen Formulierungen auf der Grundlage spieltheoretischer Überlegungen zu "erklären" sei, versuchte *J. Ossner* (Ludwigshafen) nachzuweisen. *R. Posner* (Berlin) offerierte handlungslogische Analysen zum Signal, zum Anzeichen, zur Geste, zum Gestikulieren sowie zum Mitteilungsversuch. Eine Verallgemeinerung der Griceschen Konversationsmaxime wurde von *E. Rolf* (Münster) vorgestellt. *A. Rothkegel* (Saarbrücken) informierte über die Struktur des Dialogs mit Computern sowie über den Umstand, daß bei aufgabenorientierten Dialogen auch (in übergeordnete Textmuster eingebettete) sprachliche Interaktionsmuster eine Rolle spielen. *S.F. Sager* (Hamburg) trug Überlegungen vor, die sich auf die phylogenetische Herleitung von Dialogmaximen bezogen, von denen ihrerseits anzunehmen ist, daß sie solchen interaktiven Ereignissen wie Höflichkeitsritualen zugrunde liegen. *M. Schecker* (Freiburg i. Br.) befaßte sich mit dem Insistieren und einem davon zu unterscheidenden Resistieren unter dem Aspekt strategischen Kommunizierens. Mit dem "Werbecharakter" von Gesprächen, die sog. Pharmareferenten mit niedergelassenen Ärzten zu führen versuchen, beschäftigte sich *K. Schonauer* (Köln). An zahlreichen Beispielen vorwiegend aus den Bereichen "Werbung" und "Politik" machte *K. Sornig* (Graz) auf die persuasiven Eigenschaften der entsprechenden Sprachverwendung aufmerksam. *W. Sucharowski* (Kiel) stellte Ergebnisse automatischer Dialoganalysen vor, die sich auf das an geistig behinderte Kinder adressierte Erzählen biblischer Geschichten bezogen. Durch Betrachtungen hinsichtlich der Delimitation von Texten und Dialogen versuchte *G. Tschauder* (Bochum) die "These von der generellen Dialogizität der Sprache" zu rechtfertigen. Mit Fragen der Verfremdungsstrategie Frank Wedekinds beschäftigte sich *E. Weidl* (Berlin). Daß (und inwiefern) auch Texten so etwas wie "Dialogizität" zuzuschreiben sei, dafür plädierte *E. Weigand* (Bochum). *R. Weingarten* (Bielefeld) erörterte die These, daß Untersuchungen bestimmter Dialoge einen neuen (empirischen) Zugang zu Problemen der Wortbedeutung und ih-

rer Analyse eröffnen würden. *O. Werner* (Freiburg i. Br.) legte dar, inwiefern Probleme der Referenz von Eigennamen in Dialogen, Gesprächsstrategien und -störungen ihren Niederschlag finden.

Eckard Rolf, Universität Münster (überlassen vom Arbeitskreis für Semiotik, Berlin)

ner anderen vieldiskutierten Gegenüberstellung und argumentierte, daß die Sprache sich wohl kaum entwickelt hätte, wenn es den frühen Menschen nur um eine Verbesserung der in Mimik und Gestik ja schon hoch entwickelten kommunikativen Interaktion gegangen wäre. *Juan D. Delius* (Bochum) analysierte in seinen "Überlegungen zum Ursprung der Sprache" die neurophysiologischen Voraussetzungen des Sprechens. Er wies auf die Wichtigkeit der Kategorienbildung, des latenten Lernens und des Nachahmungslernens für die sprachliche Abbildung der Welt hin und berichtete von seinen Experimenten mit Tauben, die in dieser Hinsicht bereits verblüffende Leistungen erbringen. *Horst Singer* (Bochum) problematisierte in seinem Referat über "Die Sprache des Hundes oder die 'Sprache' des Hundes" eine detaillierte Analyse des Zeichenrepertoires der Hunde und seiner kommunikativen Funktionen. *Holger Preuschhof* (Bochum) setzte dieses Thema mit seinem Referat über "'Sprache' oder 'Schrift' bei unseren nächsten Verwandten" fort und *Helmut Jachnow* (Bochum) stellte mit seinen Überlegungen zu den Erscheinungsweisen und Ursachen des Wandels nonverbalen kommunikativen Verhaltens" den Zusammenhang mit dem Menschen her. *Ilona Schwarzerlsing* (Bochum) war die erste, die die mögliche zeitliche Reihenfolge der verschiedenen Stadien der Sprachrevolution thematisierte, indem sie "Eine kritische Würdigung von Julian Jaynes' Modell des Sprachursprungs" gab. Weitere Einwürfe zu den Stadien der Sprachrevolution folgten in dem Referat von *Eric de*

Grolier (Paris) über "The Evolution of Language during the Homo Sapiens Period", in dem Referat von *Bernard Bichakjian* (Nimwegen) über "Linguistic Paedomorphosis" (die These, daß sich in der Sprachentwicklung jeweils dasjenige Verhalten durchsetzt, das von den Kindern beim Spracherwerb zuerst ausgebildet wird), in dem Referat von *Jan Wind* (Amsterdam) über "The Evolution of the Human Vocal Tract", in dem Referat von *Udo L. Figge* (Bochum) über "Aspekte der Sprachentstehung" (das die Stadien des Werkzeuggebrauchs im Mensch-Tier-Übergangsfeld in den Vordergrund stellte) und in dem Referat von *Walter A. Koch* über "Ding Dong II: From Icon to Symbol in The Phylogeny of Human Language". Koch stellte seinen großangelegten Versuch einer Synthese aller wichtigen vorhandenen Ansätze zur Erklärung der Sprachentstehung zur Diskussion und steckte damit einen Rahmen ab, auf den auch alle weiteren Überlegungen der Tagungsteilnehmer bezogen waren. Es folgten kritische Sichtungen der Hinweise, die die heutige Linguistik für Hypothesen über den Sprachursprung geben kann. Wie sich in der Referaten von *Michael Job* (Bochum) über "Struktur und Eigenschaften von Protosprachen", von *Thomas Stolz* (Bochum) über "Morphologische Natürlichkeit als sprachwandeldeterminierender Faktor" und von *Norbert Boretzky* (Bochum) über den "Kreofall: Sprachgenese oder Interferenz?" zeigte, glauben die Sprachwissenschaftler zwar, Prinzipien für den Sprachwandel angeben zu können, aber nicht mit diesen Prinzipien die Sprachentstehung erklären zu können. Die letzte Gruppe von Referaten widmete sich der Ontogenese der Sprache und zeigte damit, welche heute beobachtbaren Voraussetzungen der Sprechens sich in der Sprachrevolution einstellen mußten, damit die menschliche Sprache entstehen konnte. *Marianne Kugler-Krumme* (Bochum) untersuchte am Beispiel von Pantomime und Gestenkode der Kinder "Die Entwicklung spontan gebildeter Zeichensprachen". *Susan Vogel* (Bochum) analysierte "The Ontogenesis of the Reality-Fiction Distinction", *Jan Wind* dokumentierte auf Photos und Videoaufnahmen "Empathetical Facial Movements of Parents Accompanying Children's Speech", und *Michael Vandenbosch* (Brüssel) berichtete über das von *Roland Corluy* (Brüssel) geleitete belgische "Child-Ape Communication Project". Die von *Peter L.W. Koch* (Bochum) im Rahmenprogramm eingeleitete Vorführung des Films "La guerre du feu" stellte die Verbindung zu den populärwissenschaftlichen Versuchen einer Darstellung der Menschwerdung her. Wie sich in der Abschlusdiskussion zeigte, hat dieses Kolloquium durch die gelungene Verbindung von inspirierenden Fragestellungen und soliden Forschungsberichten allen Teilnehmern Denkanstöße gegeben, die die weitere Arbeit bestimmen werden. Auskunft erteilt: Prof. Dr. Walter A. Koch, Englisch Seminar (GB 6/129), Ruhr-Universität, Postfach 102148, D-4630 Bochum 1.

Semiotischer Arbeitskreis, Berlin

GENESE DER SPRACHE

Tagung in Zusammenarbeit mit dem Bochumer semiotischen Colloquium

Vom 5. bis 7. Februar 1986 fand in Bochum eine Tagung zur Genese der Sprache statt, das von *Walter A. Koch* in Verbindung mit dem Bochumer Semiotischen Colloquium veranstaltet wurde. Die etwa 50 Teilnehmer aus Deutschland und dem westeuropäischen Ausland aktualisierten auf der Grundlage von zwanzig 45minütigen Referaten die heute verfügbaren biologischen und historischen Befunde, die die Bildung von Hypothesen über die Entstehung der menschlichen Sprache ermöglichen. Zu Beginn stellte *W.A. Koch* die Kontinuitätsthese, nach der die Sprache im Tier-Mensch-Übergangsfeld allmählich aus Mimik, Gestik und Lautäußerungen hervorgegangen ist, der älteren Diskontinuitätsthese gegenüber, die der Sprache gegenüber allen vorher entstandenen Zeichensystemen einen Sonderstatus einräumt. Er schlug vor, im Rahmen des Kolloquiums nach Wegen zu suchen, die Argumente für beide Thesen zu einer vermittelnden Position zu verbinden und dabei die heutigen Einsichten in die Ontogenese, Pathogenese, Aktualgenese und Sprachtypologie einzubeziehen. *Elmar Holenstein* (Bochum) beschäftigte sich in seinem Referat "Vom Ursprung der Sprache - kommunikativ oder kognitiv?" mit ei-



ESPRIT-Projekt P393 (ACORD)

Construction and Interrogation of Knowledge Bases using Natural Language Text and Graphics

Forschungsstelle:

TA Triumph-Adler AG, Neue Technologien/Basisentwicklung, Fürtherstr. 212 8500 Nürnberg

Mitarbeiter:

Dr. Gerhard Heyer
Ralf Kese
Dr. Rainer Lutze

Projektpartner:

Laboratoires de Marcoussis, Paris, (Projektführer)
Groupe Bull, Paris
Universität Stuttgart Institut für Romanistik/Linguistik (IRL)
Fraunhofer Gesellschaft IAO, Stuttgart
University of Edinburgh, Center for Cognitive Science, Dept. of Architecture (EdCAAD)

Dauer: 1. 1. 85 - 31. 12. 89

Projektziele

Gegenstand des ACORD-Forschungsvorhabens ist die Konstruktion und Befragung von Wissensbasen im Bereich sog. *business communications* unter Verwendung natürlichsprachlicher Texte und Graphik.

Im einzelnen sind von linguistischer Seite die Festlegung einer Grammatik des intendierten Anwendungsbereichs fürs Deutsche, Französische und Englische und der Bau eines Parsers vorgesehen. Zur semantischen Verarbeitung des Textes (und der Graphik) ist die Definition einer für den Anwendungsbereich geeigneten semantischen Repräsentationssprache und Wissensbasis, sowie die Konstruktion eines Dialogmanager erforderlich, der die pragmatische Seite des Dialogs in geeigneter Weise zu behandeln gestattet.

Ausgewählte Entwicklungsschwerpunkte

Als erstes Anwendungsgebiet wurden Geschäftsberichte, Geschäftsmitteilungen und Geschäftsgraphiken, wie sie sich typischerweise in Zeitungsnotizen finden, ausgewählt. Die Komplexität freien natürlichsprachlichen Texts, die thematisch schwierige Eingrenzbarkeit des Gegenstandsbereichs und die ge-

ringen Möglichkeiten, Geschäftsgraphiken direkt zu verändern, haben jedoch eine Umdefinition des Anwendungsbereichs erforderlich gemacht. Als neues Anwendungsgebiet wurde die Logistik im Transportwesen ausgewählt, womit ACORD als eine graphikunterstützte natürlichsprachliche Schnittstelle zu Experten- oder Decision-Support-Systemen angesehen werden kann. Für das Deutsche wurde ein top-down Parser in PROLOG auf der Grundlage des LFG Systems implementiert. Für das Englische wurde in Edinburgh, ausgehend von PATR II, eine Unification Categorical Grammar (UCG) entwickelt, die mit einem Chart-Parser implementiert wurde. Im Französischen wird mit String Grammars gearbeitet, die als DCG implementiert sind. Sei es mit oder ohne die Hilfe des Zwischenschritts einer Generierung von F-Strukturen liefern alle Parser DRSen als Ergebnis zur weiteren Verarbeitung ab, wobei in ACORD mit einer um Sorten erweiterten DRS ("InL" für "Indexed Language") gearbeitet wird. Die wiederum auf einer Erweiterung von PROLOG basierende (hybride) Wissensbasis kann schließlich als Modell für die semantische Repräsentationssprache InL verstanden werden, das über den Knowledge-Base-Manager verwaltet wird.

Problemschwerpunkte

Es ist aufschlußreich, daß die Hauptschwierigkeiten bei der Integration aller Partnersoftware in einen 1. Prototypen (bis Februar 1987) im Bereich des Lexikons, der Morphologie und des Zusammenwirkens von Lexikon und Wissensbasis auftreten. Zur Lösung dieser Schwierigkeiten wird in ACORD die Idee eines Protolexikons verfolgt, das als neutraler Formalismus die Grundlage der einzelnen, einer bestimmten Notation verpflichteten Lexika der einzelnen Sprachen bildet.

Ausgewählte Forschungsberichte

Evaluation of Different Parsers and Syntactic Theories, Bull and IRL Stuttgart, public

Montague Grammar, Situation Semantics, and Discourse Representation Theory: A Comparison of Three Semantic Theories, IRL Stuttgart, public
Deductive Natural Language Processing for DRs, Bull, public

Evaluation of Semantic Theories Used in Existing Systems and Evaluation of Semantic Theories that are not yet implemented, Bull, public

Problems of Dialogue Parsing, CCS Edinburgh, Esprit
Temporal Information and Natural Language Processing, CCS Edinburgh, Esprit

Incremental Interpretation in Dialogue, CCS Edinburgh, Esprit
A Specification of InL, CCS Edinburgh, ACORD intern
The Protolexicon, CCS Edinburgh, ACORD intern

Research on the Semantics of Deictic Phenomena and Natural Language, FhG/IAO, Esprit

Research on the Semantics of Natural Language and Visual Situations: The Semantics of Perception Reports, FhG/IAO, public

PROJEKTE

Investigation into the Literature on Combined Graphics/Textual Systems, *FhG/IAO*, Esprit
Investigation of Direct Manipulation, *FhG/IAO*, Esprit
Investigation of the Dialogue Manager, *FhG/IAO*, ACORD intern
Review of Interactive Computer Graphics Literature, *Ed-CAAD Edinburgh*, ACORD intern
Investigation of the Scope of a Knowledge Base for Business Communication, *TA*, public

Evaluation of Necessary Properties of Graphical Objects, *TA*, public
Remarks on Representing Questions in Predicate Logic, *TA*, ACORD intern
Extending PROLOG for Processing Natural Language Semantics, *TA*, ACORD intern

Fachbereich 20 (Informatik)
Forschungsgruppe
Datenverwaltungssysteme II



PROJEKT WeBeS:

**Zugang zu Werkstoffdatenbanken:
Benutzerforschung und Systementwurf**

**Bundeswirtschaftsministerium fördert Projekt
an der TH Darmstadt**

Forschungsstelle

Fachbereich Informatik, Technische Hochschule Darmstadt, *Karolinenplatz 5, D-6100 Darmstadt*

Projektleitung

Gerhard Knorz

Mitarbeiter

Karin Ammersbach
Norbert Fuhr
Roswitha Jäger-Beck
Barbara Kienitz-Vollmer

Projektpartner

Fachinformationszentrum Werkstoffe, Berlin
Fachinformationszentrum Technik, Frankfurt
Betriebsforschungsinstitut, Düsseldorf

Förderung

gefördert vom Bundesministerium für Wirtschaft unter dem Kennzeichen ZD3 - 00 30 55/10 -

Projektdauer

August 1986 bis Juli 1987

Projektbeschreibung

Vom 1.8.86 bis 31.7.87 wird an der TH Darmstadt das Projekt zu Werkstoffdatenbanken: Benutzerforschung und Systementwurf (WeBeS) vom Bundeswirtschaftsministerium gefördert. Primäre Zielrichtung des Projekts ist die verbesserte Gestaltung von Werkstoffdatenbanken (WDBn).

Eine Überprüfung der gegenwärtigen Situation im WDB-Sektor zeigt, daß—obwohl Anbieter und (potentielle) Nutzer den Wert von Fachinformationen dieser Art zunehmend erkennen—mindestens drei Probleme einer erweiterten Nachfrage von Werkstoffinformationen entgegenwirken:

– Das Faktenretrieval stellt Anforderungen an Datenbanksysteme, die von den heute verwendeten Systemen nur zum Teil erfüllt werden (z.B. Behandlung von Wertetabellen anstelle skalarer Werte, Intervalle von Werten, Werte mit Angabe der physikalischen Einheit, fehlende Werte und Parameter bei einzelnen Datensätzen). Dadurch ergibt sich eine partiell recht umständliche Bedienung des Systems für den Benutzer.

In der Regel hat der (potentielle) Nutzer einer WDB Kenntnisse über sein spezifisches Werkstoffproblem. Will er ohne Vermittlung eines Retrievalfachmannes eine Recherche durchführen, braucht er neben der Kenntnis der Anfragesprache und der Datenbankstruktur zusätzliches Wissen über spezielle Charakteristika der gespeicherten Datenmenge (z.B. zugrundeliegende Normungsvorschriften, Herkunft der Daten). Der mit der Aneignung dieses Wissens verbundene Zeit- und Kostenaufwand hat eher negative Auswirkungen auf den Einsatz von WDBn als Informationsquelle.

– Die heute auf dem Markt befindlichen WDBn enthalten nur wenige anwendungsorientierte Informationen, die in ihnen gespeicherten Daten sind eher auf die Bedürfnisse von Werkstofflieferanten und -spezialisten zugeschnitten.

Die genannten Problembereiche markieren die Ausgangsbasis für die Präzisierung des projektierten Forschungsziels, nämlich den Entwurf eines benutzerfreundlichen Zugangs zu WDBn. Welche Forderungen an zukünftige Benutzerschnittstellen für diese Art von Faktenretrievalsystemen zu stellen sind, ist jedoch nur auf der Grundlage empirisch fundierter Forschung zu beantworten. Das Erfassen der Realität heutiger Werkstoffdatenbankpraxis erfordert ein Untersuchungsdesign, das zum einen diese Realität möglichst genau abbildet (Validitätskriterium für Untersuchungsergebnisse) und zum anderen einen möglichst großen Erkenntnisgewinn im Hinblick auf forschungsleitende Ziele mit sich bringt.

Neben Interviews mit (potentiellen) Nutzern von Werkstoffinformationen werden deshalb auch Online-Recherchen mit Hilfe einer Videokamera aufgezeichnet. Als Teilnehmer an einer solchen Recherche sind ein professioneller Rechercheur und ein in-

dustrieller Anwender mit einem "Werkstoffproblem vorgesehen. Die Auswertung der Aufzeichnungen soll mit Hilfe eines elaborierten Analyseschemas vorgenommen werden.

Die Auswertung der protokollierten Recherchen soll zunächst Aufschluß darüber geben, welche Klassen von Informationen der Endanwender benötigt und aus welchen Quellen er diese erhält (WDB, Rechercheur, zusätzliche Hilfsmittel) und welche Interaktionsformen benutzt werden. Darauf aufbauend soll ein benutzerorientiertes konzeptionelles Datenmodell entworfen werden, das die Struktur dieser Informationen und die darauf anwendbaren Operationen beschreibt. Ausgehend von diesem Datenmodell sind Anforderungen an die Datenbankarchitektur und den Inhalt (zukünftiger) WDBn abzuleiten. Auf der anderen Seite müssen für die Elemente und Operationen des Datenmodells benutzerfreund-

liche Repräsentationen entwickelt werden, wobei insbesondere auch graphische und natürlichsprachige Ein- und Ausgabemöglichkeiten berücksichtigt werden sollen.

Bei der Beurteilung, welche Datenbankinhalte und Systemschnittstellen angemessen sind, ist die Orientierung an dem Interessen und Bedürfnissen unterschiedlicher Benutzerklassen unerlässlich. Ziel des Projekts ist es, diese Benutzerklassen zu identifizieren und entsprechende Konzeptionen für benutzerfreundliche Zugänge zu entwickeln. Der für die unterschiedlichen komplexen Systemkonzeptionen notwendige Forschungs- und Entwicklungsaufwand soll grob abgeschätzt werden.

WeBeS, Fachbereich Informatik an der Technischen Hochschule Darmstadt.

Anzeige

MProlog Der sichere Weg zur intelligenten Problemlösung.

Bei der Übersetzung natürlicher Sprache, bei der Fehlersuche in technischen Geräten oder zur Investmentberatung:

In weltweit über 800 Installationen wird mit **MProlog** künstliche Intelligenz vor Ort eingesetzt.

MProlog hält den Kopf frei für die rein inhaltliche Darstellung des Problemgebiets. Ablauftechnische Details treten in den Hintergrund.

Logische Schlussfolgerungen aus aktuellen Fakten geben Antwort auf immer neue Fragestellungen.

DCGs, leistungsstarke Stringverarbeitung, rücksetzbare Eingabe, beliebige Umwandlung von Termen unterstützen speziell linguistische Anwendungen.

Wir von epsilon bieten Ihnen den sicheren Hintergrund aus Wissen und Erfahrung zur intelligenten Lösung Ihres Problems.

für PC bis Großrechner

kostengünstige Trennung von Entwicklungs- und Einsatzumgebung

Anschluß an Fortran, Pascal, C, Assembler

Programmierungsumgebung mit sprachorientiertem Editor

über 140 eingebaute Prädikate

Fehlersuche mit interaktivem Debug-System

bequeme Wartung durch modularen Programmaufbau

epsilon

epsilon Gesellschaft für Softwaretechnik und Systementwicklung mbH

Kurfürstendamm 188
1000 Berlin 15
Tel. 030-882 69 91

PROJEKT SEMSYN

INSTITUT FÜR LINGUISTIK & INSTITUT FÜR INFORMATIK

PROJEKT SEMSYN-86 ALS NACHFOLGER VON SEMSYN-83

Die Fortsetzung der erfolgreichen Arbeiten an der Universität Stuttgart bis 1988 gesichert

Mit Bescheid vom 28. 7. 1986 hat der BMFT rückwirkend das Projekt *SEMSYN-86* bewilligt, das interdisziplinär von den Instituten für Linguistik (Prof. Rohrer) und Informatik (Prof. Gunzenhäuser) der Universität Stuttgart durchgeführt werden wird (Laufzeit von März 86 bis September 88, maximal 3.5 wiss. Mitarbeiter).

Im Vorgängerprojekt SEMSYN-83 wurde das bisher umfangreichste System zur automatischen Generierung von Deutsch aus semantischen Strukturen implementiert und durch Kombination dieses Generators mit einem Parser für Japanisch (ATLAS/II-System von FUJITSU) das erste maschinelle Übersetzungssystem für das Sprachenpaar Japanisch/Deutsch geschaffen.

SEMSYN-86 soll aufbauend auf diesem Ergebnis eine umfassendere linguistische Fundierung erreichen und weitere Einsatzmöglichkeiten für

das Generierungssystem im Rahmen sowohl japanisch/deutscher als auch innereuropäischer Übersetzungssysteme (EUROTRA) schaffen. Der Schwerpunkt der Projektarbeit liegt auf der Übersetzung komplexer Nominalphrasen, wie sie in Titeln und in technischen Texten vorkommen.

Erweiterte Versionen des SEMSYN-Generators wurden zur Generierung kompletter Texte eingesetzt: Generierung von Zeitungsmeldungen (SEMTEX), Generierung von beschreibenden Texten zu geometrischen Konstruktionen (GEOTEX). Experimente zur Generierung von Hilfetexten sind geplant.

EUROTRA-D-kompatibles SEMSYN-System

In der Zeit vom 6. bis 10. Oktober 1986 weilten Vertreter von EUROTRA-D zu einem Arbeitsbesuch beim SEMSYN-Projekt: Dr. Hans Haller und Birgit Weck wollten zunächst den SEMSYN-Generator en detail studieren; der Studienphase schloß sich ein mehrtägiges Experiment an: Das SEMSYN-System wurde so ergänzt, daß es nun auch aus den von EUROTRA-D favorisierten semantischen Strukturen nach Steiner generieren kann. Ein Arbeitsbericht über dieses gemeinsame Experiment ist in Arbeit.

Nachsatz

Am 26. September 1986 hat Dietmar Rösner über "Ein System zur Generierung deutscher Texte aus semantischen Repräsentationen" promoviert. Die Redaktion gratuliert herzlich.

LINGUISTIK TOOLS LABORATORY

Projekt zur Entwicklung von PC-Werkzeugen an der Universität Erlangen

Forschungsstelle:

Universität Erlangen-Nürnberg RRZE, Martensstr. 1, D-8520 Erlangen. Network: Goerz@SUMEX.ARPA, GOERZ@DERRZE1.BITNET

Mitarbeiter:

Dr.G. Goerz
Studenten

Status: *Startphase*

Projektbeschreibung in Stichwörtern:

Parsing algorithms; CFG parser; ATN's; cascaded ATN's; context-free parsers; chart parsers; PSG; unification grammars; PATR; knowledge repre-

sentation tools; KR-schemata; KR-languages; frame system: FRL; logic-based system: DUCK; object-oriented system: FLAVORS; associative networks; truth maintenance.

Die Mainframe-Version eines portablen ATN-Systems und eines PSG-Parsers sind fertiggestellt, jedoch noch nicht auf PC's verfügbar. FRL, DUCK und Flavors dagegen sind bereits auf dem PC implementiert.

G. Goerz, Universität Erlangen-Nürnberg.

Electronic Scholar's Resource Guide (ESRG) Ein interdisziplinäres und internationales Projekt

Der Einzug des Rechners in Lehre und Forschung der verschiedensten Disziplinen in den Universitäten wird sich in den kommenden Jahren mit Sicherheit fortsetzen und es sind nicht wenige, die meinen, daß die daraus erwachsenden Änderungen und Fortschritte von größter Bedeutung sein werden. Auf Initiative von Prof. *Joseph Raben*, der neben anderen Zeitschriften bei *Paradigm Press Computers and the Humanities* oder *Scope* herausgibt, hat sich ein internationales Komitee zusammengefunden, um **Electronic Scholar's Resource Guide (ESRG)** als ein interdisziplinäres Projekt durchzuführen: Auf der Grundlage einer Fragebogenerhebung soll ein "Referenz Tool" zusammengestellt werden, das Oryx Press, Phoenix AZ als Buch und Online-Datenbank anbieten wird. Die Datenerhebung und -Auswertung soll bis zum 1. März abgeschlossen sein. Institutionen, die sich hinsichtlich einer Beteiligung angesprochen fühlen, jedoch noch keinen Fragebogen erhalten haben, sollten sich bitte *umgehend* bei der Redaktion des LDV-Forum melden, die für den deutschen Sprachbereich zuständig ist. Weitere Mitglieder des ESRG-Komitees sind *B. Brainerd, L. Burnard, Ch. Delcourt, L. Farina, M. Gaunt, J. Hughes, H. Karlgren, A. Oikawa, R. Ortali, J. Raben; V. Stevens, V. Vincent* und *A. Zampolli*.

Erklärtes Ziel des ESRG-Projektes ist es, die Situation all diejenigen Studenten, Lehrenden und Forschern zu verbessern zu helfen, die sich weltweit des Potentials des Computers bewußt werden, denen aber eine Informationsquelle fehlt, die ihnen den Zugang zu einer effektiven Nutzung erschließt.

Gerhard Knorz, Technische Hochschule Darmstadt, Fachbereich Informatik, D-6100 Darmstadt

SOFTWARE - COURSEWARE

CONDUIT (Oakdale Campus, University of Iowa, Iowa City IA 52242) is a consortium that distributes a wide range of courseware. For example for language study, both English and foreign, *CONDUIT* offers the following programs (among others):

- *DASHER* allows instructors to create **language drills** for elementary and intermediate students. The package includes programs and text materials needed for creating exercises in English, French, **German**, or Spanish.
- English: Basic mechanics consists of a series of five modules that provides 100 lessons covering the **fundamentals of English grammar and punctuation**

In German Teleconferencing, a **self-contained German instruction program** intended to replace the traditional textbook, all instructions and statements are in German, with an English translation available. It provides multiple-choice drills in grammatical concepts and terminology, exams, an English-German, German-English lexicon, English-to-German and German-to-English passages for translation. Further details from *Yoshio Tanaka, GT study group, Foreign Language Dept., San Diego State University, San Diego CA 92182*.

The following programs are available from Intellectual Software, 798 North Avenue, Bridgeport CT 06606:

- How to do research, a **step by step interactive tutorial on research methods**, includes such topics as notes, outlines, Readers Guide to Periodical Literature, bibliographies, card catalogue, and research in computerized databases.
- German Verbs: The Present Tense **teaches stem formation from the infinitive of regular verbs** and the ending, associated with the present tense. Also gives rules for use of the familiar and formal forms of the German pronouns.
- German Comprehension provides timed fill-in-the-blank tests of reading comprehension, and **teaches phrases, words and common interrogatives** often needed when travelling.

aus: *Scope 4(1986)1: S. 3*

OPTICAL DISCS FOR BIBLIOGRAPHIC DATA BASES

Atari has demonstrated a **compact (12 cm.) read-only memory (ROM) disk player** which uses machinery similar to that in compact disk home stereos, and has announced software by *Activenture* that can access data on a 550M optical disk storing encyclopedia information. The price for the compact disk player is \$ 600; it is available from *Atari Personal Computer Systems, 1265 Borregas Ave, Dept C, Sunnyvale CA 94086*.

The Library Corporation now has available a **cataloguing system** using the new compact disk technology that provides local access to all Library of Congress English and foreign language MARC bibliographic records. Further information from the *Library Corporation, Bibliographic Database Publisher, P.O.B. 40035, Washington D.C. 20016*.

BYU

CONCORDANCE

BYU Concordance is a text retrieval software package that allows rapid access and retrieval of data (from ASCII DOS text files). BYU Concordance includes two programs: BYUINDEX and BYUVIEW. BYUINDEX reads a text file and creates indexes. BYUVIEW uses these indexes and the text file to rapidly retrieve and display the text.

Features of BYU

1. Retrieve quotable data from literary, scientific, historical, or religious documents.
2. Retrieve text from private collections of information (e.g., 3x5 cards, papers, journals, etc.).
3. Create a comprehensive index to textual materials. The index can be appended to the source text.
4. Access archival records such as the minutes, memos, policies, or personal records of one or more organizations.
5. Allow rapid retrieval of product maintenance reports, product problem reports, or product upgrade information for use by customer service centers.
6. Create a concordance showing keywords-in-context. The concordance can be either printed or further modified by a word processor.
7. Examine documents for mis-spelled words or names.
8. Retrieve data, such as potential customer names, from and unformatted text source.

9. Access bibliographical data, abstracts or quotable material from a textual data base created to archive such material.
10. Study word and phrase distribution within literary documents.
11. Compare different versions of a text (e.g., different languages, different sources, different time periods, etc.).
12. Compare different textual materials thematically.
13. Synthesize information from sources such as almanacs, encyclopedias, dictionaries, thesauruses, periodicals, catalogs, journals, etc.
14. Analyze the word and phrase patterns of any textual materials.
15. Retrieve word usage examples from foreign language texts to be used by translators and language students.
16. Study historical data, such as census records, genealogical information, etc.
17. Create 'variable' usage indexes for computer programs.

Price: 684.- DM

Marketing representative Europe

**Institut für angewandte
Kommunikations- und
Sprachforschung e.V.**

Poppelsdorfer Allee 47
D-5300 BONN 1
Tel.: 02 28/73-56 45
73-56 20

The **System for Digital Storage and Retrieval** is a project being developed by two researchers at the Smithsonian Institution's Air and Space Museum in Washington DC to digitally encode, store and retrieve old archival materials, using a digitizing camera, a PC, off-the-shelf software, videodisks, a videodisk player, and a high resolution printer. This combination allows photographs, drawings, other paper documents to be saved with hand-written notes and curlicue typefaces preserved. Further details from *Hernan Otano or Rob Martella, Air and Space Museum, Smithsonian Institution, 1000 Jefferson Dr. SW, Washington DC 20007-3835.*

Aus Scope 4(1986)1: S. 4

CHINESISCH-DEUTSCHE TERMINOLOGIE-DATENBANK BEI DER GID

Ein Personal Computer und eine Speicherschreibmaschine, so könnte der Arbeitsplatz eines Fachübersetzers für chinesische Texte heute schon aussehen, wenn die Ergebnisse eines Projekts bei der Gesellschaft für Information und Dokumentation mbH (GID) in Frankfurt Eingang in die Praxis finden.

MIDASTA (Mikrocomputer gestütztes Datenverwaltungssystem für Terminologiearbeit) heißt das System, das es ermöglicht, deutsche oder englische Bezeichnungen in einen Mikrocomputer einzugeben, um dann über eine Speicherschreibmaschine die chinesischen Sprachäquivalente in chinesischer Symbolschrift ausgedruckt zu bekommen. Umgekehrt lassen sich auch die chinesischen Wörter eingeben, das Programm liefert dann die zugehörigen deutschen oder englischen Übersetzungen.

Die Benutzung des Systems ist durch die sogenannte Menütechnik, die dem Anwender ein hohes Maß an Erklärung und Unterstützung bietet, denkbar einfach.

Realisiert wurde MIDASTA auf dem Mikrocomputer RC 750 Partner mit dBase II. Als Drucker, insbesondere für die chinesischen Schriftzeichen, dient die Schreibmaschine 1011 von Olympia.

Angewendet wurde MIDASTA zunächst auf den Fachwortschatz Reprographie, den die GID bereits 1982 als deutsch-englisch-französisch-spanisches Wörterbuch in herkömmlichem Publikationsverfahren herausgegeben hat (Wörterbuch Reprographie. München: Saur 1982). Im elektronischen Publikationsverfahren ergänzt die GID dies nun durch ein Wörterbuch der Reprographie deutsch-englisch-chinesisch, das sich im wesentlichen auf dieselben etwa 500 Fachbegriffe stützt.

Yu Yong Yuan, Gesellschaft für Information und Dokumentation (GID), Sektion für Kommunikation, Tel.: (069) 6687-286

IV+V SYSTEM SOFTWARE PACKAGE

An Integrated Software for Flexible Information Management

At the Society for Information and Documentation, a publicly funded research and service organization in the Federal Republic of Germany, a software package has been developed (in cooperation with the Institute for Computerized Documentation, Graz, Austria) which provides support for automation of information and documentation services. An operational prototype of the system has been tested under real-world conditions; the feedback from these tests is currently being used to tune up the system.

The IV+V System (Informationsvermittlung und -verarbeitung (information dissemination and processing)) is conceived as a generalized data management system, intended for both bibliographic and factual information management. The IV+V System is designed as an implementation of the relational database model and is compliant with the command level specifications of SQL/DS. Basic system functions are: data entry/data editing, database creation and maintenance, retrieval, and output. The system also includes a screen processor which permits user-developed screens and user-controlled screen sequencing for specific applications, and a task scheduler, which provides operating system control of tasks. The internal file structure for tuples is directory based. Directory is established at the time the tuple is defined via data definition facility. Data elements are stored in a standard B-tree for rapid access. The creation and maintenance of libraries of all screen images is provided through the mask editor subsystem. All database management system and associated utility software is written in UCSD Pascal. The UCSD P-code interpreter and supporting modules have been rewritten (in assembler) and extended to provide communication support for auxiliary devices. The IV+V System is conceived as a hierarchy of tool boxes (hardware; basic software; application software; application) with the following highlights: portability for different hardware configurations; compatibility with various external information systems; adaptability for different types of application and usage.

I. Mistrik (IV+V Project Manager), Gesellschaft für Information und Dokumentation mbH, Tiergartenstr. 17, 6900 Heidelberg

BMFT UND KEG FÖRDERN INFORMATIONSMITTLUNG

Der europäische Informationsmarkt ist im Wachsen, aber: Mit der Zahl der unterschiedlichen Informationsanbieter (Hosts) steigt auch die Zahl der unterschiedlichen Systeme. Und es ist immer mehr der Spezialist notwendig, der über das Know How der Recherche in Datenbanken verfügt. Firmenauf- und -ausbau wird im Rahmen des am 30. Dezember 1985

durch den Bundesminister für Forschung und Technologie publizierten "Modellversuchs" Informationsvermittlung gefördert, indem über 3 Jahre degressiv Fördermittel für Personalkosten, Sachkosten und informationstechnische Ausrüstung gewährt werden. Die Entscheidung über die Auswahl der Förderung ist gefallen, die Bescheide an die Antragsteller sind versandt.

Parallel zu dieser nationalen Förderung erfolgt derzeit eine indirekte Förderung durch die Kommission der Europäischen Gemeinschaften (KEG): Die von der Kommission beauftragte "Gruppe zur Förderung des europäischen Fachinformationsmarktes" baut derzeit eine Datenbank BROKERSGUIDE auf, um internationale Zusammenarbeit und Kontakte zu fördern. Diese Datenbank ist kostenlos auf dem Host ECHO verfügbar, derzeit in der

Aufbau-Version nur unter dem Passwort BROKER. Sie enthält Angaben über etwa 500 Firmen und Organisationen, die im Auftrag für andere Datenbank-Recherchen durchführen. Im Rahmen einer Recherche in dieser Datenbank kann der Information Broker in der Bundesrepublik z.B. feststellen, ob ein Fachkollege in Italien existiert, der bei einem italienischen Host recherchieren kann, aber auch deutsch spricht. Eintragen lassen in diese Datenbank kann sich jeder - ob er nun durch das BMFT gefördert wird oder nicht.

Ausführliche Informationen sind zu erhalten bei: EURONET/DIANE ECHO, Peter Cornelius, 177 route d'Esch, L-1471 Luxembourg, Tel.: 00352-488041, Telex: 2181 EUROL LU, Mailbox: ILUX: P. Cornelius (über GEONET-INTERMAIL)

SKRIPTENFOLGE ZUR INFORMATIONSLINGUISTIK

Projekt "Informationsvermittlung" am Lehrstuhl für Informationswissenschaft in Konstanz entwickelte Lehrmaterialien

In der letzten Ausgabe des LDV-Forum, S. 91 wurde von den curricularen Aufgaben und Ergebnissen des BMFT-geförderten Projektes *Informationsvermittlung* (1982 - 1985) berichtet. Im folgenden werden, wie angekündigt, die Gliederungen der Skripte *Informationslinguistik I* und *II* sowie *Intelligente Informationssysteme* in geringfügig gekürzter Form präsentiert. Die Skripten können bei *U. Hahn, Universität Konstanz, Informationswissenschaft, Postfach 5560, 7700 Konstanz 1* angefordert werden.

U. Hahn:

INFORMATIONSLINGUISTIK I

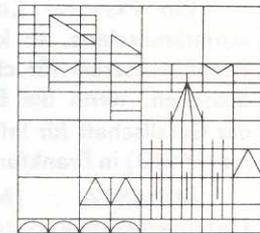
Einführung in das linguistische Information Retrieval

Bericht CURR-4/85, Januar 1985

Skript der gleichnamigen Lehrveranstaltung im WS 1982/83 am Lehrstuhl für Informationswissenschaft der Universität Konstanz.

Inhalt

0. Vorwort	4
1. Einführung in den Kurs Informationslinguistik I	6
2. Natuerlichsprachliche Effekte in Informationssystemen	11
2.1 Graphematik	14
2.2 Morphologie	15
2.3 Syntax	18
2.4 Semantik	26
3. Graphematisch-phonologische Verfahren der Informationslinguistik ..	32
3.1 Kompression fuer das Erkennen von Namensvarianten	35
3.2 Trigramm-Analyse fuer die Schreibfehlererkennung und -korrektur	44
4. Morphologische Verfahren der Informationslinguistik	55
4.1 'Longest Matching' zur Reduktion von Texttoken auf Grund- bzw Stammformen	58
4.2 'Iteratives Matching' zur Reduktion von Texttoken auf Grund- bzw. Stammformen	71
4.3 Woerterbuchabhaengiges Normalisieren zur Reduktion von Texttoken auf Standardformen	81



Universität Konstanz

Sozialwissenschaftliche Fakultät
 Fachgruppe Politikwissenschaft/
 Verwaltungswissenschaft
 Informationswissenschaft

Prof. Dr. Rainer Kuhlen

5. Syntaktische Verfahren der Informationslinguistik	109
5.1 Partielles Parsing mit dem Begrenzer-Verfahren	112
5.2 Transformation nominaler Paraphrasen mit dem Permutationsverfahren	122
5.3 Transformation komplexer Nominalkomposita in Simplex-Segmente mit dem Dekompositionsverfahren	127
5.4 Partielles Parsing durch syntaktisches Pattern Matching	141
6. Semantische Verfahren der Informationslinguistik	152
6.1 Listen-orientierte Extraktion inhaltlich signifikanter Tokens	152
6.2 Semantische Berechnungen zur Bestimmung inhaltlich signifikanter Tokens; ein an absoluten Häufigkeiten orientiertes Indexierungsverfahren	157
6.3 Semantische Berechnungen zur Bestimmung inhaltlich signifikanter Tokens; das auf gewichteten Relationen beruhende Indexierungsverfahren von PASSAT	163
7. Zur Integration informationslinguistischer Verfahren in konventionellen Informationssystemen	173

U. Hahn:

INFORMATIONSLINGUISTIK II

Linguistische und statistische Verfahren im experimentellen Information Retrieval

Bericht CURR-5/85, Januar 1985

Skript der gleichnamigen Lehrveranstaltung im WS 1983/84 am Lehrstuhl für Informationswissenschaft der Universität Konstanz.

Inhalt

0. Vorwort	5
1. Einführung in den Kurs Informationslinguistik II	8
1.1 Curriculare Verknüpfung von Informationslinguistik I und II und zentrale Lernziele des Kurses	8
1.2 Exkurs: explizite Definitionsansätze zur Informationslinguistik ...	9
1.3 Systemanforderungen im experimentellen Information Retrieval	10
1.4 Grundlagen des statistischen Information Retrieval	12
1.5 Grundlagen des linguistischen Information Retrieval	13
1.6 Curriculum Informationslinguistik: ein impliziter Definitionsversuch	16
2. Statistische Verfahren des experimentellen Information Retrieval	22
2.1 Term-orientierte Masse	22
2.1.1 Okkurrenzmasse	23
2.1.2 Assoziationsmasse	32
2.2 Dokument-orientierte Klassifikationsverfahren	39
2.2.1 Methoden des perfekten Clustering: Eigenwert- und Faktorenanalyse	40
2.2.2 Methoden des heuristischen Clustering	44
2.3 Bemerkungen zu statistischen Formen der Wissensrepräsentation und Wissenstransformation	50
3. Systeme des statistischen Information Retrieval	52
3.1 Textanalyse-Systeme des statistischen Information Retrieval	53
3.1.1 SMART	53
3.1.2 AIR / WAI	70
3.1.3 KEDMA (RESPONSA)	72
4. Linguistische Verfahren des experimentellen Information Retrieval	74
4.1 Motivation grammatikalischer Verfahren im Information Retrieval	74
4.2 Komponenten und Verfahren linguistischer Wissensrepräsentation (illustriert am Beispiel kontextfreier Phrasenstrukturgrammatiken und daran angepasster Abarbeitungsverfahren) ...	79
4.3 Typologie formaler und linguistischer Grammatikformate	91
4.4 Rekursive (RIN) und Erweiterte Uebergangnetzwerke (ATN) ...	94
4.5 Bemerkungen zu linguistischen Formen der Wissensrepräsentation und Wissenstransformation	109
5. Systeme des linguistischen Information Retrieval	112
5.1 Textanalyse-Systeme des linguistischen Information Retrieval	116

NACHRICHTEN

5.1.1	Linguistic String Project (LSP)	116
5.1.1.1	Sprachtheoretische Grundlagen von LSP	116
5.1.1.2	Systemstruktur von LSP	117
5.1.1.3	Grammatik- und Parser-Konzeption von LSP ...	119
5.1.1.4	Transformationskomponente von LSP	127
5.1.2	Juristische Dokumentenanalyse / Datenschutz (JUDO/DS) Computergestuetztes Texterschliessungssystem (CTX) ..	133
5.1.3	Saarbruecker Uebersetzungssystem / Bundessprachenamt (SUSY-BSA)	135
5.1.4	DETECT	137
5.1.5	Communication in Natuerlicher Sprache mit Dialog-Orientierten Retrievalsystemen (CONDOR)	139
5.1.6	Syntagmatic Organization Language (SYNTOL)	143
5.1.7	EPISTLE	146
5.2	Natuerlichsprachliche Zugangssysteme zu Informationssystemen	148
5.2.1	LADDER (LIFER-System)	148
5.2.2	Programmed Language-Based Enquiry System (PLANES) ...	157
5.2.3	RENDEZVOUS	159
5.2.4	Lunar Sciences Natural Language Information System (LSNLIS/LUNAR)	161
5.2.5	ROBOT / INTELLECT	163
5.2.6	Restricted English Question-Answering System (REQUEST) Transformational Question Answering System (TQA)	165
5.2.7	User Specilty Languages (USL)	167
6.	Abschaetzungen zum Einfluss statistischer und linguistischer Ver- fahren auf die Informationspraxis und das experimentelle Informa- tion Retrieval	170

U. Hahn:

INTELLIGENTE INFORMATIONSSYSTEME

Verfahren der Künstlichen Intelligenz im experimentellen Information Retrieval

Bericht CURR-6/85, Februar 1985

Skript der gleichnamigen Lehrveranstaltung im WS 1983/84 am Lehrstuhl für Informationswissenschaft der Universität Konstanz, ergänzt um Ausarbeitungen zu den Kursen *Programmierung von KI-Verfahren in PROLOG* (WS 84/85) und *Informationsmarkt* (SS 1984)

Inhalt

0.	Vorwort	9
1.	Einfuehrung in den Kurs Intelligente Informationssysteme	12
1.1	Zum funktionalen Vergleich computergestuetzter und intelligenter Informationssysteme	12
1.2	Wissensrepraesentation und -transformation in Expertensystemen	23
1.3	Zusammenfassung der Eigenschaften von DRS, FRS und EXS	30
1.4	Zentrale Lernziele und inhaltliche Charakterisierung des Kur- ses	34
1.5	Propaedeutische Modellierungsuebung an einfachen Wissenstypen	42
2.	Grundlegende Methoden der Wissensrepraesentation und Wissenstrans- formation in der Kuenstlichen Intelligenz	48
2.1	Logik: Praedikatenlogik erster Stufe	51
2.1.1	Syntax der Praedikatenlogik (1): Alphabet, atomare und komplexe Formeln	51
2.1.2	Semantik der Praedikatenlogik	55
2.1.3	Syntax der Praedikatenlogik (2): Quantoren, Inferenzregeln, Unifikation	57
2.1.4	Klausenlogik	61
2.1.5	Beweis von Theoremen: Resolutionsverfahren	67
2.1.6	"Who's Joe Smith's boss?" Widerspruchsbeweis und Antwortextraktion	68
2.2	Regel-basierte Systeme	77
2.2.1	Produktionssysteme	80
2.2.2	Transformationssysteme: regelbasierte Deduktionssysteme	84
2.2.3	AND/OR-Graphen, Vorwaerts-/Rueckwaerts-Verkettung	85
2.3	Semantische Netze	92
2.3.1	Deklarative Struktur semantischer Netze: Knoten, Kanten	94

2.3.2	Prozedurale Struktur semantischer Netze: direktes und inferentielles Matching	94
2.3.3	Strukturelle Erweiterungen einfacher semantischer Netze	101
2.4	Frames & Scripts	104
2.4.1	Deklarative Struktur von Frames: Slots, Defaults, Wertebereiche, attached procedures ..	104
2.4.2	Prozedurale Struktur von Frames: Frame-Inferenzen (Trigger, Demons)	105
3.	Vertiefung methodischer Aspekte der Wissensrepräsentation und -transformation	110
3.1	Logik als prinzipielle Wissensrepräsentationsbasis der Kuenstlichen Intelligenz	111
3.1.1	Vorteile einer Logifizierung nicht-logischer Repraesen- tationsformalismen	111
3.1.2	Logische Reformulierung semantischer Netzwerke in Klausel- Form-Logik	112
3.1.2.1	Einfaches semantisches Netzwerk	112
3.1.2.2	Zusaetzliche Ausdrucksmittel der Klausel-Form- Logik	114
3.1.2.3	Erweitertes semantisches Netzwerk: syntaktische Variante der Klausel-Form-Logik	115
3.1.2.4	Logische Inferenzregeln: Resolution	115
3.1.2.5	Inferenzen in erweiterten semantischen Netz- werken	118
3.1.3	Praedikatenlogische Reformulierung strukturierter Ob- jektDarstellungen: Frames, Scripts, semantische Netze	121
3.1.3.1	Exemplarischer Uebergang vom Praedikatenkal- kuel auf Units	121
3.1.3.2	Matching-Prozeduren auf Units	130
3.1.3.3	Deduktions-Operationen auf Units	133
3.2	Wissensrepräsentationssprachen	142
3.2.1	Frame-Repraesentationssprachen	144
3.2.1.1	Prototypische Frame-Sprache: KRL	145
3.3	PROLOG, Logik und logisches Programmieren	153
3.3.1	Prolog und Logik	153
3.3.1.1	Prolog und Horn-Klauseln	154
3.3.1.2	Prolog als Resolutionstheorembeweiser	158
3.3.1.3	Abarbeitungsstrategie in Prolog	160
3.3.2	Prolog und logisches Programmieren	162
3.3.2.1	Kontrollinformation / Steuerung des Kontroll- flusses	162
3.3.2.1.1	Backtracking	162
3.3.2.1.2	Cut-Praedikat	162
3.3.2.2	Illegale Logik	162
4.	Eine Typologie intelligenter Informationssysteme	169
4.1	Problemloesungssysteme	171
4.1.1	Expertensysteme	173
4.1.1.1	Die MYCIN-Familie	180
4.1.1.1.1	Aufgabenbereich von MYCIN	180
4.1.1.1.2	Systemstruktur von MYCIN	182
4.1.1.1.3	Wissensrepräsentation in MYCIN ..	184
4.1.1.1.4	Wissenstransformation in MYCIN: Goal-directed Backward Reasoning ..	201
4.1.1.1.5	Erklaerungsdialoeg mit MYCIN	211
4.1.1.1.6	MYCIN-Familie: MYCIN, TEIRESIAS, PUFF, CENTAUR, E-MYCIN	211
4.1.1.2	Die RI-Familie	218
4.1.2	Wissensbasierte Decision-Support-Systeme	220
4.1.2.1	Zur Konzeption wissensbasierter Decision-Sup- port-Systeme	220
4.1.2.2	Probleme bei der Modellierung formaler Organi- sationen am Beispiel intelligenter Entschei- dungsberatung in Buero-Umgebungen	221

4.1.2.3	Der 'Office Semantics'-Ansatz	223
4.1.2.3.1	Modellierung organisationaler Um- welt: Office Semantics	223
4.1.2.3.2	Die Wissensrepräsentationsprache OMEGA	224
4.1.2.4	Bueroarbeit als informelle kooperative Arbeit	225
4.1.2.4.1	Modellierung organisationaler Um- welt: eine kognitive Theorie informeller kooperativer Arbeit (Bueroarbeit)	225
4.1.2.4.2	ODYSSEY: ein wissensbasiertes DSS fuer Rei- seplanungsprozeduren	226
4.2	Natuerlichsprachliche Systeme (Informationslinguistik III)	236
4.2.1	Frage-Antwort-Systeme	243
4.2.1.1	HAM-RPM / HAM-ANS	256
4.2.1.1.1	Aufgabenbereich von HAM-RPM/HAM-ANS	256
4.2.1.1.2	Systemstruktur von HAM-RPM/HAM-ANS	256
4.2.1.1.3	Wissensrepräsentation in HAM-RPM / HAM-ANS	256
4.2.1.2	Berlin Automatic Construction of Semantic Net- works (BACON)	263
4.2.1.3	Genial Understander System (GUS)	265
4.2.1.4	CO-OP	266
4.2.1.5	Knowledge-Learning and -Using Systems (KLAUS)	268
4.2.1.6	Information Retrieval Using the RUS Parser (IRUS)	270
4.2.1.7	Transportable English Access Data Manager (TEAM)	272
4.2.1.8	Unix Consultant (UC)	274
4.2.2	Text-Transformationssysteme	276
4.2.2.1	Text-orientierte Verfahren fuer die Informa- tionsverwaltung und Kondensierung von Volltex- ten (TOPIC/TOPOGRAPHIC)	282
4.2.2.2	CYRUS + FRUMP = CyFr	323
4.3	Intelligentes Information Retrieval	325
4.3.1	Intelligentes Dokumenten-Retrieval	326
4.3.2	Intelligentes Daten-Retrieval (in wissensbasierten Datenbanksystemen)	328
4.3.3	Problemloesendes Informationssystem mit Deutsch als In- teraktionssprache (PLIDIS): Prototyp eines intelligenten Informationssystems	333
5.	Industrialisierungstendenzen in der Kuenstlichen Intelligenz: wissensindustrielle und foerderpolitische Entwicklungen auf dem Markt fuer intelligente Informationssysteme	336
5.1	Indikatoren fuer die Industrialisierung der Kuenstlichen In- telligenz	336
5.2	Organisationelle Aspekte der KI-Industrie	337
5.3	Finanzierungsaspekte der KI-Industrie: zur Rolle des Risiko-Kapitals	342
5.4	Fallstudien zur Realitaet der KI-Industrie	346
5.5	Wissenschaftspolitische Aspekte der KI-Industrie	349
6.	Zum methodischen und Anwendungsspektrum der Kuenstlichen Intelli- genz	368
6.1	Anwendungsgebiete der Kuenstlichen Intelligenz	369
6.2	Methoden/Techniken der Kuenstlichen Intelligenz	370
6.3	Intelligente Informationssysteme und ihr Bezug zur Kuenstlichen Intelligenz	372

Robert F. Simmons: Computations from the English. A Procedural Logic Approach for Representing and Understanding English Texts.

New Jersey: Prentice-Hall Inc., Englewood Cliffs, 1984. 308 Seiten.

Simmons' Buch von 1984 ist ein am Forschungsansatz des Autors orientiertes Lehrbuch und wendet sich allgemein an Forscher, Entwickler und Student(inn)en im Bereich der natürlichsprachlichen Systeme. Da eine Behandlung aller Aspekte des Buchs hier unmöglich ist, werde ich mich auf seinen Beitrag zur LDV-Forschung beschränken.

Nach einem historischen Überblick (Kap.1) entwickelt Simmons einen einheitlichen, logikorientierten Ansatz zur Analyse/Synthese und zur semantischen Repräsentation natürlicher Sprache (Kap. 2-5). In den Kapiteln 6 bis 11 werden die Eigenschaften dieses Ansatzes für sehr viele Arten solcher Softwaresysteme, die NL-Komponenten enthalten, diskutiert. Durch diese Konzeption ist es dem Autor möglich, wichtige theoretische Fragestellungen direkt für die technische Anwendung fruchtbar zu machen.

In Kapitel 1 wird die Forschungsgeschichte der natürlichsprachlichen Systeme sowie der augenblickliche Forschungsstand resümiert.

Kapitel 2 bringt eine philosophische Diskussion des Konzeptbegriffs im Rahmen des epistemologischen Monismus, dessen Zentralthese (es gibt nur Begriffe, keine Objekte) in der Tat die Grundlage für Simmons' intentionalen Repräsentationsansatz ist.

In Kapitel 3 werden der Hornklausel-Formalismus und die Resolutionstechnik eingeführt. Im Hintergrund steht dabei ein Hornklausel-Theorembeweiser HCPRVR, der dem Dec-10 PROLOG recht ähnlich ist, aber weniger syntaktischen Komfort (z.B. kein Strukturkonzept) bietet.

In Kapitel 4 wird eine annotierte cfg für ein Englisch-Fragment sowie die Repräsentationstechnik der Semantischen Relationen SR, einer Obermenge der Kasusrelationen, eingeführt. SR's einigen funktionale und syntaktische Informationen; sogenannte SSR's enthalten zusätzlich (ähnlich wie LFG-f-Strukturen) syntaktische Informationen. Es stellt sich allerdings die (im Buch nicht behandelte) Frage, inwieweit SR's selbst semantisch interpretationsbedürftig sind. Die resultierende Gesamtgrammatik hat eklektizistische Züge, wobei einige grammatische Lösungsvorschläge, z.B. die Behandlung von long-distance-Abhängigkeiten durch Pointer-Funktionen, die etwas ad-hoc wirken.

Kapitel 5 zeigt Möglichkeiten zur Generierung von S(S)R's, wobei u.a. spezielle semantische Dependenzregeln (semantic event forms) in die PSG integriert werden.

Analog zu Winograds SHRDLU wird in Kapitel 6 ein Steuerungsprogramm für einen Blöckewelt-Roboter skizziert.

Kapitel 7 zeigt die Anwendung einer HCPRVR-Grammatik in einem NL-frontend für ein Dokumentenverwaltungssystem. Bereits bei dieser kleinen Grammatik führt der Verzicht auf einen spezialisierten Grammatikinterpreter allerdings zu einer eher unübersichtlichen Grammatikdarstellung.

Die Kapitel 8 und 9 sind Fallstudien zur transphrastischen Analyse, zum Abstracting und zum inhaltsbezogenen Retrieval für einen narrativen Text. Einer attribuierten cfg werden satzübergreifende Regeln, die jeweils einem Teilbaum der story-grammar entsprechen, (Schemaregeln) hinzugefügt. Durch SSR's instantiierte story-Grammatiken (Schemata) können in einer Datenbank abgelegt, mit einer Anfragerrepräsentation abgeglichen oder zur story-grammar eines Abstracts verkürzt werden. Interessant ist die durch bottom-Anwendung erreichte Flexibilität von Schemaregeln. Trotzdem würde ein realistisches Textanalysestystem unrealistischerweise "hundreds or thousands of schemas" (Seite 171) benötigen!

Kapitel 10 bringt mehrere cfg-Interpreter in HCPRVR (top-down, bottom-up, chart) sowie Mechanismen für Paraphrasenerzeugung und Übersetzung (Paraphrasen- bzw. Transferregeln operieren auf SSR's).

Das in Kapitel 11 vorgestellte Beispiels-Expertensystem ist unabhängig vom S(S)R-Ansatz. Die folgenden Ausführungen zeigen aber, daß generalisierte SR's auch als textunabhängiger Wissensrepräsentationsformalismus aufgefaßt werden können.

Simmons Buch ist mit verschiedenen Anhängen (Dokumentation und Listing zum System HCPRVR, Satzgrammatik und Schemata des Beispieltexes) und mit den üblichen Indizes ausgestattet. Erfreulicherweise gibt es zu jedem Kapitel gute Experimentieraufgaben. Als gravierender Schönheitsfehler bleibt der mit ca. US \$ 90.- hohe Preis, der die Verbreitung des Buchs hafthaft beschränken dürfte.

Sebastian Goeser, Universität Ulm

Jiri Panyr: Automatische Klassifikation und Information Retrieval: Anwendung und Entwicklung komplexer Verfahren in Information-Retrieval-Systemen und ihre Evaluierung.

Sprache und Information, Band 12. Tübingen: Niemeyer, 1986. 416 Seiten.

Panyrs Arbeit beschreibt das von ihm entwickelte Klassifikationsverfahren STEINADLER, das zum größten Teil im Rahmen des Forschungsprojektes CONDOR der Firma SIEMENS entstand; zusätzlich gibt das Buch einen Überblick über die Anwendung von Klassifikationsverfahren im Information Retrieval. Die Arbeit wurde 1984 an der Philosophischen Fakultät der Universität des Saarlandes (Fachrichtung Informationswissenschaft) als Dissertation angenommen.

Das umfangreiche Werk ist recht verständlich und gut lesbar geschrieben, ärgerlich ist nur die Unsitte, die Anmerkungen am Schluß des Buches zu sammeln, statt sie als Fußnoten direkt beim Text aufzuführen.

Von den rund 340 Seiten Text sind 190 Seiten fast ausschließlich der Beschreibung von STEINADLER gewidmet. Wie Panyr ausführt, bestehen die wesentlichen Vorzüge dieses Klassifikationsverfahrens zum einen in dem relativ geringen Aufwand bei seiner Anwendung, zum anderen in der Möglichkeit, die Klassifikation aufgrund von neu hinzugekommenen oder gelöschten Dokumenten zu korrigieren, anstatt den gesamten Dokumentenbestand neu klassifizieren zu müssen.

Die vorliegende Beschreibung des Verfahrens genügt allerdings kaum den Ansprüchen, die man an eine wissenschaftliche Arbeit stellt. Die breite Darstellung des doch recht komplexen Systems enthält so gut wie keine Begründungen für irgendwelche gewählten Vorgehensweisen; stattdessen finden sich nur Bemerkungen der Art "Oft genügt, wie sich auch gezeigt hat ..." (Anm. 7 zu Kapitel 5), "Anschließend können (es ist aber nicht zwingend) ..." (S. 156) oder "Bei der CONDOR-Anwendung wurde meist gewählt: ..." (S. 219). Bei der ganzen Beschreibung bleibt die empirische Fundierung der Arbeit im Dunkeln. Auf den Seiten 146-149 finden sich sogar (schematische?) Darstellungen von Term-Häufigkeitsverteilungen, die wohl außer Panyr noch niemand in der quantitativen Linguistik beobachten konnte. Würde man nicht auch aus anderen Quellen von experimentellen Anwendungen der Klassifikationskomponente von CONDOR (in der Arbeit sind neben einem kurzen Hinweis auf solche Anwendungen nur einige Beispiele von Klassifikationen mit STEINADLER wiedergegeben), man müßte vermuten, daß STEINADLER nicht über die konzeptionelle Phase hinausgelangt ist.

Die Ausführungen zur Einordnung und zur Anwendung von Klassifikationsverfahren im Information Retrieval beeindrucken in erster Linie durch

ihre Breite (das Literaturverzeichnis weist die stolze Zahl von 358 Einträgen auf), doch vermißt man hier eine kritische Diskussion der Vielzahl von zusammengetragenen Zitaten. So werden z.B. die Begriffe Relevanzgrad, -wahrscheinlichkeit und Systemrelevanz eingeführt, ohne jedoch darauf einzugehen wie diese bei der Anwendung von Retrievalverfahren, die mit Klassifikationen arbeiten, zu interpretieren sind. Auf Seite 285 wird fälschlicherweise behauptet, alle mit Relevanzfeedback arbeitenden Retrievalverfahren seien den probabilistischen Retrievalstrategien zuzurechnen—hier fragt sich, ob Panyr die Bedeutung der von ihm erläuterten Begriffe wirklich voll erfaßt hat. Weit schwerer als solche Oberflächlichkeiten wiegt aber die Tatsache, daß zwischen den allgemeineren Betrachtungen und der Darstellung von STEINADLER so gut wie keine Verbindungen gezogen werden. So beschreibt Panyr zwar ausführlich die Evaluierungsproblematik bei Retrievalverfahren, die Evaluierung von STEINADLER handelt er jedoch in wenigen Zeilen mit der Angabe von zwei Zahlenwerten ab, die aber, da sie für sich alleine stehen, völlig nichtssagend sind. Gleiches gilt für die unterschiedlichen Bewertungskriterien von Klassifikationsverfahren (z.B. Effektivität, Effizienz, Stabilität), die er zwar allgemein diskutiert, bei der Beschreibung von STEINADLER gibt er aber hierzu nur eine Aufwandsabschätzung des Verfahrens.

Besonders bei den von Panyr gepriesenen Vorzug von STEINADLER, die Möglichkeit zur Anpassung des Klassifikationsnetzes an Änderungen des Dokumentenbestandes, sollte man zumindest eine Diskussion des Problems der daraus resultierenden Stabilität des Verfahrens erwarten, besser wäre natürlich eine entsprechende experimentelle Untersuchung. Allein—Panyr geht mit keinem Wort auf diese Problematik ein.

Welchen Gewinn hat nun ein Leser von der Lektüre des vorliegenden Werkes? Er erhält einen breiten, aber relativ unkritischen und wenig tiefschürfenden Überblick über den Einsatz von Klassifikationsverfahren in Information Retrieval, und zusätzlich die ausführliche Beschreibung eines neuen, komplexen Klassifikationsverfahrens, dessen Wert er jedoch aufgrund des Gebotenen überhaupt nicht einschätzen kann.

Warum ein derart umfangreiches Werk mit so wenig Gehalt in der doch sonst recht anspruchsvollen Reihe "Sprache und Information" erschienen ist, bleibt wohl ein Geheimnis der Herausgeber.

Norbert Fuhr, Technische Hochschule Darmstadt

Christopher Habel (Hrsg.): Künstliche Intelligenz. Repräsentation von Wissen und natürlichsprachliche Systeme.

Frühjahrsschule Dassel(Solling), März 1984. Informatik Fachberichte, Subreihe Künstliche Intelligenz, Band 93. Berlin: Springer 1985. 320 Seiten

Wolfgang Hoepfner, Universität Hamburg, Fachbereich Informatik, Projektgruppe WISBER

Der Band **Künstliche Intelligenz. Repräsentation von Wissen und natürlichsprachliche Systeme** entstand aus einem Teil der Kurse, die auf der Frühjahrsschule 'Künstliche Intelligenz' (KIFS) in Dassel (Solting) im März 1984 angeboten wurden. Diese Veranstaltung wird seit 1982 jährlich von der Gesellschaft für Informatik (insbesondere von deren Fachausschuß **Künstliche Intelligenz und Mustererkennung**) durchgeführt und dient in erster Linie dazu, dem Nachholbedarf in der Ausbildung auf dem Gebiet KI im deutschsprachigen Raum zu begegnen. Die Teilnehmer dieser Frühjahrsschule sind in etwa zu gleichen Teilen Studenten und Mitarbeiter industrieller Institutionen oder Forschungseinrichtungen.

Der Band, der aus der ersten Frühjahrsschule hervorging (BIBEL, SIEKMANN 1982), enthielt Einführungen in wichtige Teildisziplinen der Künstlichen Intelligenz: **Einführung in die Künstliche Intelligenz, Expertensysteme, Deduktionsverfahren, Computer-Unterstützter Algorithmenentwurf, Natürlichsprachliche Systeme, Bildverstehen und Robotertechnologie**. Der vorliegende Band enthält sechs Aufsätze, in denen nicht versucht wird, das gesamte Spektrum der KI abzudecken, sondern es werden speziellere Aspekte dargestellt. Auch im Programm der Frühjahrsschule sind diese Themen als Aufbaukurse angeboten worden. Grundkenntnisse im Gebiet Künstliche Intelligenz sind deshalb hier vorausgesetzt. Alle während der KIFS-84 angebotenen Grund-, Aufbau- und Spezialkurse sind im Vorwort des Herausgebers aufgeführt.

Versucht man, die sechs Aufsätze den Arbeitsschwerpunkten der KI zuzuordnen, dann ist das grundlegende Gebiet **Wissensrepräsentation** mit den ersten beiden Aufsätzen vertreten, die nächsten drei können dem Gebiet **Natürlichsprachliche Systeme** zugeordnet werden und der letzte Beitrag beschäftigt sich mit **Maschinellen Lernen**. Im Gegensatz zum ersten Band der KI-Frühjahrsschulen nimmt der hier vorliegende eine *"Zwischenstellung zwischen Lehrbuch*

und Aufsatzsammlung" ein (S.V), wobei der Charakter einer Aufsatzsammlung hier deutlich dominiert. Nun aber zu den Bestandteilen dieser Sammlung im einzelnen.

Wolfgang Bibel: Inferenzmethoden (S. 1-47)

In diesem Aufsatz wird versucht, *"einen klärenden Überblick über das Phänomen des Schließens in seinen verschiedenen Manifestationen unter möglichst einheitlichem Gesichtspunkt zu geben"* (S. 1). Als einheitlicher Gesichtspunkt wird die mathematische Logik eingeführt, deren Grundlagen im ersten Kapitel knapp vorgestellt werden (Prädikatenlogik erster Stufe, Resolutionsprinzip, Konnektionsmethode, Modallogik (auf nur etwas mehr als einer Seite!) und mathematische Beweisführung). Das zweite Kapitel hat *nicht-monotone Inferenz* zum Thema, ein für die KI zentrales Phänomen, bei dem es kurz gesagt darum geht, Kalküle zu entwickeln, die das Zurücknehmen von ursprünglich gültigen Aussagen gestatten. Anhand des klassischen Beispiels vom Konzept *Vogel*, dem man die Eigenschaft *kann-fliegen* zubilligt, gleichzeitig aber mit den bekannten Ausnahmen *Strauß* und *Pinguin* leben muß, stellt Bibel verschiedene nicht-monotone Repräsentationsformen und Inferenzmethoden vor. Datenbank-formalismen, PROLOG-Lösungen, und - als Erweiterung der Prädikatenlogik - Zirkumskription, Modallogik und fuzzy logic werden als Beispiele kurz eingeführt. Im dritten Kapitel werden schließlich drei Spezialformen von Inferenzen skizziert: Ableitungen über dem Wissen verschiedener Akteure, Metainferenzen und induktives Schließen.

Insgesamt kann der Aufsatz von Bibel als - allerdings äußerst knapper - Überblick über das Gebiet mathematisch-logischer Inferenzkalküle gelten. Die zahlreichen Verweise auf weitere Literatur sind für ernsthaft Interessierte sicher der Hauptgewinn. Neben der teilweise sehr idiosynkratischen Terminologie (*Intellektik* statt *KI*, *Ermangelungsinferenz* für *default reasoning*) erschweren fehlende Sonderzeichen (z.B. Indizes) und eine inkorrekte Abbildung (S. 33) die Lektüre etwas.

Joachim Laubsch: Techniken der Wissensrepräsentation (S. 48-93)

Eine Unausgewogenheit des ersten KIFS-Bandes (BIBEL, SIEKMANN 1982) bestand darin, daß kein Überblick zu dem für alle Bereiche der KI-Forschung zentralen Gebiet der *Wissensrepräsentation* gegeben wurde. Diese Lücke wird durch den Aufsatz von Laubsch im vorliegenden Band geschlossen.

Nach einer kurzen Einführung (*Was ist Wissen und wie wird es angewandt?*) und einer noch kürzeren Vorstellung von vier Expertensystemen (EL, MOLGEN, MYCIN und R1), die verschiedene Repräsentationsformalismen (Constraints, Produktionsregeln) einsetzen, werden im dritten Kapitel verschiedene prozedurale Lösungen für ein exemplarisches Problem vorgestellt (Auffinden einer Verschmutzungsquelle in einem Wasserleitungsnetz). Für dieses Beispiel wird eine Lösung durch ein LISP-Programm mit einer Lösung durch das regelbasierte System AMORD verglichen, wobei der Expliziteitsanspruch des zweiten, wissensbasierten Ansatzes sehr anschaulich vor Augen geführt wird. Das vierte und umfangreichste Kapitel stellt zunächst die wichtigsten Repräsentationsklassen der KI vor (Prädikatenlogik, Semantische Netze, Frames). Als neuerer Ansatz, der eine Vereinigung von logischen Formalismen, die als Kalkül bereits Verwendungsvorschriften enthalten, mit deklarativen Formalismen anstrebt, wird die Sprache KRYPTON vorgestellt und ein Hauptproblem, nämlich die Subsumptionsrelation, an mehreren Beispielen illustriert. Inferenzen und Meta-inferenzen zur Steuerung von Ableitungen werden an mehreren Beispielen eingeführt (Blocks-Welt, Kontrolle in Produktionssystemen, AMORD, EMYCIN). Der Aufsatz schließt mit einer wiederum an Beispielen orientierten Diskussion von Backtracking-Verfahren.

Insgesamt gesehen gibt der Aufsatz von Laubsch einen kompakten Überblick über das Thema Wissensrepräsentation. Positiv hervorzuheben ist die Vielzahl der Beispiele, deren detaillierte Präsentation oft bis auf die LISP-Ebene hinuntergeht. Dies setzt beim Leser eine gewisse Vertrautheit mit LISP und KI-Programmiertechniken voraus, eine Voraussetzung, die für den Leserkreis jedoch nicht ungerechtfertigt erscheint. Ergänzend zur rein deskriptiven Darstellung wären an vielen Stellen Bewertungen der vorgestellten Formalismen (*Was kann man mit ihnen nicht repräsentieren, und warum ist dem so?*) hilfreich gewesen. Ein zusammenfassendes Kapitel, in dem auf aktuelle Forschungstrends und Zukunftsperspektiven hingewiesen wird, fehlt ebenfalls.

Camilla B. Schwind: Semantikkonzepte in der Künstlichen Intelligenz (S. 94-158)

Ziel dieses Aufsatzes ist es, "die wichtigsten in der Künstlichen Intelligenz (KI) formulierten und benutzten Konzepte zur Darstellung der Semantik natürlicher Sprache" vorzustellen und zu diskutieren (S. 95). Nach allgemeinen

Äußerungen über die Problematik der Definition von *Semantik* - dies läuft letztlich auf eine nicht näher erläuterte *konzeptuelle Struktur* hinaus - gliedert sich der Aufsatz in drei Kapitel: *Rolle und Aufgabe einer semantischen Komponente* (8 S.), *Grundformalismen* (3 S.) und *Semantikkonzepte* (45 S.).

An einer allgemeinen Systemarchitekturgraphik (Fig. 1, S. 97), die Analyse- und Generierungsprozesse in einem natürlichsprachlichen System aufzeigen soll, aber schon allein für die Analyseprozesse nicht hinreichend ist, wird versucht, Semantik als Übersetzungsprozeß von syntaktischen Repräsentationen in eine *Semantiksprache* zu definieren, wobei semantische Relationen, Tiefenkasus und Inferenzregeln pauschal als Elemente des konzeptuellen Wissens aufgezählt werden, jedoch nur dem ersten dieser Elemente semantisches Wissen zugebilligt wird. Wozu man nun eine *Semantikkomponente* (die in der Systemarchitektur der Fig. 1 als solche nicht vorkommt) in einem natürlichsprachlichen System braucht, soll an einzelnen Phänomenen gezeigt werden. Es sind dies Auflösung syntaktischer Mehrdeutigkeiten, Pronomenanalyse, definite Kennzeichnungen und Ellipsenanalyse. Als Beispielssammlung mögen diese Phänomene dem Leser intuitiv klar werden, wenn er sie nicht bereits vorher kannte; was jedoch spezifisch semantische Probleme sind und wie man sie lösen könnte, wird hierdurch nicht erhellt.

Kapitel 3 soll einige Grundformalismen einführen. Was dann tatsächlich geschieht, entpuppt sich als Einführung der Definite-Clause-Grammar und ihrer Implementierung in PROLOG. Dieser Formalismus wird für die Syntaxanalyse eingeführt, wofür er ja auch gedacht ist. Semantik als Thema des Aufsatzes tritt nur in folgender Bemerkung in Erscheinung: "Wenn man während der Syntaxanalyse semantische Informationen manipulieren will, führt man weitere Stellen für Terme in die syntaktischen Prädikate ein" (S. 108).

Das umfangreichste Kapitel, das vierte, beschäftigt sich mit *Semantischen Netzwerken, Frames und Logik*. Für den ersten Repräsentationsmechanismus wird als Beispiel unglücklicherweise eine Darstellung lexikalischer Merkmale angeführt (Wortarten, Tempusformen etc., Fig. 2, S. 110) und auch die weiteren Beispiele für Satzsemantik mittels Tiefenkasusrelationen kehren wichtige Probleme unter den Teppich (z.B. Quantifizierung). Es folgen formale Definitionen

semantischer Netze, einige Anwendungsbeispiele (Erzeugung von Netzdarstellungen aus Sätzen, automatische Indexation (!), CD-Graphen) und schließlich eine Bewertung und Diskussion; das Geradebiegen der dort vorgebrachten Argumente würde allerdings mindestens doppelt so viel Raum beanspruchen und unterbleibt aus diesem Grunde hier.

Nach den semantischen Netzen werden *Frames* eingeführt, wobei die Autorin offenbar dem Irrtum unterlegen ist, daß Kasusrahmen und *scripts* ("Handlungsrahmen") prototypische Anwendungsfälle für Frame-Repräsentationen sind. Zur knappen Bewertung und Diskussion des Frame-Ansatzes kann nur das wiederholt werden, was zu dieser Rubrik bezüglich semantischer Netze oben gesagt wurde.

Als dritter Formalismus für Semantikdarstellungen tritt nun *Logik* auf den Plan, die durch Zitierung des ersten Satzes aus dem Johannesevangelium in ihrer historischen Bedeutung gewürdigt wird. Klassische Prädikatenlogik und ihre Beziehung zu natürlichsprachlichen Äußerungen wird formal ausführlich dargestellt, wobei allerdings viele Einzelheiten fragwürdig bleiben. Gleiches gilt für den Exkurs über Prädikatenlogik als Datenbankschnittstelle und die weiteren Abschnitte (dreiwertige Logik, Modallogik). Andere, speziell für natürlichsprachliche Systeme relevante Logikvarianten, wie Sortenlogik, Fuzzy Logic (ohne Literaturverweis!, ZADEH (1981) wäre hier beispielsweise angebracht gewesen) und nichtmonotone Logiken werden nur namentlich genannt. Die Wertung logikorientierter Ansätze beginnt mit dem Axiom (1): "*Logik ist bereits eine Theorie der Bedeutung.*" (S. 151) und endet mit dem Axiom (8): "*Das Problem der Primitive stellt sich auch in logikorientierten Ansätzen. Wenn frau einen Logikformalismus benützt, muß sie nicht-logische Symbole und ihre Bedeutung formulieren, d.h. eine Menge solcher Symbole auswählen.*" (S. 152). Glücklicherweise findet man im Beitrag von Laubsch eine adäquatere Darstellung von semantischen Netzen, Frames und Logikformalisten.

Insgesamt hätte man sich gewünscht, daß der Herausgeber kraft seiner Fachkompetenz auf eine verbesserte Version des Aufsatzes zu dem zweifellos interessanten Themenbereich gedrungen hätte, beispielsweise durch Hinweise auf neuere KI-Arbeiten (gut zwei Drittel der zitierten Literatur ist vor 1980 erschienen), oder daß er den Aufsatz in dieser Form einfach abgelehnt hätte.

Thomas Christaller: Parser als integraler Bestandteil von Sprachverarbeitungssystemen. Eine Materialiensammlung. (S. 159-183)

In diesem Beitrag geht es darum, aus der Sicht einer Komponente eines natürlichsprachlichen Systems generelle Prinzipien für die Architektur solcher Systeme zu gewinnen. Die stärkere Integration verschiedener, traditionell sequentiell arbeitender Komponenten ist das Hauptanliegen des Autors, das durch "*eine ganzheitliche Sicht von Sprachverarbeitungssystemen*" motiviert ist (S. 159).

Nach der Einleitung werden im zweiten Kapitel generelle Probleme der Sprachanalyse sowie die derzeitige Forschungssituation skizziert. Das dritte Kapitel beschäftigt sich mit Analysekomponenten in Sprachverstehenssystemen und den ihnen zu Grunde liegenden Architekturprinzipien. Kapitel 4 enthält eine Taxonomie von integrierten Analyseverfahren (heterarchisches Modell, Integrationsmodell, Kaskadenmodell, Blackboard-Modell) und weist auf kognitive Indizien für ein integriertes Vorgehen hin. Für drei Vertreter dieser Klassifikation werden im fünften Kapitel implementierte Systeme vorgestellt: Winograds SHRDLU als Kaskadenvertreter, HEARSAY-II als Blackboard-System, der Word-Expert-Parser als Integrationsmodell und das DFG-Projekt 'Prozedurale Dialogmodelle' als weiteres Kaskadenmodell, dem der Autor generell den Vorzug gibt (S. 180). Das Projekt 'Prozedurale Dialogmodelle' bildet im sechsten Kapitel den Anlaß für eine Diskussion empirisch orientierter Ansätze in der KI-Forschung. Das abschließende siebte Kapitel enthält einige an den Erfahrungen eines KIFS-Kurses orientierte didaktische Hinweise auf den Einsatz von Rollenspielen als Simulationsverfahren. Als Wissensrepräsentationsformalismus werden parti-tionierte Netzwerke sehr knapp eingeführt, gleiches geschieht mit der Lexical Functional Grammar als Grammatikformalismus.

Dieser Aufsatz entspricht wohl am ehesten dem tatsächlichen Ablauf eines KIFS-Kurses, und dürfte deshalb am fruchtbarsten für die Teilnehmer selbst sein. Andererseits ist er eine brauchbare Diskussionsgrundlage für Modularisierungsaspekte in sprachverarbeitenden Systemen. Textverarbeitende Systeme sind für den Autor sicher ein hilfreiches Instrument bei der Erstellung von Druckvorlagen. Damit der Einsatz eines solchen Systems für die Leser nicht zum Nachteil wird, wäre bei diesem Aufsatz eine augenfreundlichere, vor allem größere Schrifttype notwendig gewesen.

Uta M. Quasthoff: Textverstehen und Textproduktion (S. 184-248)

Dieser Aufsatz ist der Teil des gesamten Bandes, der linguistischer Vorgehensweise am ähnlichsten ist. Er hätte so auch in jedem Linguistik-Sammelband abgedruckt sein können, und die linguistische Grundorientierung wird von der Autorin auch explizit hervorgehoben (S. 184).

Im ersten Kapitel werden die Grundlagen aus textlinguistischer Sicht unter dem Gesichtspunkt einer umfassenden Kognitionswissenschaft vorgestellt. Durch eine, auch graphisch repräsentierte *disziplinäre Flußlandschaft* (S. 188) wird der Zusammenhang der an Texttheorien beteiligten Disziplinen verdeutlicht. Spezifisch linguistische Probleme, wie Koreferenz, Ellipsen, Thema/Rhema-Gliederung, Topikalisierung, Metasprache, Mehrdeutigkeiten u.a.m., bilden die Basis für einen Vergleich zwischen dem wissenschaftlichen Vorgehen der Disziplinen Linguistik, Psychologie, Soziologie und KI. Im zweiten Kapitel werden zwei aus der Linguistik stammende Textstrukturmodelle vorgestellt und verglichen (*Geschichtengrammatik* und *Relationsstruktur*), die beide einen Anspruch auf kognitive Adäquatheit erheben. Auf neuere KI-Ansätze in diesem Bereich (z.B. McKEOWN 1982, LEHNERT 1982, MANN, THOMPSON 1983) wird leider nicht eingegangen. Das dritte Kapitel befaßt sich mit der Generierung von Texten und ist auf insgesamt vier Seiten relativ kurz geraten. Die genauere Vorstellung des Systems TALE-SPIN (MEEHAN 1980, die Angabe fehlt im Literaturverzeichnis und sei deshalb hiermit nachgereicht) hätte die Probleme, mit denen KI-Systeme zu kämpfen haben, sehr gut demonstrieren können, insbesondere was den pragmatischen Aspekt der Planung angeht.

Zusammenfassend ist dieser linguistisch orientierte Beitrag zu Problemen der KI zu begrüßen, wenn auch eine etwas stärkere Integration mit KI-Lösungsansätzen wünschenswert gewesen wäre.

Christopher Habel, Claus-Rainer Rollinger: Lernen und Wissensakquisition (S. 249-320)

Der letzte Beitrag dieses Bandes befaßt sich mit einem Themengebiet, das in der KI-Forschung gerade in den letzten Jahren an Aktualität zugenommen hat: wie kann man Methoden für Lernen und Wissenserwerb auf Rechenanlagen modellieren?

Das einleitende Kapitel führt in die Probleme des maschinellen Lernens ein. Das zweite

Kapitel behandelt verschiedene Lernmethoden, wobei auf das *Lernen durch Beispiele* und auf *Lernen durch Wahrnehmung (Entdeckung)* im folgenden dritten Kapitel näher eingegangen werden soll. Es finden sich dann allerdings nur Vorstellungen von Systemen für den ersten Lernansatz. Diese Systeme werden mit Ausnahme der Generalisierungsstrategie (Kap. 3.2) und der Systeme LEX und LEX2 (Kap. 3.7) sehr anschaulich mit zahlreichen Abbildungen und Beispielen vorgestellt. Generell kann man sagen, daß Maschinelles Lernen erst am Anfang steht, und dies wird im knappen 4. Kapitel dieses Aufsatzes ('Abschließende Bemerkungen', 13 Zeilen) auch zum Ausdruck gebracht.

Die bereits zu Anfang geäußerte Einschätzung des vorliegenden (und vergleichsweise preiswerten) Bandes als einer Aufsatzsammlung sollte sich nach den Einzelbesprechungen bestätigen haben. Mit Ausnahme des Artikels von C. Schwind sind alle Beiträge durchaus lesenswert und als Ergänzungen des vorhergehenden KIFS-Bandes (BIBEL, SIEKMANN 1982) oder als Einführungen in Spezialthemen geeignet. Insbesondere in den letzten drei Aufsätzen des Bandes (Christaller, Quasthoff, Habel/Rollinger) wird auch auf die Grenzen der vorgestellten Lösungen hingewiesen, was eigentlich für jede wissenschaftliche Publikation gelten sollte, gerade im Bereich der KI aber oftmals - veranlaßt durch euphorische oder übergeneralisierende Aussagen - aus dem Blickfeld gerät.

Beziehungen zwischen den einzelnen Aufsätzen sind leider nicht explizit, sondern müssen vom Leser selbst hergestellt werden. Daß dies nicht geschehen ist, ist sehr zu bedauern, zumal es für den vorliegenden Band wegen der zahlreichen inhaltlichen Berührungspunkte zwischen den Aufsätzen für den Herausgeber nicht allzu schwierig gewesen wäre.

Literatur:

BIBEL, SIEKMANN 1982

Bibel, W., Siekmann, J. (eds.): Künstliche Intelligenz. Frühjahrsschule Teisendorf, März 1982, Berlin: Springer 1982. (= Informatik Fachberichte 59)

LEHNERT 1982

Lehnert, W. G.: Plot Units: A Narrative Summarization Strategy. In: W.G. Lehnert, M.H. Ringle (eds.): Strategies for Natural Language Processing. Hillsdale: Erlbaum 1982, S. 375-414

Gesellschaft für Linguistische Datenverarbeitung (GLDV)

Die Gesellschaft für Linguistische Datenverarbeitung wurde 1975 unter dem Namen 'LDV-Fittings. Verein zur Förderung der wissenschaftlichen linguistischen Datenverarbeitung' gegründet: Sie hat heute über 220 Mitglieder, die sich in Forschung, Lehre, Anwendung und im Studium mit Linguistischer Datenverarbeitung im weitesten Sinne befassen. GLDV-Mitglieder arbeiten ebenso in der philologischen Datenverarbeitung, wie in der automatischen Übersetzung, in der sprachbezogenen Informationswissenschaft und -praxis, in der Grundlagenforschung zur Mensch-Maschine-Kommunikation und in der sprachorientierten Künstlichen Intelligenz.

Die GLDV versteht sich als Kooperationsorgan ihrer Mitglieder. Sie veranstaltet Tagungen und gibt das zweimal jährlich erscheinende LDV-Forum heraus. Der Schwerpunkt ihrer fachlichen Arbeit liegt bei den von den Mitgliedern initiierten Arbeitskreisen.

Die GLDV fördert die Zusammenarbeit mit den Nachbardisziplinen (Informatik, Künstliche Intelligenz, Informationswissenschaft, Sprachwissenschaft, usw.). Sie arbeitet mit den einschlägigen Fachgesellschaften zusammen. International kooperiert die GLDV mit Organisationen wie der Association for Literary and Linguistic Computing (ALLC) und der Association for Computational Linguistics/European Chapter (ACL/EC).

Studienführer LDV

Lutz-Hensel, M.: Studienführer Linguistische Datenverarbeitung (LDV) für die wissenschaftlichen Hochschulen der Bundesrepublik Deutschland, 1985.

Zu beziehen über

Prof. Dr. Jürgen Krause
Universität Regensburg
Linguistische Informationswissenschaft
Postfach 397
8400 Regensburg

Publikationen

Krallmann, D. (Hrsg.): Dialogsysteme und Textverarbeitung (Vorträge der Jahrestagung 1979). Essen 1980.

Krallmann, D.; Krause, J. (Hrsg.): Linguistische Datenverarbeitung und Informationswissenschaft in der Bundesrepublik Deutschland. Studiengänge, Berufsaussichten, Förderung (Vorträge der Jahrestagung 1980). Essen/Regensburg 1981.

Bátori, I.; Krause, J.; Lutz, H.D. (Hrsg.): Linguistische Datenverarbeitung. Versuch einer Standortbestimmung im Umfeld von Informationslinguistik und Künstlicher Intelligenz (Vorträge Jahrestagung 1982). Tübingen 1982.

Informationszentrum Sozialwissenschaften; Institut für deutsche Sprache (Hrsg.): Linguistische Datenverarbeitungssoftware. Standard- und Entwicklungsprogramme. Bonn, Mannheim 1982.

Krause, J.; Niederehe, H.-J.: Mikros und Textverarbeitung (Vorträge der Jahrestagung 1983, Sektion I). Hamburg 1984.

Rostek, L.; Schulz, G. (Hrsg.): Computerunterstützte Inhaltsanalyse (Vorträge der Jahrestagung 1983, Sektion II). Frankfurt 1985.

Hellwig, P.; Lehmann, H. (Hrsg.): Trends in der Linguistischen Datenverarbeitung (Vorträge der Jahrestagung 1984). Hildesheim, 1986.

Endres-Niggemeyer, B.; Krause, J. (Hrsg.): Sprachverarbeitung in Information und Dokumentation (Vorträge der Jahrestagung 1985). Berlin, Heidelberg, New York, 1985.

Tagungen und Workshops

Dez. 1979 Kolloquium zur linguistischen und literarischen Datenverarbeitung (in Kooperation mit ALLC). Bonn.

Dez. 1980 Kolloquium «Linguistische Datenverarbeitung - Ausbildung und Berufsperspektiven». Saarbrücken.

März 1981 Workshop «Datenbanken». Sektion: Datenbanken und natürliche Sprache (in Kooperation mit der GI). Heidelberg.

April 1981 Expertengespräch «Automatisches Indexieren» (in Kooperation mit der GI). Frankfurt.

Nov. 1981 Workshop «Philologische Anwendungen der LDV». Tübingen.

März 1982 Tagung «LDV und Nachbarn». Koblenz.

März 1983 LDV-Kolloquium Trier (Sektionen: Mikroelektronik und Textverarbeitung / Computerunterstützte Inhaltsanalyse / Maschinelle Übersetzung).

Febr. 1984 «Trends in der Linguistischen Datenverarbeitung». Heidelberg.

März 1985 «Sprachverarbeitung in Information und Dokumentation». Hannover.

Febr. 1986 «LDV und Geisteswissenschaften». Göttingen.

MANN, THOMPSON 1983

Mann, W.C., Thompson, S.A.: Relational Propositions in Discourse. Univ. of Southern Carolina, Information Sciences Institute, Rep. ISI/RR-83-115, Marina del Rey, November 1983

McKEOWN 1982

McKeown, K. R.: Generating Natural Language Responses to Questions about Database Structure. PhD Thesis, University of Pennsylvania, Philadelphia, Technical Report MS-CIS-82-5, May 1982

MEEHAN 1980

Meehan, J.R.: The Metanovel. Writing Stories by Computer. New York: Garland 1980

ZADEH 1981

Zadeh, L.A.: Test-score Semantics for Natural Languages and Meaning-Representation via PRUF. In: B. Rieger (ed.): Empirical Semantics. Bochum: Brockmeyer 1981, S. 281-349

Wolfgang Hoepfner, Universität Hamburg, Fachbereich Informatik, Projektgruppe WISBER

Betrifft: Rezensionen

Der Redaktion liegen folgende Bücher als Rezensionsexemplare vor:

- I. Batori, H. Weber (Hrsg.): Neue Ansätze in Maschinellem Sprachübersetzung: Wissensrepräsentation und Textbezug. Tübingen: Niemeyer, 1986 (*Bemerkung: bestens geeignet für den nächsten Themenschwerpunkt!*).
- T. Herrmann: Gestaltung der Mensch-Computer-Interaktion: Systemklärung als kommunikatives Problem. Tübingen: Niemeyer, 1986
- R. Kuhlen: Informationslinguistik. Tübingen: Niemeyer, 1986
- W. Lenders, G. Willée: Linguistische Datenverarbeitung. Ein Lehrbuch. Opladen: Westdeutscher Verlag, 1986

Wissenschaftler(innen), die im LDV-Forum ein Buch besprechen möchten, werden gebeten, sich mit der Redaktion in Verbindung zu setzen. Dabei sind auch hier nicht genannte Bücher als Vorschläge willkommen. Die Redaktion wird sich in letzterem Fall um die Zusendung eines Rezensionsexemplars bemühen.

Für die Bibliographie "**NEUERE VERÖFFENTLICHUNGEN AUS DEM BEREICH DER LINGUISTISCHEN DATENVERARBEITUNG**" werden dem LDV-FORUM vom
GID-Informationszentrum (GID-IZ)
Lyoner Strasse 44-48
D-6000 Frankfurt 71

die Standardprofile **Linguistische Datenverarbeitung** und **Künstliche Intelligenz** der Datenbasis INFO-DATA regelmäßig zur Verfügung stellt. Die weitere Verarbeitung dieser Standardprofile umfaßt

eine inhaltliche Selektion, um Arbeiten ohne Bezug auf die Verarbeitung natürlicher Sprache auszuschließen (die Grenze ist fließend und subjectiv gezogen).

eine formale Selektion (nur deutsch- und englischsprachige Arbeiten, nur Arbeiten, die nach 1983 erschienen sind).

eine radikale Verdichtung auf wesentliche bibliographische Angaben.

Es sei ausdrücklich darauf hingewiesen, daß Bezieher der genannten Standardprofile eine weitaus umfangreichere Unterlage in die Hand, bekommen, die insbesondere auch die Kurzfassungen enthält!

Leser, die hier aufgeführte Literatur bestellen wollen, wenden sich an die Bibliothek des GID/IZ:

GID-IZ Bibliothek, Lyonerstr.44-48, F-71

Tel.: (069) 66 87-374 oder 66 87-373

Verarbeitete Ausschnitte der INFODATA-Datenbank: Standardprofile vom April bis August 1986 (ab Erf-Jahr 8512, Erf-Nr 5401 bis Erf-Jahr 8605, Erf-Nr 1600).

Aufbau einer Dokumentationseinheit:

Autor, A.: **Der Titel in Fettdruck, mit einem Punkt abgeschlossen.** Die Quelle, mit Seitenangabe, *Deskriptor1; Deskriptor2; ...; Deskriptor-n.* Bestellnummer

Morik, K.: **Memorandum zur Situation der Künstlichen Intelligenz angesichts der neuen Förderungspolitik des BMFT.** LDV-Forum, 3(1985)1: 8-15. *Research promotion; Artificial intelligence; University; Industry; Cooperation; Criticism.* 85-05353

Slagle, J. R.; Hamburger, H.: **An expert system for a resource allocation problem.** Communications of the ACM, 28(1985)9: 994-1004, *Expert system; Question answering system; Military matters and technology; Computer network.* 85-05388

Zimmermann, H. H.: **Stand und Perspektiven der praxisorientierten maschinellen Übersetzung.** Sprache und Datenverarbeitung, 9(1985)1: 5-10, *Automatic translation; Requirement; Language barrier; Word processing; Telecommunications.* 85-05402

Wheeler, P.: **LOGOS.** Sprache und Datenverarbeitung, 9(1985)1: 11-21, *Automatic translation; English; German; Dictionary; User support; Word processing.* 85-05403

Pigott, I. M.: **SYSTRAN. Maschinenübersetzung bei der Kommission der EG. Gegenwärtiger Zustand und Geschichte.** Sprache und Datenverarbeitung, 9(1985)1: 22-27, *Automatic translation; EC; Word processing; Translation service; EC-countries.* 85-05404

Ducrot, J. M.: **TITUS IV. System zur automatischen und gleichzeitigen Übersetzung in vier Sprachen.** Sprache und Datenverarbeitung, 9(1985)1: 28-36, *Automatic translation; Multilingual; Dictionary; Syntax; Dialog system; Man-machine communication; Textile engineering; Documentation.* 85-05405

Wilms, F. J. M.: **SUSANNAH. Ein praxisorientiertes maschinelles Übersetzungssystem** Sprache und Datenverarbeitung, 9(1985)1: 37-46, *Automatic translation; Word processing; Microcomputer; Terminology; Pilot application.* 85-05406

Gerhardt, T. C.: **ITS - Informatives Textübersetzungs-System. Leistungskriterium.** Sprache und Datenverarbeitung, 9(1985)1: 47-51, *Automatic translation; Effectiveness; Information value; Text; German; English.* 85-05407

Maas, H. D.: **Organisation und Realisierung des multilingualen Übersetzungssystem EUROTRA.** Sprache und Datenverarbeitung, Tübingen, 9(1985)1: 52-57, *Automatic translation; Multilingual; Project; Computational linguistics; Grammar; Software; EC-countries.* 85-05408

Coward, H.; Standera, O.: **Refereeing and editing problems in electronic journal publications.** Computer compacts, 3(1985)2: 48-51, *Electronic publishing; Journal; Publishers; Author; Information-seeking behavior; Work flow; Man-machine communication; Humanities; Social sciences; Canada.* 85-05411

Univ. des Saarlandes, Sonderforschungsbereich Elektronische Sprachforschung (SFB 100): **Computational linguistics projects at the University of Saarbrücken** Multilingua, 4(1985)3: 165-169, *Computational linguistics; Automatic translation; Man-machine communication; Text analysis; Research and development.* 85-05417

- Univ. of Manchester, Inst. of Science and Technology (UMIST): **Computational linguistics projects at the Centre for Computational Linguistics**. *Multilingua*, 4(1985)3: 170-171, *Automatic translation; Database; Terminology; Multilingual; Japanese; English; EC; Computational linguistics; Research and development*. 85-05418
- Neshitov, V. V.: **System of continuous distributions in the information science and linguistics**. *Automatic documentation and mathematical linguistics* 18(1984)2: 1-12, *Information science; Computational linguistics; Mathematical model; Bibliometrics; Language statistics*. 85-05419
- Volodin, K. I.; Nizker, E. M.: **A system for automatic document indexing using a morphologic analyzer and descriptor dictionary**. *Automatic documentation and mathematical linguistics*, 18(1984)2: 13-19, *Automatic indexing; Dictionary; Text analysis; Morphology; Russian*. 85-05420
- Valkovskii, V. A.: **Semantics of intelligent systems. Database manipulation**. *Automatic documentation and mathematical linguistics*, 18(1984)2: 24-35, *Systems research; Data structure; Logic; Artificial intelligence*. 85-05421
- Belonogov, G. G.; Duganova, I. S.; Kalinin, Yu. P.; Kuznetsov, B. A.; Pozdnyak, M. V.; Shturman, Ya. A.: **An experimental system for automatic detection and correction of spelling errors in texts**. *Automatic documentation and mathematical linguistics* 18(1984)2: 36-40, *Orthography; Correction; Computer program; Morphology; Text analysis; Russian; Testing*. 85-05422
- Bondarenko, G. V.; Yarovenko, O. I.: **Using structural patterns of text in automatic processing of information**. *Automatic documentation and mathematical linguistics* 18(1984)2: 41-52, *Text analysis; Automatic indexing; Word formation; Testing; Comparison*. 85-05423
- Zhilin, S. A.: **Subject indexing as a process of perception**. *Automatic documentation and mathematical linguistics* 18(1984)2: 53-57, *Linguistics; Communication; Context; Semantics; Information retrieval; Indexing*. 85-05424
- Rubashkin, V. Sh.: **Logical language for description of combinability of property terms**. *Automatic documentation and mathematical linguistics* 18(1984)2: 58-69, *Vocabulary; Relation; Semantics; Logic; Model; Lexicology; Computational linguistics*. 85-05425
- Lavrenova, O. A.: **Semantic representation of texts on the basis of a model of knowledge system**. *Automatic documentation and mathematical linguistics* 18(1984)2: 70-82, *Computational linguistics; Text analysis; Semantics; Artificial intelligence; Model*. 85-05426
- Rozenberg, V. Ya.: **Plotting semantic information quantity vs. recipient's thesaurus size**. *Automatic documentation and mathematical linguistics* 18(1984)2: 83-89, *Text analysis; Semantics; Thesaurus structure; Mathematical model; Communication theory*. 85-05427
- Bakulov, A. D.; Cernyj, A. J.; ZIID, Berlin: **Sprachliche Sicherstellung moderner automatisierter Datenbanken**. ZIID-Schriftenreihe (1985): 59, *Information retrieval system; Database; Data structure; Documentary language; Language; Terminological control, STAIRS; DIALOG; RECON; ORBIT/ELHILL; MISTRAL*. 85-05481
- Steels, L.; Campbell, J. A. (Hrsg.): **Progress in artificial intelligence**. Horwood u. Halsted Press (1985): 320 *Artificial intelligence; Expert system; Systems research; Programming; Text analysis; Learning*. 85-05484
- Schank, R. C.: **Looking at learning**. S. 17-29, *Learning; Model; Artificial intelligence*. 85-05485
- Bundy, A.; Byrd, L.; Mellish, C.S.: **Special-purpose, but domain-independent, inference mechanisms** S. 93-111, *Artificial intelligence; Inference; Syntax; Frame*. 85-05487
- Berliner, H. J.: **Multiprocessing and duality in intelligence**. S. 112-128, *Artificial intelligence; Intelligence; Frame; Comparison*. 85-05488
- Chouraqui, E.: **Construction of a model for reasoning by analogy**. S. 169-183, *Artificial intelligence; Inference; Model*. 85-05491
- Pfeifer, R.; Nicholas, D. W.: **Toward computational models of emotion**. S. 184-192, *Artificial intelligence; Human factor; Model; Systems research* 85-05492
- De Mori, R.: **An expert system for speech decoding**. S. 271-289, *Expert system; Speech processing; Algorithm; Frame*. 85-05497
- Maier, D.; Salveter, S. C.: **Supporting natural language updates in database systems**. S. 291-302, *Systems research; Database; Command language; Language*. 85-05498
- Roesner, D.: **Schemata for understanding of argumentation in newspaper texts**. S. 303-311, *Text analysis; Reading; Learning; Model; Artificial intelligence* 85-05499
- Tait, J. I.: **Generating summaries using a script-based language analyser**. S. 312-318, *Text analysis; Method; Semantics*. 85-05500
- Neet, H. E.: **Assoziationsrelationen in Dokumentationslexika für die verbale Sacherschliessung**. Geneva, CH (1984): 78 *Thesaurus research; Thesaurus construction; Lexicology; Content analysis; Reference; Relation; Linguistics*. 85-05514
- Ullman, J. D.: **Implementation of logical query languages for databases**. *ACM transactions on database systems*, 10(1985)3: 289-321, *Database; Command language; Systems research; Implementation; Method; Artificial intelligence* 85-05580
- Bierfert, H.: **Automatische Spracherkennung - von der Theorie zur Praxis. Die Konstruktion eines Spracherkennungssystems als thematischer, methodischer und pragmatischer Prozess**. Niemeyer (1985): 353, *Speech processing; Pattern recognition; Phonetics; Heuristics; Algorithm*. 85-05606
- Bristow, G.: **Electronic speech synthesis. Techniques, technology and applications**. Granada (1984): 358, *Speech processing; Method; Technology; Computational linguistics; Language*. 85-05614

- Helbig, H.: **Forschungsergebnisse auf dem Gebiet der natürlichsprachigen Kommunikation mit dem Rechner.** Neue Technik im Büro, 29(1985)5: 140-142, *Speech processing; Language; Question answering system; Computational linguistics; Trend.* 85-05629
- Hahn, U.; Hammwöhner, R.; Kuhlen, R.; Reimer, U.; Thiel, U.: **TOPIC II/TOPOGRAPHIC II** Automatische Textkondensierung und text-orientiertes Informationsmanagement. Projektziele. State-of-the-art Konstanz, (1984) 72 S., TOPIC-12-84; TOPOGRAPHIC-3-84, *Text analysis; German; Automatic abstracting; Frame; Computational linguistics; Information retrieval; Command language; Computer graphics; Man-machine communication.* 85-05660
- Erdman, H. P.: **The impact of an explanation capability for a computer consultation system.** Methods of information in medicine, 24(1985)4: 181-191, *Expert system; Medicine; Model; Evaluation.* 85-05730
- Schwarze, C.; Wunderlich, D. (Hrsg.): **Handbuch der Lexikologie.** Athenaeum (1985) 491 S., *Lexicology; Word meaning; Semantics; Word formation; Learning; Artificial intelligence* 85-05741
- Blaser, A.; Pistor, P. (Hrsg.): Gesellschaft für Informatik: **Datenbank-Systeme für Büro, Technik und Wissenschaft.** Proceedings, GI-Fachtagung, Karlsruhe, März 1985. Berlin: Springer (1985): 519, *Systems research; Database; Model; Software engineering; CAD; Office organization; Microcomputer; Expert system.* 85-05758 86-00056
- Lesmo, L.; Torasso, P.; Siklossy, L.: **Semantic and pragmatic processing in FIDO. A flexible interface for data-base operations.** Information systems, 10(1985)2: 219-238, *Artificial intelligence; Computational linguistics; Man-machine communication; Database; Relation; Semantic network.* 86-00140
- MacKeown, K. R.: **Discourse strategies for generating natural-language text.** Artificial intelligence, 27(1985)1: 1-41, *Artificial intelligence; Question answering system; Decision support; Generation; Method; Model.* 86-00141
- Saito, S.; Shirai, K.: **Voice I/O system.** S. 289-303, *Speech processing; Speech recognition; Speaker identification; Code; Data input; Data output; Japan.* 86-00169
- Ein-Dor, P.: **The answerability of database queries.** Information systems, 10(1985)3: 261-270, *Artificial language; Man-machine communication; Database; Search query; Command language; Relation; Model; Draft.* 86-00172
- Chiku, K. S.: **The near future of science and technology resources in Japan.** In: Fjaellbrant, N. (Hrsg.): The future of information resources for science and technology and the role of libraries. Proceedings of the 11th meeting of IATUL, Oxford, England, April 1985. Göteborg, (1985) S. 77-87, *Special library; Science and technology; Stock; Grey literature; Information center; Dialog service; Library network; Automatic translation; Japanese; English; Library automation; Japan.* 86-00207
- Lee, R. M.: **Database inferencing for decision support.** Decision support systems, 1(1985)1: 57-68, *Information system; Manager; Decision support; Systems research; Inference.* 86-00224
- Klingbiel, P. H.: **Phrase structure rewrite systems in information retrieval.** Information processing and management 21(1985)2: 113-126, *Automatic indexing; Dictionary compilation; Vocabulary; Terminological control; Thesaurus construction; Phrase-structure grammar;* 86-00232
- Shepherd, M. A.; Watters, C.: **A common interface for accessing document retrieval systems and DBMS for retrieval of bibliographic data.** Information processing and management 21(1985)2: 127-138, *Information retrieval system; Reference database; Man-machine communication; Database; Command language; Information retrieval; Software.* 86-00233
- Smith, F. J.; Devine, K.: **Storing and retrieving word phrases.** Information processing and management 21(1985)3: 215-224, *Dictionary compilation; Method; Storage; Database organization; Data organization; Text analysis; Computational linguistics; Frequency analysis.* 86-0024
- Willett, P.: **An algorithm for the calculation of exact term discrimination values.** Information processing and management 21(1985)3: 225-232, *Automatic indexing; Descriptor; Indicator; Algorithm.* 86-00241
- Gordon, D. B.; Sager, N.: **A method of measuring information in language, applied to medical texts.** Information processing and management 21(1985)4: 269-289, *Text analysis; Frequency analysis; Computational linguistics; Question answering system; Medicine.* 86-00243
- Rousseau, R.: **On relative indexing in fuzzy retrieval systems.** Information processing and management, 21(1985)5: 415-417, *Weighting; Descriptor; Automatic indexing; Information retrieval system; Model.* 86-00249
- MacBride, D. K.; Lambert, J. V.; Lane, N. E.: **The development of a standardized, computer-based abbreviation algorithm.** Behaviour and information technology, 4(1985)3: 239-247, *Speech recognition; Abbreviation; Algorithm; Software; Testing* 86-0026
- Slocum, J.: **A survey of machine translation. Its history, current status, and future prospects.** Computational linguistics, 11(1985)1: 1-17, *Automatic translation.* 86-00261
- Isabelle, P.; Bourbeau, L.: **TAUM-AVIATION - Its technical features and some experimental results.** Computational linguistics, 11(1985)1: 18-27, *1 Automatic translation; Method; Model; Implementation; Evaluation; Canada* 86-00262
- Vauquois, B.; Boitet, C.: **Automated translation at Grenoble University.** Computational linguistics, 11(1985)1: 28-36, *Automatic translation; Method; Model; Implementation; France.* 86-00263
- Devadason, F. J.: **Computerization of deep structure based indexes.** International classification, 12(1985)2: 87-94, *Indexing techniques; Documentary language; Structure; Automatic classification; Code.* 86-00325

- Harper, K.: **The architecture of speech understanding systems.** Information technology training, 3(1985)2: 56-61, *Speech processing; Information system; Information retrieval; Speech input; Speech recognition.* 86-00326
- Hewitt, R. A.: **Signal processing.** Information technology training, 3(1985)2: 62-66, *Speech processing; Speech recognition; Communication engineering; Model.* 86-00327
- Graham, J. A.: **Synthetic speech from English language text.** Information technology training, 3(1985)3: 91-98, *Speech output; Word processing; English; Software.* 86-00328
- Enser, P. G. B.: **Automatic classification of book material represented by back-of-the-book index.** Journal of documentation, 41(1985)3: 135-155, *Automatic classification; Book; Index; Analysis; Method; Testing.* 86-00382
- Dym, E. D. (Hrsg.): **Subject and information analysis.** Dekker (1985) 507 S., *Alphabetical cataloging; Computational linguistics; Automatic translation; Coordinate indexing; Information retrieval; Subject heading; Thesaurus; Classification research; Abstracting.* 86-00411
- Hicks, C. E.; Rush, J. E.; Strong, S. M.: **Content analysis.** S. 57-109, *Computational linguistics; Method; Text analysis; English; Grammar; Syntax; Automatic classification; Information retrieval; Question answering system* 86-00413
- Lehmann, W.: **Machine translation.** S. 110-123, *Automatic translation; Grammar; Trend; Evaluation* 86-00414
- Sager, N.: **Natural language analysis and processing.** S. 154-168, *Computational linguistics; Grammar; Syntax; Semantics; History.* 86-00415
- Stevens, M. E.: **Automatic analysis.** S. 169-178, *Character recognition; Pattern recognition; Speech processing; Text analysis; Automatic translation.* 86-00416
- Daily, J. E.: **Classification and categorization.** S. 327-349, *Subject heading; Classification; Classification research; Context; Semantics; Artificial language; Class; Notation; Faceted classification; Indexing.* 86-00421
- Daily, J. E.: **Natural classification.** S. 350-369, *Classification; Semantics; Context; Notation; Classification system; Preparation; Classification research.* 86-00422
- Mathis, B. A.; Rush, J. E.: **Abstracting.** S. 445-484, *Abstracting; Abstracting service; Automation; Selection; Consistency; Automatic abstracting; Evaluation; Mathematical method.* 86-00426
- Kent, A.; Belzer, J.; Kurferst, M.; Dym, E. D.; Shirey, D. L.; Bose, A.: **Relevance predictability in information retrieval systems.** S. 487-498, *Relevance; Testing; Abstract; Bibliographic record; Document; Selection; Text analysis; Comparison; Information retrieval.* 86-00427
- Hawkins, D. T.; Levy, L. R.: **Front end software for online database searching part 1 - Definitions, system features, and evaluation.** Online, 9(1985)6: 30-37, *Software; User friendly; Automatic searching; Online; Man-machine communication; Microcomputer; Computer.* 86-00439
- Vossen, G.; Brosda, V.: **Die Einrelationen-Benutzerschnittstelle des MEMODAX-Datenbank-Systems.** Angewandte Informatik, 27(1985)12: 534-540, *Database; Man-machine communication; Software; User support.* 86-00475
- Shieber, S. M.: **Criteria for designing computer facilities for linguistic analysis.** Linguistics, 23(1985)2: 189-211, *Computational linguistics; Artificial intelligence; Systems research; Software.* 86-00479
- Evans, R.: **ProGram - a development tool for GPSG grammars.** Linguistics, 23(1985)2: 213-243, *Computational linguistics; Phrase-structure grammar; Software.* 86-0048
- Phillips, J. D.; Thompson, H. S.: **GPSGP - a parser for generalized phrase structure grammars.** Linguistics, 23(1985)2: 245-261, *Computational linguistics; Phrase-structure grammar; Man-machine communication; Software.* 86-00481
- Golding, A. R.; Thompson, H. S.: **A morphology component for language programs.** Linguistics, 23(1985)2: 263-284, *Computational linguistics; Dictionary compilation; Morphology; Software; English; Finnish; French.* 86-00482
- Johnson, M.: **Computer aids for comparative dictionaries.** Linguistics, 23(1985)2: 285-302, *Computational linguistics; Dictionary; Preparation; Computer aided; Software.* 86-00483
- Aarts, J.; Heuvel, T. van den: **Computational tools for the syntactic analysis of corpora.** Linguistics, 23(1985)2: 303-335, *Computational linguistics; Software; Syntax.* 86-00484
- Biber, D.: **Investigating macroscopic textual variation through multifeature/multidimensional analyses.** Linguistics, 23(1985)2: 336-360, *Computational linguistics; Text analysis; Method; Typology; Evaluation.* 86-00485
- Habermann, F. W. A.: **MT with Systran at Karlsruhe Nuclear Research Center.** Terminologie et traduction, (1985)2: 1-8, *Automatic translation; French; English; Text; Time factor; Effectiveness; Nuclear technology.* 86-0053
- Lehmann, W. P.; Bennett, W. S.: **Human language and computers.** Computers and the humanities, 19(1985)2: 77-83, *Computational linguistics; Syntax; Grammar; Phrase-structure grammar; Automatic translation; Semantics; English; Programming language.* 86-00533
- Sedelow, S. Y.: **Computational lexicography.** Computers and the humanities, 19(1985)2: 97-101, *Lexicology; Dictionary compilation; Language statistics; Syntax; Semantics.* 86-00535
- Slocum, J.: **Machine translation.** Computers and the humanities, 19(1985)2: 109-116, *Automatic translation; Translation; Computer aided; Computational linguistics; Artificial intelligence; Translator.* 86-00537

- Rich, E.: **Artificial intelligence and the humanities.** Computers and the humanities, 19(1985)2: 117-122, *Artificial intelligence; Question answering system; Computational linguistics; Humanities.* 86-00538
- Wall, R.: **The 'new' logics and natural language processing.** Computers and the humanities, 19(1985)2: 123-129, *Computational linguistics; Logic; Syntax; English.* 86-00539
- Lowden, B. G. T.; Roeck, A. de: **Generating English paraphrases from relational query expressions.** Behaviour and information technology, 4(1985)4: 337-348, *Dialog system; Man-machine communication; Search; Command language; Language; Generation; English; Software.* 86-00581
- Gazdar, G. ; Pullum, G. K.: **Computationally relevant properties of natural languages and their grammars.** New generation computing, 3(1985)3: 273-306, *Computational linguistics; Grammar; Language; Syntax; Morphology; Artificial language.* 86-00586
- Clancey, W. J.: **Heuristic classification.** Artificial intelligence, 27(1985)3: 289-350, *Expert system; Classification method; Heuristics; Inference.* 86-00587
- Shoval, P.: **Principles, procedures and rules in an expert system for information retrieval.** Information processing and management, 21(1985)6: 475-487, *Search query; Expert system; Semantic network.* 86-0059
- Univ. Hamburg, Forschungsstelle für Informationswissenschaft und Künstliche Intelligenz: **The project HAM-ANS (Hamburg Application-Oriented Natural Language System).** ECCAI newsletter, 2(1985)1: 2-14, *Artificial intelligence; Expert system; Speech processing; Project; Man-machine communication* 86-00597
- Salton, G.: **Thoughts regarding the RIAO-85 conference in Grenoble.** SIGIR forum 18(1985)2/4: 5-9, *Information retrieval; Database; Software; Automatic indexing; Man-machine communication; Artificial intelligence.* 86-00647
- Lesk, M.: **SIGIR 1985.** SIGIR forum, 18(1985)2/4: 10-15, *Information retrieval; Logic; Mathematical method; Search query; Fulltext searching; Semantics; Trend.* 86-00648
- Salton, G.: **Some characteristics of future information systems.** SIGIR forum, 18(1985)2/4: 28-39, *Information retrieval system; Database; Software; Unification; Requirement; Computational linguistics; Question answering system; Weighting; Mathematical model.* 86-0065
- Pratt, G. E. C.: **Using the micro-computer to simplify database access. Designing interfaces to complex files.** Journal of information science, 10(1985)3: 131-138, *Man-machine communication; Micro-computer; Software; Search; Chemistry; Patent documentation.* 86-00681
- Gebhardt, F.: **Querverbindungen zwischen Information-Retrieval- und Experten-Systemen.** Nachrichten für Dokumentation, 36(1985)6: 255-263, *Expert system; Information retrieval system; Information retrieval; Trend; Software; Comparison.* 86-00735
- Curwen, A. G.: **A decade of PRECIS, 1974-84.** Journal of librarianship, 17(1985)4: 244-267, *Automatic indexing; Software; National bibliography; UK PRECIS.* 86-00746
- Steels, L.: **Second generation expert systems.** Future generations computer systems, 1(1985)4: 213-221, *Expert system; Systems research; Heuristics; Learning; Model; Trend.* 86-00784
- Wiederhold, G.: **Knowledge bases.** Future generations computer systems, 1(1985)4: 223-235, *Expert system; Systems research; Heuristics; Learning; Trend.* 86-00785
- Benskin, M.; Cowham, R. H.; Doyle, A.: **A computer-aided system for printing variable text-blocks on crowded maps.** ALLC journal, 5(1985)1/2: 1-24, *Computational linguistics; Software; Typesetting; Method; Map; Language; Geography.* 86-00869
- Eeg-Olofsson, M.: **A probability model for computer-aided word-class determination.** ALLC journal, 5(1985)1/2: 25-30, *Computational linguistics; Word; Class; Mathematical model.* 86-0087
- Lewis, D.: **The development and progress of machine translation systems.** ALLC journal, 5(1985)1/2: 40-52, *Automatic translation; Software; History; Trend; Semantics; Artificial intelligence.* 86-00871 In: UC1 86-0320
- Evans, N.; Pisciotta, H.: **Search Helper. Testing acceptance of a gateway software system.** In: Williams, M. E. (Hrsg.); Hogan, T. H. (Hrsg.); Online Review (Veranst.): National Online Meeting 1985. Proceedings of the sixth National Online Meeting, New York, April/May 1985. Medford, (1985) S. 131-136, *Man-machine communication; Search; Software; Microcomputer; Evaluation; Empirical study; Search Helper.* 86-00913
- Obermeier, K. K.; De Hilster, D.: **DIID - data independent interface for databases.** S. 333-341, *Man-machine communication; Database; Prototype; Speech input; Expert system; Software.* 86-00931
- King, R.; MacLeod, D.: **A database design methodology and tool for information systems.** ACM transactions on office information systems, 3(1985)1: 2-21, *Office organization; Information system; Model; Semantics; Draft; Method; Prototype.* 86-00947
- Marsh, E.; Friedman, C.: **Transporting the Linguistic String Project system from a medical to a Navy domain.** ACM transactions on office information systems, 3(1985)2, S. 121-140, *Text analysis; Computational linguistics; Medicine; Military matters and technology.* 86-00953
- Hafner, C. D.; Godden, K.: **Portability of syntax and semantics in Datalog.** ACM transactions on office information systems, 3(1985)2, S. 141-164, *Question answering system; Computational linguistics; Syntax; Semantics; Datalog.* 86-00954
- Damerou, F. J.: **Problems and some solutions in customization of natural language database front ends.** ACM transactions on office information systems, 3(1985)2, S. 165-184, *Man-machine communication; Question answering system; English.* 86-00955

- Thompson, B. H.; Thompson, F. B.: **ASK is transportable in half a dozen ways.** ACM transactions on office information systems, 3(1985)2, S. 185-203, *Question answering system; Man-machine communication; Change.* 86-00956
- Slocum, J.; Justus, C. F.: **Transportability to other languages. The natural language processing project in the AI program at MCC.** ACM transactions on office information systems, 3(1985)2, S. 204-230, *Computational linguistics; Project; Draft; Language; Model; Multilingual.* 86-00957
- Nicholson, R. T.: **Usage patterns in an integrated voice and data communications system.** ACM transactions on office information systems, 3(1985)3, S. 307-314, *Office organization; Information system; Integrated; Electronic mail; Text; Speech input; Speech output; Use.* 86-00962
- Laubsch, J. (Hrsg.); Gl, FA 1.2: **GWAI-84. 8th German Workshop on Artificial Intelligence, Wingst/Stade, October 1984.** Springer (1985) 292 S., *Artificial intelligence; Method; Expert system; Text analysis; Language; Man-machine communication; Logic; Programming.* 86-00968
- Kilbury, J.: **A modification of the Early-Shieber algorithm for direct parsing of ID/LP grammars.** S. 39-48, *Phrase-structure grammar; Algorithm; Effectiveness; Implementation J. Early; S. M. Shieber; PROLOG.* 86-00972
- Frey, W.: **Syntax and semantics of some noun phrases.** S. 49-63, *Text analysis; Semantics; Syntax; Grammar; Algorithm.* 86-00973
- Granger, R. H.; Eiselt, K. P.; Holbrook, J. K.: **ATLAST: a three-level language analysis system.** S. 64-77, *Text analysis; Model; Language; Syntax; Semantics; Pragmatics; Inference.* 86-00974
- Gehrke, M.: **Ellipsenrekonstruktion in aufgabenorientierten Dialogen.** S. 78-87, *Man-machine communication; Case grammar; Syntax; Semantics; Pragmatics; Question answering system.* 86-00975
- Kobsa, A.: **VIE-DPM: a user model in a natural-language dialogue system.** S. 145-153, *Man-machine communication; User friendly; Language; Dialog system; Inference.* 86-00979
- Shafer, D. G.: **Microcomputer-based expert systems. Where we are, where we are headed.** Expert systems, 2(1985)4: 188-195, *Expert system; Microcomputer; Programming language; Hardware; Trend.* 86-01005
- Bolc, L.; Kowalski, A.; Kozłowska, M.; Strzalkowski, T.: **A natural language information retrieval system with extensions towards fuzzy reasoning.** International journal of man-machine studies, 23(1985)4: 335-367, *Expert system; Medicine; Man-machine communication; Language; Semantic network; Deduction; Software.* 86-01027
- Lindquist, T. E.; Fainter, R. G.; Hakkinen, M. T.: **GENIE: a modifiable computer-based task for experiments in human-computer interaction.** International journal of man-machine studies, 23(1985)4: 391-406, *Communication science; Man-machine communication; Dialog system; Information-seeking behavior; Testing; Software; Speech output.* 86-01029
- Symposium K: **Künstliche Intelligenz/KI-Forschung in der Praxis: Methoden, Werkzeuge und Anwendungen.** In: Online: Fortschritt der Computertechniken im Büro und Betrieb. (1986) 1K1-7K11, *Artificial intelligence; Expert system; Commercial; Effectiveness; Trend; Microcomputer; Man-machine communication.* 86-01064
- Brueckner, T.: **MOLEX-Generator für Verben.** LDV-Info, (1985)5: 5-41, *Dictionary compilation; Lemmatization; Text analysis; Computer program; MOLEX.* 86-01088
- Kolvenbach, M.; Dickgiesser, S. (Mitarb.): **Regeln zur automatischen Erkennung von Satzenden sowie zur Bereinigung der Grossschreibung am Satzanfang.** LDV-Info, (1985)5: 42-81, *Text analysis; Character recognition; Guideline; Context; Dictionary compilation; Computer program; Mannheimer Konvention.* 86-01089
- Brueckner, T.: **Wörterbuchartikel als formale Strukturen.** LDV-Info, (1985)5: 82-89, *Dictionary compilation; Database; Data structure; Computer program.* 86-01090
- Kolvenbach, M.: **Lexikographische Datenbank (LEDA) - Ergebnis-Komponente. Überlegungen zu Verfahren auf der Basis eines vorhandenen Lexikons (DUDEN).** LDV-Info, (1985)5: 90-136, *Database; Lexicology; Dictionary; Structure; Fulltext; Computational linguistics.* 86-01091
- Kupersmidt, Yu. Ya.; Lyubarskii, Yu. Ya.: **Dialogue procedures in automated dispatcher-controlled information systems.** Automatic documentation and mathematical linguistics, 19(1985)1: 23-34, *Information system; Artificial intelligence; Man-machine communication; Semantic network; Energy; USSR YADRO* 86-01129
- Belonogov, G. G.; Zagika, E. A.; Novoselov, A. P.; Khoroshilov, A. A.: **Automatic normalization of words and word combinations.** Automatic documentation and mathematical linguistics, 19(1985)1: 60-67, *Word formation; Unification; Dictionary compilation; Text analysis; Russian.* 86-01132
- Liebowitz, J.: **Expert systems and telecommunications.** Telematics and informatics, 2(1985)3: 195-206, *Expert system; Telecommunications; Trend.* 86-01163
- Parkhurst, C. A. (Hrsg.); Cooper, L. (Mitarb.); Heller, P. (Mitarb.); American Society for Information Science (ASIS): **ASIS '85. Proceedings of the 48th ASIS annual meeting, Las Vegas, October 1985.** Knowledge Industry Publ. (1985) 403 S., *Information retrieval; Man-machine communication; Dialog service; Electronic publishing; User analysis; Expert system; Software; Training in information science.* 86-01231
- Micco, H. M.: **A comparison of subject access systems in medicine: LCSH, MeSH, PRECIS.** S. 41-53, *Automatic indexing; Method; Comparison; Dialog service; Medicine; PRECIS; MeSH; LCSH.* 86-01236

- Tague, J.: **The two-faced interface.** S. 81-85, *Man-machine communication; User friendly; Semantics; Model; Software.* 86-01240
- Devadason, F. J.: **Computer generation of different types of subject index entries based on deep structure of subject indexing languages.** S. 88-96, *Automatic indexing; Syntactic indexing; Generation.* 86-01241
- Crawford, R. G.; Becker, H. S.: **Toward the development of interfaces for untrained users.** S. 236-239, *Man-machine communication; User support; User friendly; Software.* 86-01257
- Jamison, W.; Metzler, D.: **An expert system for statistical consulting.** S. 293-296, *Expert system; Natural language processing; Mathematical method; Software.* 86-01262
- Metzler, D. P.; Noreault, T.; Haas, D. F.; Cosic, C.: **An expert system approach to natural language processing.** S. 301-307, *Computational linguistics; Expert system; Speech processing; Syntax.* 86-01263
- Nemzer, D. E.: **Chemical structure input methods.** S. 308-310, *Dialog system; Chemical formula; Data input; Software; Method; Man-machine communication; Comparison; CAS ONLINE; DARC; CHEMLAB; MedChem.* 86-01264
- Keil, G. C.; Wilms, F. J. M.; Univ. des Saarlandes; BMFT: **Untersuchungen zur anwenderorientierten maschinellen Übersetzung natürlicher Sprache auf der Grundlage des Saarbrückener Übersetzungssystems SUSY. Erfahrungsbericht und Systemdokumentation. Schlussbericht.** Saarbrücken, (1985) 180 S., (Vertrieb: Fachinformationszentrum Energie, Physik, Mathematik: Egenstein-Leopoldshafen, BMFT-FB-ID-85-003), *Automatic translation; Computer aided; Database; Terminology; Software; User friendly; SUSY-BSA.* 86-01280
- Mochmann, E.: **Inhaltsanalyse in den Sozialwissenschaften.** Sprache und Datenverarbeitung, 9(1985)2: 5-10, *Computational linguistics; Text analysis; Communication science; Dictionary compilation; Language statistics; Social sciences General Inquirer; WORDS.* 86-01306
- Mohler, P. P.: **Computerunterstützte Inhaltsanalyse - zwischen Algorithmen und Mythen.** Sprache und Datenverarbeitung, 9(1985)2: 11-15, *Text analysis; Manually operated; Computer aided; Dictionary; Clustering; Computer program; Social sciences; TEXTPACK V.* 86-01307
- Mergenthaler, E.: **Anwendung der Textanalyse bei gesprochener Sprache am Beispiel psychoanalytischer Gespräche.** Sprache und Datenverarbeitung, 9(1985)2: 16-20, *Text analysis; Speech processing; Lemmatization; Psychiatry.* 86-01308
- Mergenthaler, E.; Kächele, H.: **Changes of latent meaning structures in psychoanalysis.** Sprache und Datenverarbeitung, 9(1985)2: 21-28, *Text analysis; Computer aided; Word formation; Word meaning; Class; Model; Psychiatry.* 86-01309
- Sponholz, G.; Baitsch, H.: **Genetische Beratung - Texte, Vokabular, Kategorien und ein Wörterbuch als Grundlagen der computerunterstützten Analyse.** Sprache und Datenverarbeitung, 9(1985)2: 30-34, *Dictionary compilation; Text analysis; Language statistics; Medicine; Psychology.* 86-01310
- Austin, D.: **Vocabulary control and information technology.** Aslib proceedings, 38(1986)1: 1-15, *Terminological control; Automatic indexing; Man-machine communication; Automation; Trend.* 86-01325
- Wothke, K.: **Text-to-speech-systems for German. A short survey.** LDV-Forum, 3(1985)2: 17-23, *Speech output; Transcription; German; Trend* 86-01330
- Meder, N.: **Artificial intelligence as a tool of classification. Or: the network of language games as cognitive paradigm** International classification, 12(1985)3: 128-132, *Classification research; Artificial intelligence; Automatic classification; Model; Comparison.* 86-01335
- Brooks, C. P.; Newell, A. F.: **Computer transcription of handwritten shorthand as an aid for the deaf. A feasibility study.** International journal of man-machine studies, 23(1985)1: 45-60, *Transcription; Manuscript; Machine-readability; Speech output; Method; Social groups.* 86-01339
- Niemann, H. (Hrsg.); DAGM; GI; Univ. Erlangen-Nürnberg: **Mustererkennung 1985. Proceedings, 7. DAGM-Symposium, Erlangen, September 1985.** Springer (1985) 351 S., *Pattern recognition; Speech processing; Expert system; Production engineering; Medicine.* 86-01353
- Schukat-Talamazzini, E. G.: **Schnelle Präselektion von Wörtern aus kontinuierlich gesprochener Sprache.** S. 170-174, *Speech processing; Pattern recognition; Method.* 86-01358
- Hoegen, H.; Marschall, E.; Schmidbauer, O.; Sommer, R.: **Worthypothesengenerierung im Projekt SPICOS.** S. 175-179, *Speech processing; Pattern recognition; Word; Generation; Algorithm.* 86-01359
- Katterfeldt, H.: **Sprachgrenzenbestimmung mit Polynomklassifikatoren.** S. 180-184, *Speech processing; Pattern recognition; Speech recognition; Method.* 86-01360
- Bross, F.: **Der Einfluss von Sprachsignalstörungen auf rechnergestützte Verfahren zur Sprechererkennung.** S. 224-228, *Speech processing; Speech recognition; Pattern recognition.* 86-01361
- Mercier, G.: **Expert systems approach to acoustic-phonetic decoding and word recognition.** S. 290-312, *Speech processing; Expert system; Pattern recognition; Trend.* 86-01363
- Yu, Y. Y.: **Aufbau eines Deutsch-Chinesischen Terminologie-Datenbanksystems.** Nachrichten für Dokumentation, 37(1986)1: 9-13, *Database; Terminology; Microcomputer; German; English; Chinese; Reprography; Testing; Federal Republic of Germany; China.* 86-01391
- Belonogov, G. G.; Kuznetsov, B. A.; Krichevskii, V. K.: **Automatic document indexing for dialogue information systems.** Automatic documentation and mathematical linguistics, 18(1985)4: 71-78, *Automatic indexing; Information retrieval system; Abstract; Freetext search; Effectiveness; Testing.* 86-01421

- Rudkevich, A. V.; Temnoeva, T. A.; Markov, Yu. V.: **Frame model of a subject area and its application in a factographic information system.** Automatic documentation and mathematical linguistics, 18(1985)4: 79-86, *Expert system; Frame; Scope; Facts.* 86-01422
- Manukyan, G. G.; Manukyan, E. N.; Mkrtchyan, O. E.: **Probability-based recognition of input messages in dialogue information systems and methods of its acceleration.** Automatic documentation and mathematical linguistics, 18(1985)4: 101-105, *Computer program; Natural language processing; Text analysis; Algorithm; Time factor; Mathematical model.* 86-01425
- Ivashko, V. G.: **A formalization of inductive reasoning for expert information systems.** Automatic documentation and mathematical linguistics, 18(1985)4: 106-115, *Expert system; Inference; Induction; Logic; Model; Pharmacology.* 86-01426
- Miller, J. J. H. (Hrsg.): **PROTEXT 2. International conference on text processing systems 2, Dublin, 1985.** Boole Press (1985) 220 S., *Layout of publications; Electronic publishing; Typesetting; Word processing; Man-machine communication; Software; Pattern recognition; Programming language.* 86-01427
- Bosc, P.; Chaffaut, A.; Hardy, B.: **EREDIA: a system to develop servers including text processing techniques.** S. 133-144, *Word processing; Layout of publications; Man-machine communication; Software.* 86-01436
- Nanard, J.; Nanard, M.; Sallantin, J.; Haiech, J.: **Using a document preparation system as interface to an expert system.** S. 175-180, *Layout of publications; Academic approach; Man-machine communication; Expert system; Biology; Software.* 86-01439
- Technische Hochschule Ilmenau, Inst. für Informationswissenschaft, Erfindungswesen und Recht (INER): **14. Kolloquium über Information und Dokumentation. Nutzung von Informationstechnologien zur Beschleunigung des wissenschaftlich-technischen Fortschritts und zur Durchsetzung der ökonomischen Strategie. Themenkreis 2b** Ilmenau, DDR (1985) 153 S., *Content analysis; Thesaurus updating; SDI; Updating; Text analysis; Pattern recognition.* 86-01480
- Gering, E.: **Frage der Formalisierung des Erschliessens von Sachverhaltsinformationen aus gesellschaftswissenschaftlichen Texten.** S. 26-34, *Content analysis; Computer aided; Text; Social sciences.* 86-01481
- Dierich, E.: **Zur Anwendung von Methoden und technischen Mitteln der Mustererkennung für die Erschließung von Fachtexten.** S. 35-42, *Content analysis; Classification method; Pattern recognition; Clustering.* 86-01482
- Dobrowski, M.; Jendrzey, U.; Iwicki, R.; Ceslik, R.: **Automatische Erkennung grammatikalischer Kategorien.** S. 58-66, *Content analysis; Text analysis; Grammar; Word; Method.* 86-01484
- Elkerton, J.; Williges, R. C.: **A performance profile methodology for implementing assistance and instruction in computer-based tasks.** International journal of man-machine studies, 23(1985)2: 135-151, *Systems research; Man-machine communication; User support; Software; Draft.* 86-01511
- Wojcik, A. S. (Hrsg.): **AFIPS: National computer conference 1985. July 1985, Chicago.** AFIPS Press (1985) 672 S., *Artificial intelligence; Mainframe; Software engineering; Man-machine communication; Microcomputer; Computer network; Software; Systems research.* 86-01515
- Godden, K.: **Categorizing natural-language queries for intelligent responses.** S. 67-73, *Natural language processing; Evaluation; Search strategy; Man-machine communication; English; DIALOG.* 86-01516
- Thompson, B. H.; Thompson, F. B.: **How to get a large natural-language system into a personal computer.** S. 75-85, *Natural language processing; Implementation; Microcomputer; Data organization; Software; System environment; Software engineering. UNIX.* 86-01517
- Swartout, W. R.: **Knowledge needed for expert system explanation.** S. 93-98, *Expert system; Systems research; Effectiveness; Software engineering; Software documentation; English; XPLAIN.* 86-01518
- MacDonald, D. D.: **Surface generation for a variety of applications.** S. 105-110, *Expert system; Natural language processing; Knowledge representation.* 86-01519
- Bargielski, M.; Zychowski, P.: **Informationswert des Wortes im Verhältnis zum Informationswert des Textes.** In: Technische Hochschule Ilmenau, Inst. für Informationswissenschaft, Erfindungswesen und Recht (INER): 14. Kolloquium über Information und Dokumentation vom November 1985. Themenkreis 1. 1. Ilmenau, DDR (1985) S. 72-78, *Information value; Word; Text; Language statistics; Text analysis; Information theory.* 86-01545
- Mildner, H.: **Textinformation - eine Herausforderung an die Wissenschaftskommunikation.** S. 79-89, *Fulltext database; Text; Model; Selection; Text analysis; Journal; Fulltext searching.* 86-01546
- Jarosch, H.; Müller, H. D.: **Eine Methode zur Intensivierung der Mensch-Maschine-Kommunikation bei der Nutzung des Datenbank-Systems DIALOG.** S. 153-162, *Database system; Man-machine communication; User support; Software; Model; German Democratic Republic; DIALOG.* 86-01549
- Mater, E.: **Automatische Umwandlung von natürlichsprachigen Anfragen in eine Dialog-Recherchesprache.** S. 221-226, *Factual data bank; Data structure; Database system; Natural language processing; Search strategy; Inference.* 86-01555

NEUE BERICHTE, PAPERS, MEMOS ...**... VON DER TECHNISCHEN HOCHSCHULE DARMSTADT**

Karin Ammersbach: *Auswertung protokollierter Online-Recherchen. Projektabschlussbericht (Förderungsmaßnahme F204.08 der GID) (DV-II-86-5), Technische Hochschule Darmstadt, Fachbereich Informatik, August 1986.*

Neben Belkin/Brooks (1983) "Using Discourse Analysis for the Design of Information Retrieval Mechanisms" liegt mit dem Abschlußbericht (DV-II-86-5) für das Projekt "Auswertung protokollierter Online-Recherchen" eine weitere, im interdisziplinären Bereich von Linguistik, sprachorientierter KI und Psychologie angesiedelte Arbeit vor, die mit Hilfe von gesprächsanalytischem Datenmaterial die empirischen Grundlagen für die benutzerfreundliche Gestaltung von Information-Retrieval-Systemen zu erweitern sucht.

Untersuchungsgegenstand der Arbeit sind "reale", auf Tonband protokollierte Retrievalgespräche, die im Design einer Feldstudie im Dezember 1985 an der Landeshochschulbibliothek Darmstadt zwischen Studenten der TH Darmstadt und Rechercheuren des FIZ Technik e.V./Frankfurt aufgezeichnet wurden. Für eine präzisere Einordnung in linguistische Gesprächsklassifikationen wird der Retrievaldialog, dessen Beratungscharakter schon Knorz (1983) und Belkin/Brooks (1983) betonen, mit Hilfe eines elaborierten Schemas (Henne/Rehbock 1982) mit der Gesprächskategorie Beratung verglichen, ein adäquates Analysekonzept (Schank 79:81) wird vorgestellt und unter Rückgriff auf ein, aus einer früheren Untersuchung (Ammersbach, K.: Benutzermodelle für Information-Retrieval-Systeme. Diplomarbeit, 1986) abgeleitetes Ablaufschema wird das transkribierte Gesprächskorpus auf makrostruktureller Ebene untersucht.

Die Organisation der protokollierten Retrievalgespräche in Gesprächsphasen (Eröffnung - Pre-Search Phase - eigentliche Recherche-phase - Vorbereitungen zum Abschluß - Beendigung) wird im einzelnen dargestellt und mit einer reichhaltigen Auswahl von Beispielen belegt. Ferner werden die Auswirkungen von Variablen wie z.B. fachlich stark divergierende Fachgebiete, schriftliche Vorinformationen, Beteiligung/Anwesenheit weiterer "Experten", Zeitdruck und technische Störungen auf den kommunikativen Handlungsablauf an konkreten Gesprächspassagen beobachtet. Daneben wird über die Ergebnisse einer Fragebogenaktion an der gleichen Benutzerstichprobe berichtet, mit der Daten über bestimmte Erwartungen und Erfahrungen der Kunden vor und nach der Recherche erfaßt wurden.

Mit der Beobachtung realer Informationsvermittlungssituationen schafft die Arbeit eine konkrete empirische Grundlage, die für die Weiterentwicklung kooperativer Retrievalsysteme eingesetzt werden kann.

Literatur

- Ammersbach, K. (1986): *Benutzermodelle für Information Retrieval Systeme.* Diplomarbeit. TH Darmstadt, Fachbereich Informatik, Fachgebiet Datenverwaltungssysteme II.
- Belkin, N.J.; Brooks, H. (1983): *Using Discourse Analysis for the Design of Information Retrieval Mechanisms.* I: ACM 1983, S. 31- 49.
- Henne, H., Rehbock, H. (1982): *Einführung in die Gesprächsanalyse.* Berlin, New York: de Gruyter 1982.
- Knorz, G. (1983): *Kooperatives (Referenz-) Retrieval - eine Herausforderung für KI- und IR-Forschung.* (DVII83-1), TH Darmstadt, Fachgebiet Datenverwaltungssysteme II, 1983.
- Schank, G. (1979): *Zum Ablaufmuster von Kurzberatungen - Beschreibung einer Gesprächsstruktur.* In: Dittmann, J. (Hrsg.). *Arbeiten zur Konversationsanalyse.* Tübingen 1979. S. 176-197.
- Schank, G. (1981): *Untersuchungen zum Ablauf natürlicher Dialoge.* München. Max Huber Verlag, 1981.

**Fachbereich 20 (Informatik)**

Forschungsgruppe

Datenverwaltungssysteme II

6100 Darmstadt, Alexanderstr. 22a

Telefon (06151) 16 29 53

PROJEKT SEMSYN
**UNIVERSITÄT
STUTT GART**

Institut für Informatik
Azenbergstr. 12
D-7000 Stuttgart 1

NEUE BERICHTE, PAPERS, MEMOS...
... VON DER UNIVERSITÄT STUTT GART

Dietmar Rösner: When Mariko talks to Siegfried – Experiences from a Japanese/German Machine Translation Project.

In this paper we report on experiences from a 2 1/2 year project that designed and implemented a prototypical Japanese to German machine translation system for titles of Japanese scientific papers. The analysis of the Japanese input and its transformation into a semantic representation is done by FUJITSU's system ATLAS/II. SEMSYN's part is to produce a correct and understandable German text from these interface structures.

Dietmar Rösner: Linguistic Tools and Software Tools of the SEMSYN Project. (discussion material for the Saarbrücker Expert Talk on MT, April 3-4, 1986).

This paper gives an overview of the tools that have been developed during the implementation of the SEMSYN generator for German on a SYMBOLICS lisp machine. These tools include:

- - **interface tools** that provide easy and comfortable communication with the system
- - **experimentation tools**: e.g. SEMNET-EDIT, a tool for interactively editing or creating semantic nets and generating German from them
- - **lexicon tools**: e.g. menu-based interfaces for lexicon maintenance
- - **linguistic tools**: e.g. a formalism for specifying the intended utterances as functional structures and a language for manipulation those structures.

Dietmar Rösner: Von Titeln zu Texten - Zur Entwicklung des Textgenerators SEMTEX.

In diesem Beitrag stellen wir SEMTEX vor, ein am Institut für Informatik der Universität Stuttgart implementiertes experimentelles System zur Generierung von Texten in deutscher Sprache.

Die derzeitige Version von SEMTEX generiert Zeitungsmeldungen wie z.B. die folgende:

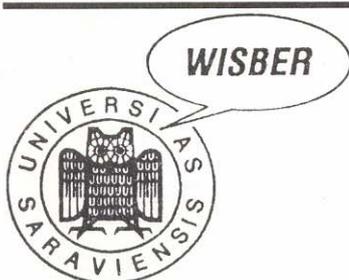
**Geringfügige Reduzierung der Arbeitslosenzahl
Nürnberg/Bonn (cpa)**

Die Zahl der Arbeitslosen in der Bundesrepublik Deutschland hat sich während des Oktober nur sehr wenig verringert. Sie ist von 2151600 auf 2148800 zurückgegangen. Die Arbeitslosenquote hatte Ende Oktober einen Wert von 8.6 Prozent. Sie hatte am Ende des Vergleichszeitraumes des Vorjahrs ebenfalls bei 8.6 Prozent gelegen. Regierungssprecher Ost bewertete die Verringerung der Arbeitslosenzahl positiv. Der stellvertretende DGB-Vorsitzende Muhr erklärt, daß der Rückgang der Zahl der Arbeitslosen nicht darüber hinwegtäuschen dürfe, daß sie jetzt unverändert unerträglich hoch sei.

NEUE BERICHTE, PAPERS, MEMOS ...
... VON DER UNIVERSITÄT SAARBRÜCKEN

H.-U. Block, M. Gehrke, H. Haugeneder, R. Hunze: Neuere Grammatiktheorien und Grammatikformalimen. Report No. 1, Universität des Saarlandes, FR. 10 Informatik IV, 1986

R. Busche, S. op de Hipt, M.-J. Schachter-Radig: Entwurf eines Erhebungsschemas für Geldanlage. Report No. 2, Universität des Saarlandes, FR. 10 Informatik IV, 1986



PUBLIKATIONEN

H.-U. Block, R. Hunze: Incremental Construction of C- and F-Structure in an LFG-Parser. Report No. 4, Universität des Saarlandes, FR. 10 Informatik IV, 1986 (in: Proc. of the 11th Intern. Conf. on Computational Linguistics, COLING'86, Bonn, S. 490-493)

H.-U. Block, H. Haugeneder: The Treatment of Movement Rules in an LFG-Parser. Report No. 5, Universität des Saarlandes, FR. 10 Informatik IV, 1986 (in: Proc. of the 11th Intern. Conf. on Computational Linguistics, COLING'86, Bonn, S. 482-486)

M. Gehrke, H.-U. Block: Morpheme-based Lexical Analysis. Report No. 6, Universität des Saarlandes, FR. 10 Informatik IV, 1986

H.-U. Block, M. Gehrke, H. Haugeneder, R. Hunze: Probleme der Wissensrepräsentation in Beratungssystemen. Report No. 7, Universität des Saarlandes, FR. 10 Informatik IV, 1986

H. Bergmann, M. Gerlach, W. Höppner, H. Marburger: Beratung und natürlichsprachlicher Dialog - eine Evolution von Systemen der Künstlichen Intelligenz. Report No. 8, Universität des Saarlandes, FR. 10 Informatik IV, 1986

M. Fliegner, M.-J. Schachter-Radig: Form der Ergebnisse der Wissensakquisition in WISBER-XPS4 Report No. 9, Universität des Saarlandes, FR. 10 Informatik IV, 1986

INFORMATIONSWISSENSCHAFT UND -PRAXIS IN DER DEUTSCHEN DEMOKRATISCHEN REPUBLIK

Mit einem neuen Band der Reihe Beiträge und Berichte, die über den Entwicklungsstand der Fachinformation im In- und Ausland informieren, legt die GID erstmalig eine Arbeit über ein sozialistisches Land vor:

Wolfgang G. Stock: Informationswissenschaft und -praxis in der Deutschen Demokratischen Republik. (Beiträge und Berichte Reihe B Nummer 7), Frankfurt am Main: IDD Verlag, 1986: 346 S. (DIN A 5), ISBN 3-8151-0035-6

Der Zeitpunkt für eine solche Veröffentlichung ist nach Ansicht des Autors mehr als reif: Die Informations- und Dokumentationswissenschaft, in der DDR auch als "Informatik" geführt, hat sich Mitte der 80er Jahre zu einer anerkannten Wissenschaftsdisziplin entwickelt, die einen Vergleich mit der "information science" in USA oder der "informatika" der UDSSR nicht zu scheuen braucht, über die aber außerhalb der DDR wenig bekannt ist.

Stock will diese Lücke schließen. Er unternimmt dies mit wissenschaftlicher Gründlichkeit und unverhohlener Achtung vor einer ganzheitlichen Konzeption des Informationswesens, der es gelungen ist, eine solide theoretische Fundierung mit Praxisnähe zu verbinden. Er beschreibt die verschiedenen Aspekte der Informationswissenschaft (u.a. die grundlegenden informatischen Theorien, die Meßmethoden der Informatik, die Konzeption einer Informationsökonomie sowie das Verhältnis von wissenschaftlich-technischem Fortschritt und Informatik), um sich dann der Technik und Methodik der wissenschaftlichen Informativität, der Organisation der Informationswissenschaft und -praxis im nationalen wie im internationalen Rahmen und der Ausbildung von Informationsfachkräften zuzuwenden.

Ein umfangreiches, nach Sachgebieten gegliedertes Literaturverzeichnis schließt das Buch ab.

Dr. Johannes Arz
Universität des Saarlandes
FR. 10 Informatik IV
Im Stadtwald 15
D-6600 Saarbrücken 11



Veranstaltungen

BEISPIELEINTRAG

Tagungsbeginn - Tagungsende **STADT**
Titel der Veranstaltung. *Kontaktadresse.*

1987

JANUAR

7. - 9. LAS CRUCES

Third Workshop on Theoretical Issues in Natural Language Processing. *Yorick Wilks, Computing Research Laboratory, NMSU, Box 3CRL, Las Cruces NM 88001*

FEBRUAR

17. - 19. ST. LOUIS

CSC 87: ACM Computer Science Conference. *Arlan DeKnoek, Computer Science Department, University of Missouri-Rolla, Rolla, MO 65401*

22. - 28. ORLANDO

Third IEEE Conference on Artificial Intelligence Applications. *Jan Aikins, 3rd IEEE Conference on Artificial Intelligence Applications, Aion Corp., 101 University Ave., Palo Alto CA 94301, Tel.: 415/328-9595*

26. - 28. HEIDELBERG

Symposium über Semiotik in den Kulturwissenschaften. *Prof. Dr. Dietrich Harth, Germanistisches Seminar, Universität Heidelberg, Hauptstr. 207-209, D-6900 Heidelberg*

MÄRZ

4. - 6. BONN

Jahrestagung der Gesellschaft für Linguistische Datenverarbeitung. *Dr. G. Willé, Institut für Kommunikationsforschung und Phonetik (IKP), Poppelsdorfer Allee 47, Bonn*

3. - 11. DALLAS

OAC 87: Office Automation Conference. *AFIPS, 1899 Preston White Dr., Reston, VA 22091*

16. - 18. ORLANDO

Fifth Annual Conference on Application of Artificial Intelligence. *Jane Huston, SPIE, P.O. Box 10, Bellingham WA 98277, Tel.: 206/676-3290*

23. - 25. SAN DIEGO

Sixth ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database Systems. *Data), ACM SIGART (Artificial Intelligence) Ashok K. Chandra, IBM T.J. Watson Research Center, P.O. Box 218, Yorktown Heights NY 10598, Tel.: 914/945-1752*

APRIL

1. - 3. KOPENHAGEN

Third Conference of the European Chapter of the ACL. *Bente Maegaard, Institut for Anvendt og Matematisk Lingvistik, Kobenhavns Universitet, Njalsgade 96, DK-2300 Kobenhavn S, Denmark*

1. - 3. HOUSTON

Conference on Management and Information Technologies. *Conferences Dep., ACM Headquarters, 11 W. 42nd street, New York, NY 10036*

1. - 3. DARMSTADT

Datenbanken in Büro, Technik und Wissenschaft (BTW '87). *Prof. Dr. H.-J. Schek Fachbereich Informatik, Technische Hochschule Darmstadt, Karolinenplatz 5, D-6100 Darmstadt*

4. - 6. AUGSBURG

Jahrestagung der Deutschen Gesellschaft für Sprachwissenschaft: Sprachliche Strukturen und psychische Prozesse. *Dr. Bruno Strecker, Friedrich-Karl-Straße 12, D-6800 Mannheim*

5. - 9. TORONTO

CHI + GI 1987. *William Buxton, Ronald Baecker, University of Toronto, Computer Systems Research Institute, Toronto, Ontario, Canada, M5S 1A4*

7. - 8. ERLANGEN-NÜRNBERG

Expertensysteme '87 - Konzepte und Werkzeuge (german acm chapter / GI). *Dr. H. Balzert, c/o TA Triumph-Adler AG, Fürtherstr. 212, D-8500 Nürnberg 80*

8. - 10. NEW ORLEANS

Third Workshop on the Mathematical Foundations of Programming Semantics. *Michael Mislove, Mathematica Department, Tulane University, New Orleans, LA 70118, Tel.: 504/865-5727*

9. - 10. SAN JOSE

Western Educational Workshop. *Alexia Devlin, CECC Treasurer, San Francisco State University, Accounting Data, NADM-358, 1600 Holloway Ave., San Francisco CA 94132*

16. - 18. ITHACA

Analyzing the Inchoate: Complex Interrelations in the Humanities and Sciences. *Jonathan Culler, Cornell University Society for The Humanities, Andrew D. White House, 27 East Ave., Ithaca NY 14853, Tel.: 607/255-4086*

21. - 25. AMSTERDAM

International Conference on Word and Image. *Prof. S.A.Varga, CADR: Faculteit der Letteren, Vrije Universiteit, Postbus 7161, NL-1007 MC Amsterdam*

27. - 29. BERLIN

Software-Ergonomie '87: Nützen Informationssysteme dem Benutzer. *Dipl.-Ing. D. Knickriem, awfi, Arbeitswissenschaftliches Forschungsinstitut Berlin, Bayreuther Str. 3, D-1000 Berlin 30*

27. - 29. GAITHERSBURG

Symposium on Office Automation: Integration, Interconnection, and Use of Personal Computers. *Vincent Lum, Dept. of Computer Science, Naval Postgraduate School, Monterey, CA 93942, Tel.: 408/646-2449*

27. - 30. MÜNSTER

Arbeitstagung über Semiotische Theorie im späten Mittelalter und der Renaissance. *Dr. Klaus D. Dutz, Institut für Allgemeine Sprachwissenschaft, Universität Münster, Bispinghof 17, D-4400 Münster*

MAI

im Mai PARIS

Symposium über Semiotik und Epistemologie. *Prof. Dr. Roland Posner, Institut für Linguistik, TU Berlin, Sekr. TEL 6, Ernst-Reuter-Platz 7, D-1000 Berlin 10*

3. - 7. CALGARY

International Conference on Computer-Assisted Learning. *Heidi Guenter, Conference Office, Education Tower Room 102, University of Calgary, Alberta T2N 1N4, Canada*

8. - 10. PITTSBURGH

Third International Conference on Artificial Intelligence and Education, "Beyond Task Analysis". *Joyce Holl, Third Intern. Conf. on Art. Intell. and Educ., L.R.D.C. Bldg, University of Pittsburgh, Pittsburgh PA 15260 Tel.: 412/624-4881*

11. - 15. HAMBURG

International Conference on Advanced Computer Technology, Systems and Applications (CompEuro '87). *Prof. Dr. W. E. Proebster, IBM, Postfach 800880, D-7000 Stuttgart 80*

13. - 15. AVIGNON

7th International Workshop on Expert Systems and Their Applications (Avignon 87). *Agence de l'Informatique, Tour Fiat Cedex 16, 92084 Paris - La Defense, France*

18. - 22. CHICAGO

IEEE Computer Society Tutorial Week, Chicago 87. *Tutorial Week Chicago 87, IEEE Computer Society, 1730 Massachusetts Ave., N.W., Washington, DC 20036-1903, Tel.: 202/371-0101*

25. - 29. PARKVILLE

Fourth International Conference on Logic Programming. *John Lloyd Dept. of Computer Science, University of Melbourne, Parkville, Victoria 3052, Australia*

27. - 29. SAN FRANCISCO

1987 ACM SIGMOD Annual Conference on Management of Data. *Mivhael Stonebraker, 549 Evans Hall, University of California, Berkeley, CA 94720, Tel.: 415/642-5799*

JUNI

3. - 5. NEW ORLEANS

10th Annual International Conference on Research and Development in Information Retrieval. *Donald H. Kraft, Dept. of Computer Science, Louisiana State University, Baton Rouge, LA 70803, Tel.:504/388-1495*

14. - 16. CHICAGO

1987 International Conference on Chinese and Oriental Language Computing. *Patrick S. P. Wang, College of Computer Science, Northeastern University, 360 Huntington Ave., Boston, MA 02115, tel.:617/437-3711*

VERANSTALTUNGEN

22. - 24. UMEA

Nordic Conference on Computers and Computing. *Barbro Heyman, Liaison Office, University of Umea, S-901 87 Umea, Sweden*

22. - 25. KOPENHAGEN

5th Nordic Conference on Bilingualism: Two Languages, two Cultures - Educational Studies. *Royal Danish School of Educational Studies, Department of Danish, Emdrupvej 101, DK-2400 Copenhagen NV*

24. - 26. BEIJING

Second International Conference on Computers and Applications. *IEEE Computer Society, 1730 Massachusetts Ave., N.W., Washington, DC 20036-1903, Tel.: 202/371-0101*

29. - 1. 7. AACHEN

First Conference of the International Federation of Classification Societies. *Prof. Dr. H. H. Bock, Institut für Statistik und Wirtschaftsmathematik, Technische Universität Aachen, Wuellnerstr. 3, D-5100 Aachen*

29. - 8. 7. STANFORD

1987 Summer Institute of the Linguistic Society of America, Contextual and Computational Dimensions of Language. *Dept. of Linguistics, Stanford University, Stanford CA 94305*

JULI

6. - 9. STANFORD

25th Annual Meeting of the Association for Computational Linguistics. *Don Walker, Bell Communications Research, 445 South St., MRE 2A379, Morristown, NJ 07960, Tel.: 201/829-4312*

10. - 24. URBINO

Semiotisches Sommerinstitut. *Christina Chatani, Centre Internationale di Semiotica e de Linguistica, Piazza del Rinascimento 7, I-61029 Urbino*

13. - 17. SEATTLE

AAAI 87. (American Association for Artificial Intelligence) *AAAI 87, 445 Burgess Dr., Menlo Park, CA 94025, Tel.: 415/328-3123*

13. - 17. KARLSRUHE

Fourteenth International Colloquium on Automata, Languages, and Programming (14th ICALP). *ICALP 87, Institut für Angewandte Informatik und Formale Beschreibungsverfahren, Universität Karlsruhe, Postfach 6980, D-7500 Karlsruhe*

19. - 24. LUND

4 Internationaler Kongress zur Erforschung der Kindersprache. *Bodie Martensson, Institut für Linguistik, Universität Lund, Helgonabacken 12, S-22362 Lund*

27. - 31. KASSEL

2. Kongress der Internationalen Gesellschaft für Angewandte Psycholinguistik. *Prof. Dr. H. W. Dechert, Universität Kassel, Heinrich-Plett-Straße 40, D-3500 Kassel*

AUGUST

10. - 15. HONOLULU

Second International Conference on Human-Computer Interaction. *Gavriel Salvendy, Hawaii 1987, School of Industrial Engineering, Purdue University, West Lafayette, IN 47907*

10. - 15. BERLIN (DDR)

XIV Internationaler Linguistenkongress. *Akademie der Wissenschaften der DDR, Otto-Nuschke-Straße 22/23, Postfach Linguistenkongress, DDR-1086 Berlin*

16. - 21. SYDNEY

AILA Congress (Association Internationale de Linguistique Appliquee). *GAL-Sekretariat, Universität Trier, Postfach 3825, D-5500 Trier*

17. - 22. ANTWERPEN

International Pragmatics Conference. *Jef Verschueren, Linguistics, University of Antwerp, Universiteitsplein 1, B-2610 Wilrijk*

23. - 28. MAILAND

IJCAI 1987

24. - 28. SEATTLE

Sixth National Conference on Artificial Intelligence. *Quelle: C.ACM 28(1985)7*

SEPTEMBER

1. - 4. STUTTGART

INTERACT '87. *INTERACT '87 Sekretariat, FHG-IAO, Holzgartenstr. 17, D-7000 Stuttgart 1, Federal Republik of Germany*

21. - 25. BERLIN

7th International Conference on Distributed Computing Systems. *R. Popescu-Zeletin, Hahn-Meitner-Institut Berlin, Glienicke Str. 100, D-1000 Berlin 39, Federal Republik of Germany*

OKTOBER

im Oktober: ESSEN

5. Kongress der Deutschen Gesellschaft für Semiotik (DGS). *Prof. Dr. Werner Enninger CADR: FB 3, Universität GHS Essen, Universitätsstraße 2, D-4300 Essen*

15. - 18. PENSACOLA

12th Annual Meeting of the Semiotic Society of America. *CADR: Semiotic Society Secretariat, P.O.B. 10, Bloomington, IN 47402, U.S.A.*

NOVEMBER

19. - 20. SAN FRANCISCO

Eleventh Western Educational Computing Conference. *Judah Rosenwald, Extended Education, San Francisco State University, 1600 Holloway, San Francisco CA 94132*

DEZEMBER

im Dezember: SALZBURG

6. Symposium der Österreichischen Gesellschaft für Semiotik (OGS). *Mag. Gloria Withalm, Viktoriagasse 14 B/45, A-1150 Wien*

6. - 9. PITTSBURGH

Eighth Annual International Conference on Information Systems. *William R. King, Graduate School of Business, University of Pittsburgh, Pittsburgh, PA 14260, Tel.: 412/648-1587*

1988

JANUAR

4. - 8. CANBERRA

FTPLV-Weltkongress. *FIPLV Head Office, Seestraße 247, CH-8038 Zürich*

MÄRZ

8. - 13. CHICAGO

22nd Annual TESOL Conference. *CADR: TESOL Central Office, 201 D.C. Transit Building, Georgetown University, Washington D.C. 20057, U.S.A.*

JULI

8. - 22. URBINO

Semiotische Sommerinstitut. *Christina Catani, Centre Internationale di Semiotica e de Linguistica, Piazza del Rinascimento 7, I-61029 Urbino*

1989

APRIL

9. - 14. PERPIGNAN und BARCELONA

4. Kongress der Internationalen Vereinigung für Semiotik (IASS/AIS). *Prof. Dr. Gerard Deledalle, Place de l'Eglise, F-34140 Montbazin*

JULI

7. - 21. URBINO

Semiotisches Sommerinstitut. *Christina Catani, Centro Internazionale di Semiotica e di Linguistica, Piazza del Rinascimento 7, I-61029 Urbino*

SEPTEMBER

8. - 10. CAMBRIDGE, MASS.

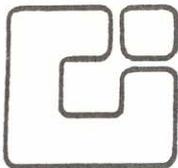
Internationales Colloquium zum 150 Geburtstag von Charles S. Peirce. *Prof. Dr. Kenneth L. Ketner, Institute for Studies in Pragmatism, Texas Tech University, Lubbock TX 79409, U.S.A.*

1990

JULI

13. - 27. URBINO

Semiotisches Sommerinstitut. *Christina Catani, Centro Internazionale di Semiotica e di Linguistica, Piazza del Rinascimento 7, I-61029 Urbino*



BTW 87

Datenbanksysteme für Büro, Technik und Wissenschaft

1. - 3. April 1987

an der TECHNISCHEN HOCHSCHULE DARMSTADT

Ziele der Veranstaltung

In den letzten Jahren wird in wachsendem Maße Datenbankunterstützung für Anwendungsgebiete gefordert, für welche traditionelle Datenbanksysteme nicht gedacht waren. Unter Schlagworten wie "non-standard" oder "nicht-konventionell" lassen sich hier Anwendungen in den Bereichen der Büroinformationssysteme, der Technik (CAD/CAM) sowie der Wissenschaft erwähnen. Charakteristisch für diese Anwendungen ist, daß sie im Zusammenhang mit globalen und lokalen Rechnernetzen auftreten und häufig eingebettet sind in einer integrierten Konzeption von Datenbank- und Expertensystemen.

Vor dem Hintergrund der neuartigen Anforderungen werden Realisierungsansätze für Datenbankgrundkonzepte wie Datenmodelle und Sprachen, Transaktionsverwaltung und Speicherungsstrukturen und das Zusammenfügen dieser Komponenten zu Systemarchitekturen behandelt. Aktuelle Fragestellungen von Büro-Datenbanken, Geo-Datenbanken, CAD-Datenbanken und Software-Produktions-Datenbanken werden entsprechend der Zielsetzung der Tagung vertreten. Außerdem wird über die Erfahrungen im Einsatz von herkömmlichen Datenbanksystemen in diesen Gebieten berichtet.

Das Programmkomitee hat aus den vielen eingereichten Arbeiten eine Mischung aus Langbeiträgen und zahlreichen Kurzbeiträgen zusammengestellt, um sowohl wissenschaftlich gesicherte Ergebnisse zu behandeln, als auch über Projekte zu berichten, die sich zum Teil noch in der Anfangsphase befinden. Es wird so den Teilnehmern aus Industrie und Hochschule Gelegenheit zum Kennenlernen und zu konstruktiver Diskussion geboten.

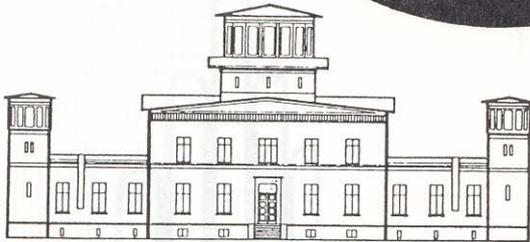
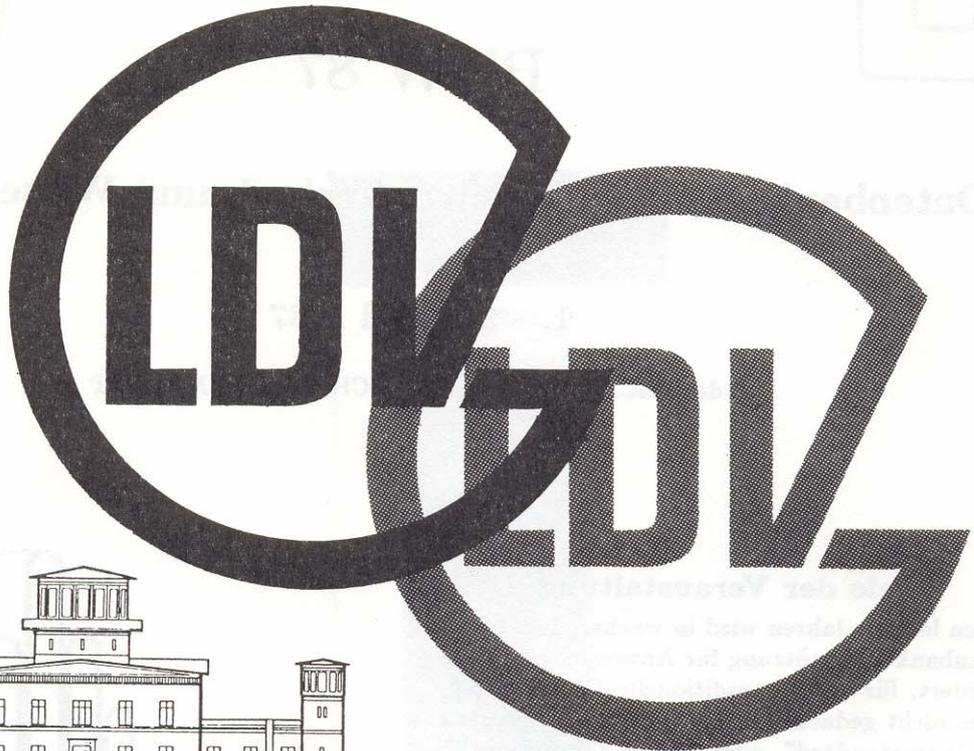


... bitte möglichst frühzeitig anmelden!

Weitere Informationen zur Tagung, insbesondere Tagungsprogramm und Anmeldungsunterlagen, erhalten Sie gleichermaßen über die Redaktion des LDV-Forum und über

Prof. G. Lausen, Fachbereich Informatik, Technische Hochschule Darmstadt, Alexanderstraße 22a, D-6100 Darmstadt

Tagungsgebühr	Mitglieder	Nichtmitglieder	Studenten*
Anmeldung bis 27.02.1987	DM 150,- ○	DM 190,- ○	DM 20,- ○
Anmeldung ab 27.02.1987	DM 200,- ○	DM 240,- ○	DM 20,- ○



JAHRESTAGUNG'87

der Gesellschaft für Linguistische Datenverarbeitung (GLDV)

“Analyse und Synthese gesprochener Sprache”

4. 3.—6.3. 1987 am Institut für
Kommunikationsforschung und Phonetik
der Universität Bonn
Poppelsdorfer Allee 47
5300 Bonn 1

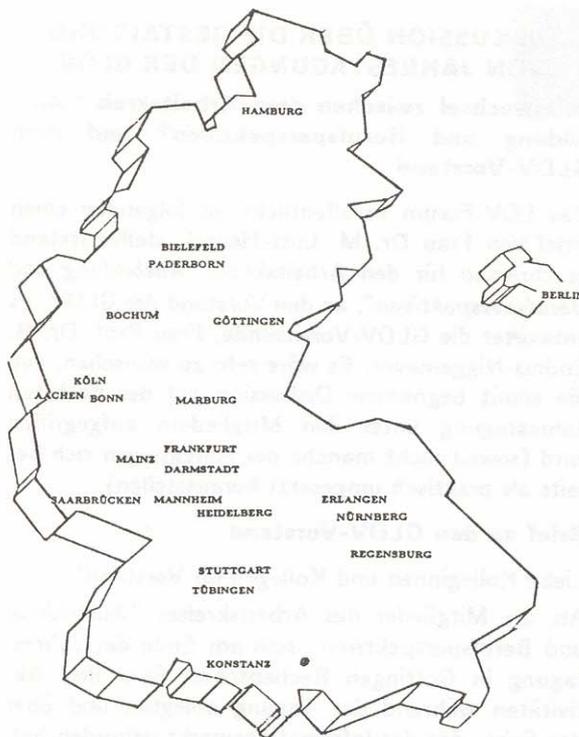


In Zusammenarbeit der FG Sprachanalyse der Gesellschaft für Informatik.

Arbeitskreise

MAP ON CURRENT RESEARCH PROJECTS IN THE FRG RELATED TO THE FIELD OF COMPUTATIONAL LINGUISTICS

July 1986



This map of current CL-research projects in the FRG is to appear in *SPRACHE UND DATENVERARBEITUNG* (1986)2. It is a revised version of that one, completed in March 1986 and sent to all partici-

pants of COLING 86 together with the last circular. The information given in this map is compiled from a special form which was filled by the institutes and projects themselves. The completeness of this map is a function of the initiative of the addressed project groups and of the effort I was able to spend.

It will not be perfect and I apologize for any mistakes and for missing information. Additional information about activities in the field of computational linguistics can be found in the *STUDIENFÜHRER LINGUISTISCHE DATENVERARBEITUNG* (für die wissenschaftlichen Hochschulen der Bundesrepublik Deutschland), compiled by Magdalene Lutz-Hensel and edited by the Gesellschaft für Linguistische Datenverarbeitung e.V. (GLDV). Graduate programs in computational linguistics are available at the universities of Bielefeld, Bonn, Regensburg, Stuttgart, Trier and at the EWH Koblenz. Regular courses in computational linguistics are offered by the universities of Berlin, Bochum, Darmstadt, Düsseldorf, Erlangen-Nürnberg, Essen, Freiburg, Göttingen, Hamburg, Heidelberg, Kassel, Kiel, Köln, Konstanz, Mainz, Marburg, München, Münster, Saarbrücken, Tübingen, Würzburg, Wuppertal (*Studienführer Linguistische Datenverarbeitung*, 1985).

Arbeitskreis "LDV und Nachbarn" der Gesellschaft für Linguistische Datenverarbeitung; Gerhard Knorz, Technische Hochschule Darmstadt.

ARBEITSKREIS SPRACHERKENNUNG, SPRACHGENERIERUNG UND PHONETISCHE DATENBANK

An der für das erste Jahr des Bestehens formulierten Aufgabe des Arbeitskreises wird fleißig weitergearbeitet: Die GLDV-Jahrestagung in Bonn 1987 nimmt konkrete Formen an. Das Programmkomitee kann unter vielen Kurzfassungen auswählen, Prof. Tillmann hat ein kompetentes Podium beisammen und das Organisationskomitee hat die schönen Räume des Rheinischen Landesmuseums für uns bestellt. Herr Knorz hat dem Arbeitskreis eine Dauer-Rubrik im LDV-Forum zugesichert.

Auf Wiedersehen in Bonn, bei der nächsten Arbeitskreissitzung (während der Jahrestagung)

K.G. Schweisthal, Institut für Phonetik und sprachliche Kommunikation der Universität München

ERSTES ARBEITSTREFFEN "LDV-STUDIENGÄNGE"

Inhalte, Ausbildungsprofil und Berufsqualifikation der LDV-Studiengänge in der Diskussion

Die Zahl der LDV-Studiengänge nimmt stetig zu. Die Unterschiedlichkeit der fachlichen und institutionellen Einbindung und die Verschiedenartigkeit der Studiengangstypen mit entsprechend verschiedenen Abschlüssen erleichtert nicht gerade den Über- und Durchblick für

- in der Lehre Tätige,
- Studienanfänger und Studierende
- mit der Studienberatung Betraute und
- das akademische und nicht-akademische Umfeld (z.B. Personalchefs).

Hinzukommt, daß das Berufsprofil für und die Berufsqualifikation von LDV-Absolventen alles andere als klar ist.

Beirat und Vorstand der GLDV haben sich (auf Anregung der Arbeitskreise *Ausbildung und Berufsperspektiven* und *LDV und Nachbarn*) deshalb entschlossen, eine Reihe von Arbeitstreffen zu initiieren. In diesem Rahmen sollen Vertreter etablierter und Initiatoren einzuführender Studiengänge, Vertreter von LDV-Studenten und LDV-Absolventen (im Beruf) gemeinsam die Ausrichtung der Studiengänge, deren Inhalte, Ausbildungsprofil und Berufsqualifikation kritisch gegenseitig unter die Lupe nehmen. Das erste Arbeitstreffen wurde von *P. Hellwig* und *H.D. Lutz* vorbereitet und fand am 30. und 31. August 86 statt. Als (langfristige) Zielvorstellung war vorgegeben,

- fachliche Minimalvorstellungen (relativ zum Studiengangstypus) zu benennen,
- Lehr- und Lernziele bzgl. einzelner Veranstaltungen zu formulieren,
- die "Durchlässigkeit" zwischen einzelnen Studiengängen zu erhöhen,
- das Profil des Faches stärker zu konturieren,
- das Ausbildungsprofil eines Studiengangs herauszuarbeiten,
- die Berufsqualifikation der Absolventen eines Studiengangs konkret angebar zu machen,
- die Studienberatungssituation zu verbessern,
- das Angebot am Stellenmarkt formulieren zu können.

Jeder der insgesamt 24 Teilnehmer hatte zur Vorbereitung des Treffens auf der Basis eines vorbereiteten Fragenkatalogs eine Stellungnahme erarbeitet, die den übrigen Teilnehmern vorab zugeschiedt worden war. Aus den Diskussionen des ersten und insbesondere des zweiten Tages (bei dann allerdings nur noch 13 verbleibenden Teilnehmern) sind zwei kleinere Arbeitsgruppen hervorgegangen, die sich mit *Inhalten und Methoden der LDV/CL (AG1)* und mit *Fähigkeiten von LDV/CL-Absolventen (AG2)* beschäftigten und weiter beschäftigen werden. Über Ergebnisse dieser Arbeit wird in nächster Zeit zu berichten sein. Wer dies nicht abwarten mag oder sein eigenes Engagement einbringen möchte, kann sich an die Ansprechpartner *Hitzenberger, Regensburg (AG1)* bzw. *Lutz, Koblenz (AG2)* wenden.

Red. (unter extensiver Verwendung der Arbeitsunterlagen des Arbeitstreffens)

ARBEITSKREIS TEXTANALYSE: STARTSCHWIERIGKEITEN

Der Arbeitskreis *Textanalyse*, auf der Jahrestagung 1986 unter reger Beteiligung von Interessenten ins Leben gerufen, hatte Startschwierigkeiten. Das erste (vorbereitete) Treffen, das am 19. September in Ulm geplant war und von *S. Göser* organisiert wurde, mußte mangels Zusagen ausfallen. Es mag sein, daß die unterschiedlichen Vorstellungen über das, was man sich unter Textanalyse vorstellen sollte, der tiefere Grund des enttäuschenden Starts waren. Es

kann andererseits aber auch an einer unglücklichen Terminlage gelegen haben. Auf der nächsten Jahrestagung sollte sich klären, ob sich der Arbeitskreis arbeitsfähig einrichten läßt. Kontaktadresse: *S. Göser, Sonderforschungsbereich 129, Universität Ulm, Am Hochsträß 8, D-7900 Ulm*

Red.

DISKUSSION ÜBER DIE GESTALTUNG VON JAHRESTAGUNGEN DER GLDV

Briefwechsel zwischen dem Arbeitskreis "Ausbildung und Berufsperspektiven" und dem GLDV-Vorstand

Das LDV-Forum veröffentlicht im folgenden einen Brief von Frau Dr. M. Lutz-Hensel, stellvertretend geschrieben für den Arbeitskreis "Ausbildung und Berufsperspektiven", an den Vorstand der GLDV. Es antwortet die GLDV-Vorsitzende, Frau Prof. Dr. B. Endres-Niggemeyer. Es wäre sehr zu wünschen, daß die somit begonnene Diskussion auf der nächsten Jahrestagung unter den Mitgliedern aufgegriffen wird (soweit nicht manche der Anregungen sich bereits als praktisch umgesetzt herausstellen).

Brief an den GLDV-Vorstand

Liebe Kolleginnen und Kollegen im Vorstand!

Als die Mitglieder des Arbeitskreises "Ausbildung und Berufsperspektiven" sich am Ende der Jahrestagung in Göttingen Rechenschaft über ihre Aktivitäten während der Tagung ablegten und über das Echo, das der Informationsmarkt gefunden hat, Überlegungen anstellten, da kamen zugleich auch Wünsche zur Sprache, die die Studenten für sich und andere als neue Teilnehmer des Vereins äußerten.

Ich bin gebeten worden, Ihnen diese Wünsche, die Thematisches und Organisatorisches für die nächste Jahrestagung betreffen, vorzutragen:

Als Mitglieder des Arbeitskreises möchten Studierende an der Diskussion beteiligt sein, die sich mit der Verwendung von Begriffen wie "LDV", "CL", "Sprachdatenverarbeitung", "Informationslinguistik", "Informationswissenschaft" usw. befaßt und die eine Standardisierung von Lehrveranstaltungen zur LDV zur Folge haben soll.

Wenn man bedenkt, daß Studierende von ihren Eltern nach diesen neuen Studiengängen befragt werden und daß Studenten bei Studienortwechsel die Argumentation mit Vertretern des BAFÖG-Amtes zu führen haben, wird verständlich, wie dringlich es ist, die Terminologiefragen einer Lösung zuzuführen.

Die studentischen Mitglieder möchten, daß der Austausch über Studiererfahrungen auf Jahrestagungen belebt wird.

Die neuen Vereinsmitglieder haben gebeten, in den Verein wirklich aufgenommen zu werden. Dazu gehört einerseits, daß sie bisherige Mitglieder, auch was deren Arbeitsgebiet und Fachinteressen angeht,

kennenlernen. Dazu gehört andererseits, daß sie selbst Gelegenheit erhalten, sich bisherigen Mitgliedern bekannt zu machen. Das geschieht wohl am besten auf Jahrestagungen.

Wie Kontakte zwischen Vereinsmitgliedern hergestellt werden könnten, das soll noch weiter bedacht werden. Als Vorschläge gebe ich folgendes weiter:

- Mitglieder des Vereins sollten darüber befragt werden, ob sie mit der Gestaltung von Jahrestagungen zufrieden sind oder was sie verändert sehen möchten.

Es wird angeregt, vielleicht in Form einer Wandzeitung auf der Jahrestagung eine Rückmeldung über die Tagungsgestaltung zu erhalten.

- Das Organisationskomitee der künftigen Jahrestagung sollte mindestens von den letztjährigen Organisatoren über Nachahmenswertes und zu Vermeidendes informiert werden. Das betrifft zum Beispiel die Terminplanung für die Veranstaltungen, verteilt auf den Tag, verteilt auf Wochentage, bezogen auf An- und Abreise usw. Ein Mitgliedertreffen am Abend des ersten Tages oder am Vorabend der Tagung wird für sinnvoll gehalten, damit man während der Tagung Bekanntschaften intensivieren kann, - dazu sollte dann auch Zeit vorhanden sein.

Auch wird empfohlen, während der gesamten Tagungsdauer einen Raum als ständiges Kommunikationszentrum vorzusehen, in das man sich bei einer Tasse Kaffee zu einem ruhigen Gespräch zurückziehen kann. (Der Raum um das Tagungsbüro, d.h. um einen Tisch herum, hat ja schon manchmal diese Lücke füllen müssen.)

Sowohl die Gesprächsgruppen in Hannover, die sich nach der Podiumsdiskussion bildeten, als auch die Informationsstände in Göttingen haben gezeigt, daß das Umfeld für ein Auf-Einander-Zugehen vorbereitet sein muß; zweifellos besteht ein Bedürfnis nach den verschiedensten Arten von Informationsaustausch, von Studienberatung bis zu Forschungsplanung, jedoch entsteht Gesprächsbereitschaft während der vortragsfreien Pausen sozusagen nicht "auf Kommando" und schon gar nicht Bereitschaft, aus einer Zuhörerhaltung in eine längere Diskussion zu einem gemeinsamen Thema einzutreten.

- Der Vorstand möge, um das Kennenlernen zu fördern, eine Mitgliederliste anfertigen lassen, die auch Auskunft über die jeweiligen Fachinteressen und die berufliche Tätigkeit in der Industrieforschung, in der Lehre usw. gibt.

Daß von einer Jahrestagung eines Vereins offensichtlich etwas anderes als eine Fütterung mit Vortragstoff erwartet wird - der oft in überaus

gedrängter Form präsentiert wird - hat sicher auch mit der größeren Anzahl von Mitgliedern zu tun, die sich nicht mehr selbstverständlich kennen und über die berufliche Tätigkeit des Gesprächspartners grundsätzlich Bescheid wissen.

Ogleich die Jahrestagungen von Monsterkongressen weit entfernt sind, haben sie mit ihnen doch vieles im Ablauf gemeinsam, so daß es vielleicht gar nicht so verwunderlich ist zu sehen, daß die genannten Wünsche mit Vorschlägen gegen die Sprachlosigkeit konvergieren, wie sie von anderen beobachtet wird (*hier wird auf eine beigefügte Kopie verwiesen, d. Red.*).

Wir möchten auch den Beirat von unseren Vorschlägen in Kenntnis setzen und geben deshalb eine Kopie dieses Briefes mit der Bitte um Weiterleitung an den Sprecher des Beirats, Herrn Prof. Dr. Bátori. (*M. Lutz-Hensel*)

Antwort des GLDV-Vorstands

Liebe Kolleginnen und Kollegen des Arbeitskreises "Ausbildung und Berufsperspektiven",

der Vorstand der GLDV hat sich mit ihrem Brief vom 10.4.86 befaßt. Er begrüßt es, wenn zu einem wichtigen Thema wie der Gestaltung von Jahrestagungen Vorschläge und Meinungsäußerungen aus der Mitgliedschaft kommen. Der Vorstand antwortet Ihnen im LDV-Forum, weil er hofft, dadurch noch weitere Kolleg(inn)en in die Meinungsbildung einzubeziehen.

Natürlich ist dies die Aufgabe des Vorstandes, möglichst das gesamte Meinungsspektrum innerhalb der GLDV zur Entfaltung kommen zu lassen. Dazu gehört auch, daß er sich für das gegenseitige Verständnis unter den Mitgliedern einsetzt.

Ihr Brief enthält beherzigenswerte Vorschläge und auch solche, die der Vorstand realisieren kann. Insbesondere wird an der von Ihnen geforderten Mitgliederliste gearbeitet: Die Mitglieder-Datenbank der GLDV soll zusätzlich zu den Adressen Angaben zu den fachlichen Schwerpunkten und Interessengebieten der Mitglieder aufnehmen. Zu diesem Vorhaben haben andere GLDV-Mitglieder datenschutzrechtliche Bedenken geltend gemacht. Das Thema wird auf der nächsten Mitgliederversammlung diskutiert.

Die Ratschläge, die der AK Ausbildung und Berufsperspektiven an die Veranstalter von Jahrestagungen richtet, sind an das Organisationskomitee der nächsten Jahrestagung in Bonn weitergegeben worden. Da die Mitglieder des AK sich im Raum Koblenz/Bonn konzentrieren, haben sie bei der nächsten Jahrestagung eine besonders günstige Gelegenheit, tatkräftig bei der Ausrichtung der Tagung mitzuwirken und dabei eigene Vorstellungen einzubringen. Das gilt auch für die Belegung des Austausches über Studienerfahrungen. Wenn Sie, liebe Kolleg(inn)en vom AK Ausbildung, als die

in diesem Punkt interessierteste und kompetenteste Gruppe das Thema nicht zur Diskussion stellen - wer sollte es dann tun?

In einem anderen Punkt muß Ihnen der Vorstand höflich, aber bestimmt widersprechen: Es ist kein terminologisches Problem, zwischen "LDV", "Computerlinguistik", "Sprachdatenverarbeitung", "Informationslinguistik" usw. zu vermitteln oder zu entscheiden. Unterschiedliche Benennungen für das Fachgebiet sind nur die Spitze des Eisberges, der aus der wissenschaftlichen Diskussion um richtige Konzepte und richtige Theorien besteht. Diese Diskussion kann nicht so geführt werden, daß es für die Argumentation mit BAFÖG-Ämtern und Eltern zweckmäßig ist. Alle LDV-Studenten sind nachdrücklich eingeladen, sich an der theoretischen Weiterentwicklung der LDV zu beteiligen. Wer LDV studiert, muß sich mit Chancen und Problemen einer jungen Wissenschaft auseinandersetzen - er/sie muß wissen, was er/sie tut. Zu den Problemen einer jungen Wissenschaft gehört es, daß konkurrierende Sichten auf das Fachgebiet vertreten werden. Solange das so ist, stößt auch eine Normierung von LDV-Lehrinhalten an offensichtliche Grenzen.

Noch ein eher persönliches Wort an diejenigen unter Ihnen, die mit einer "Fütterung mit Vortragsstoff" bei Jahrestagungen nicht zufrieden sind. Beispielsweise mir ist am Hören auch von vielen Vorträgen bei einer Tagung gelegen. Im Alltag produziere ich 18 SWS Lehre und habe kaum Freiraum, mich an der wissenschaftlichen Diskussion zu beteiligen. Meine berufliche Situation, die durch dauerndes Produzieren gekennzeichnet ist, unterscheidet sich sehr von der rezeptiven Grundsituation von Studenten. Sie ist aber vergleichbar mit der Lage der meisten aktiven GLDV-Mitglieder. Ich bitte die studentischen Mitglieder der GLDV, sich die anderen Bedürfnisse der beruflich aktiven GLDV-Mitglieder klarzumachen und sie zu respektieren. Eine Jahrestagung der GLDV ist in erster Linie ein Ort wissenschaftlicher Auseinandersetzung. Wer dies als unbefriedigende "Fütterung mit Vortragsstoff" erlebt, hat mehrere Möglichkeiten, diesem Zustand abzuweichen:

- aktiv zuhören und mitdiskutieren
- ein eigenes Rezeptionsprogramm aufstellen und dabei Vorträge minderen Interesses auslassen
- Vorträge halten

In der Tendenz Ähnliches ist zu empfehlen, wenn in den Vortragspausen die erwartete Gesprächsbereitschaft nicht "auf Kommando" aufkommt. Gerade die Student(inn)en unter Ihnen muß ich wieder um Verständnis dafür bitten, daß die berufstätigen Tagungsteilnehmer(innen) dieselbe Situation anders erleben. In den Konferenzpausen wird koordiniert, nachgefragt usw. Mit einer Liste von Gesprächsthemen und Personen im Kopf geht man von einem zum anderen und versucht, alle Fragen und Probleme zu erledigen. Diese

gezielte Gesprächsarbeit steht natürlich in einem gewissen Gegensatz zu einer einfach offenen Gesprächsbereitschaft. Wer findet, daß er/sie in den Konferenzpausen nicht genug ins Gespräch kommt, hat wieder mehrere Möglichkeiten zur Abhilfe:

- Fragen vorbereiten, die man sinnvollerweise in den Pausengesprächen klären kann
- Diskussionsgruppen mit Bekannten bilden, Dritte in das Gespräch einbeziehen
- selbst auf andere zugehen, Gespräche in Gang bringen, aufgeschlossen zuhören
- Vortragspausen in Eigenregie verlängern, wenn einem die Diskussion wichtiger ist

Natürlich können die Veranstalter einer Tagung etwas tun, um günstige Voraussetzungen für das Gespräch zu schaffen. Im Verein können der Vorstand und besonders engagierte Mitglieder mit gutem Beispiel vorangehen, um ein Gesamtklima zu schaffen, in dem Freundlichkeit, Aufgeschlossenheit, gegenseitiges Verständnis und persönliche Initiative selbstverständlich sind. Trotzdem muß der wesentliche Beitrag zum menschlichen Profil der GLDV von den einzelnen Mitgliedern kommen. In diesem Sinne bitte ich Sie, liebe Kolleg(inn)en vom AK Ausbildung, ebenso wie die anderen GLDV-Mitglieder dazu beizutragen, daß zunächst die nächste Jahrestagung der GLDV wissenschaftlich und menschlich ein Erfolg wird. (B. Endres-Niggemeyer)

Arbeitskreis "Ausbildung und Berufsperspektiven" und Vorstand der GLDV

INFORMATIONSMARKT AUF DER JAHRESTAGUNG 1986 Gedanken zu einer Premiere

Eine Reaktion (von studentischer Seite) auf die Diskussion "LDV und Ausbildung" auf der Jahrestagung '85 in Hannover war der Wunsch, im Zeitplan der Tagung einen Freiraum vorzusehen, um den Teilnehmern die Möglichkeit zu geben, Erfahrungen und Informationen auszutauschen.

Dieser Wunsch drückt sicher das Interesse der meisten Tagungsteilnehmer aus, da eine Tagung ja von vielen auch als ein "gesellschaftliches Ereignis" angesehen wird, das heißt, daß der Aspekt der Kommunikation den Teilnehmern wichtig ist.

Gespräche zwischen den studentischen und nichtstudentischen Teilnehmern kommen aber oft trotz Interesses auf beiden Seiten nicht ohne weiteres von selbst zustande. Typische Reaktionen sind beispielsweise: "Nö, zu den Studenten setze ich mich nicht, die wollen doch sicher unter sich bleiben" oder "Ich würd' den/die ja gerne einmal ansprechen, aber ich trau' mich nicht recht. Ich will denen doch auch mit meinen blöden Fragen nicht auf den Wecker fallen".

Auf der diesjährigen Tagung in Göttingen fand auf Anregung von Frau Lutz-Hensel ein "Informationsmarkt" statt, auf dem "Stände" zu bestimmten Themen zu finden waren. Wir hatten uns erhofft, den Marktbesuchern den Einstieg in Gespräche über ein konkretes Thema zu erleichtern. Als "Marktfrauen bzw. -männer" wollten wir Ansprechpartner sein und bei konkreten Fragen weiterhelfen und/oder mit den Interessenten über das jeweilige Thema diskutieren.

Themen der Stände waren:

- Studium: Studiengänge und Studienortwechsel, ...
- Praktikum: Erfahrungsaustausch
- Was bin ich? - das "Etikettenproblem" in der LDV (?)
- Auslandstudium: Informationsmaterial
- O (wie Orientierungs) - phase
- Erfahrungen mit LISP
- Studium der LDV in Göttingen

Die beiden letztgenannten Stände wurden kurzfristig von Studenten aus Göttingen organisiert.

Es kamen in den für den Informationsmarkt vorgesehenen drei Stunden recht viele Besucher und die, die nicht kamen, haben diesen Freiraum sicherlich auch nicht bedauert, so daß zumindest die für Gespräche gewonnene Zeit von allen positiv bewertet wurde. An den einzelnen Ständen kamen einige interessante Gespräche zustande - zum Teil konnte auch bei konkreten Fragen weitergeholfen werden.

Auf besonderes Interesse stieß das Thema "Studium". Der Erfahrungsaustausch dazu schien den Teilnehmern besonders wichtig zu sein, zum einen sicher, weil dieses Thema das naheliegendste ist, zum anderen mußten wir aber auch feststellen, daß die Interessenten so gut wie gar nicht darüber informiert waren, was man wo wie studieren kann - und das, wo es doch DEN STUDIENFÜHRER gibt!

Mit der Organisationsform als solcher waren wir aber doch noch nicht zufrieden. Zwar erleichtert es ein Thema, miteinander ins Gespräch zu kommen, aber reicht es denn aus, Sitzgruppen aufzustellen und Plakate darüberzuhängen?

Was haben die Teilnehmer von uns erwartet: Mehr Hilfestellung, andere Themen, eine Art Podiumsdiskussion?

In Frage stellen wollen wir aber auch unsere Zielsetzung, zumal wir auf der nächsten Tagung wieder "etwas" veranstalten möchten - und sei es nur, um die größten Informationslücken zu schließen.

Wir wollten versuchen, die Studenten in die GLDV miteinzubeziehen. Jetzt fragen wir uns, ob dies überhaupt sinnvoll bzw. notwendig ist und wer sich durch z.B. einen Informationsmarkt angesprochen fühlt. Sind das "nur" Studenten, oder auch nicht-studentische Mitglieder?

Was interessiert die Studenten: Erfahrungsaustausch auf verschiedenen Ebenen oder seminarähnliche Veranstaltungen? Wenn letzteres der Fall ist, welche Themen sollen diskutiert werden? Uns z.B. interessiert die "Etikettenfrage" immer noch; wir versuchen, uns zu diesem Thema als "zukünftige LDV-ler (?)" eine Meinung zu bilden. Dabei wäre es sicher hilfreich, mit den Mitgliedern, die sich schon mit dem Thema befaßt haben, diskutieren zu können. In welcher Form man so etwas sinnvoll organisieren könnte, wissen wir nicht.

Für uns wäre auch wichtig zu erfahren, wie denn die "Nichtstudenten" so eine Veranstaltung wie den Informationsmarkt auf der diesjährigen Tagung bewerten: *Was halten sie überhaupt davon, daß sich Studenten auf der Tagung engagieren?* (Überflüssig? Störend? Sinnvoll?) Gäbe es Fragen, die Sie gerne von den Studenten beantwortet haben möchten, vielleicht bezüglich Ihrer Lehrveranstaltungen, vielleicht sind für Sie die Studenten aber auch schon "Wesen der dritten Art", die Sie (wieder) kennenlernen möchten?

Wir hoffen, daß Sie sich durch diesen Fragenkatalog herausgefordert fühlen, uns Ihre Meinung zu schreiben:

Angelika Franzke / Carla Capellmann, Moselweißerstr. 40, D-5400 Koblenz, Tel. 0261/46411

Ulrike Rhein / Beate Firzlaff, Roonstr. 34, 5400 Koblenz, Tel. 0261/17193

URHEBERRECHTLICHE PROBLEME BEI DER NUTZUNG VON MASCHINENLESBAREN WÖRTERBÜCHERN UND TERMINOLOGIEDATENBANKEN

Am 7. 7. 1986 fand im WZH (IBM, Heidelberg) das 6. Treffen des Arbeitskreises "Maschinelle Lexikographie" statt. Das Thema dieser Veranstaltung ist für die praktische Arbeit sicher vieler GLDV-Mitglieder relevant: die Frage nach dem urheberrechtlichen Schutz maschinenlesbarer Corpora und Wörterbücher. Als Sachverständiger hatte sich freundlicherweise Herr Dr. Jürgen Göbel von der GID zur Verfügung gestellt. Dabei konnte es zunächst nur um die Klärung grundsätzlicher Prinzipien gehen; Herr Dr. Göbel wies nachdrücklich darauf hin, daß auch ähnlich erscheinende Fälle sich gerade in denjenigen Details unterscheiden können, die für die jeweilige Rechtslage entscheidend sind, sodaß ohne genaue Prüfung eines Einzelfalls keine Beurteilungen getroffen werden können. Deshalb soll hier auf eine inhaltliche Zusammenfassung der Sitzung in knapper Form verzichtet werden: zum einen ist das Thema dafür zu komplex, zum anderen werden bestimmte Punkte unter den Juristen kontrovers diskutiert. In der nächsten Ausgabe des LDV-Forum wird jedoch Martin Hackemann (GID) die Frage der rechtlichen Probleme bei der Nutzung von lexikographischen und terminologischen Datenbanken für wissenschaftliche Zwecke umfassend behandeln. Die Ausarbeitung kann vorab bereits angefordert werden.

Stephan Mehl, EWH Koblenz; Ulrich Heid, Universität Stuttgart

Mitteilungen aus der GLDV

AUS DEN GREMIEN DER GLDV

Am 27.8.86 tagten in Bonn der Vorstand und der Beirat der GLDV.

Vorstandssitzung

Der GLDV-Vorstand beschäftigte sich insbesondere mit der Verbesserung der Zusammenarbeit mit der Deutschen Gesellschaft für Sprachwissenschaft. Ab 1988 werden die Tagungstermine von GLDV und DGfS wirksam koordiniert. Die DGfS hat Prof. Günter Öhlschlager mit den Kontakten zur GLDV beauftragt. Seitens der GLDV hält P. Hellwig Verbindung mit der DGfS. Beide Gesellschaften planen eine gemeinsame (kleinere) Tagung. Gemeinsame Interessengebiete sind u.a. Berufsfelduntersuchungen und Lexikographie. K. Wothke hat einen Vorschlag für die inhaltliche Erschließung einer Mitgliederdatenbank der GLDV gemacht. Der AK Lexikographie hat Unterlagen für das geplante Tutorial zusammengestellt. Organisatorische Details sind noch zu klären.

Der AK Ausbildung arbeitet an einer Bestandsaufnahme der bestehenden LDV-Studiengänge. Zu diesem Zweck fand am 30.8.86 ebenfalls in Bonn ein Treffen der LDV-Studiengänge statt. Die GLDV-Reihe beim Olms-Verlag umfaßt nun 5 Bände. Zwei weitere stehen kurz vor der Abgabe. Das Design der Reihe soll verbessert werden.

Die GLDV-Jahrestagung 1985 in Hannover wurde mit einem Verlust von DM 136,39 abgeschlossen. Die Abrechnung der Jahrestagung 1986 in Göttingen ergab einen Fehlbetrag von DM 64,60. Wenn die Kosten für das LDV-Forum 2/1986 und eventuell eine Neuauflage des Faltblattes in Rechnung gestellt werden, sind die Finanzen der GLDV erschöpft. Der Call-for-Papers der Jahrestagung 1987 in Bonn ist erschienen. Die Arbeitskreise der GLDV sollen gebeten werden, das Spektrum der Tagung mit einem Parallelprogramm zu bereichern.

Der Vorstand nahm 14 neue Mitglieder in die GLDV auf und stimmte einem Vereinsaustritt zu.

Beiratssitzung

Der Beirat der GLDV beschäftigte sich ausführlich mit dem LDV-Forum. Der erreichte fachliche Stand des LDV-Forums wurde als sehr befriedigend empfunden. Der Umfang pro Heft sollte sich bei etwa 100 Seiten einpendeln. Der Beirat übernimmt mit dem Redakteur die Referee-Funktionen für das LDV-Forum. G.Knorz legt in der nächsten Beiratssitzung einen Entwurf für ein Redaktionsstatut des LDV-Forums vor. Sobald das LDV-Forum konsolidiert ist, kann ihm ein Newsletter als schnelleres Informationsmedium an die Seite gestellt werden. R.Drewek kümmert sich schon um die Vorbereitung.

Über die Arbeit an einem Konzept für einen DFG- Förderschwerpunkt wurde berichtet. Er soll nach übereinstimmender Meinung des Beirates "Verarbeitung sprachlich repräsentierter Information" heißen. J.Krause überarbeitet in Zusammenarbeit mit der DGfS die Blätter zur Berufskunde für Allgemeine Sprachwissenschaft und Computerlinguistik.

26.10.86

Brigitte Endres-Niggemeyer

DER SCHATZMEISTER INFORMIERT, BITTET, MAHNT, ...

- (1) Wer seine Anschrift/Bankverbindung wechselt, ist herzlich aufgefordert, mir dies mitzuteilen.
- (2) Wer sich dem Einzugsverfahren anschließen möchte, kann dies gerne tun, indem er mir eine Einzugsermächtigung zuschickt, auf der neben Name, Anschrift, Bankverbindung und Betreff (also Jahresmitgliedsbeitrag GLDV) nur noch der Satz notiert sein muß: "Hiermit ermächtige ich ...". Im übrigen verfügt Ihre Bank über entsprechende Vordrucke.
- (3) Wer nun partout darauf besteht, nicht abbuchen zu lassen, sondern selbst überweisen zu wollen, möge dies (sofern nicht offensichtlich längst geschehen) **möglichst umgehend tun!**
- (4) Wer bereits für 1987 überweisen will, möge bitte die geänderten Mitgliedsbeiträge beachten (siehe z.B. im Impressum, S. 2)

Burkhard Schaefer, Fachbereich 3, Universität Essen Gesamthochschule, Postfach 103764, D-4300 Essen 1

NEUE MITGLIEDER IN DER GLDV

Folgende Personen wurden auf der Vorstandssitzung am 29. 8. 86 neu als Mitglieder in die GLDV aufgenommen:

- Brigitte Bartsch-Spörl, München
- Thomas M. Biedassek, Heidelberg
- Birgit Boss, Mühlheim
- Karin Eckert, Quierschied
- Athena Gavrides, Darmstadt
- Frank Gerlinger, Hamburg
- Gerhard Heyer, Nürnberg
- Christel Kemke, Saarbrücken
- Ingo Herbert Kropac, Graz
- Ronald Kühn, Rüsselsheim
- Charsten Malchow, Hildesheim
- Peter Ovenhausen, Koblenz
- Ellen Pröpper, Rüsselsheim
- Wiebke Ramm, Hamburg
- Manfred Hanke

Herr Dr. H. Wulz ist nicht mehr Mitglied der GLDV. Die Mitglieder H.-U. Block, M. Butt, M. Diestelmann und G. Hoffmann waren beim letzten Versand des LDV-Forum unter der der GLDV bekannten Adresse nicht erreichbar. Für Hinweise auf die geänderte Postadresse dieser Personen ist die Redaktion (und der Verein) dankbar.