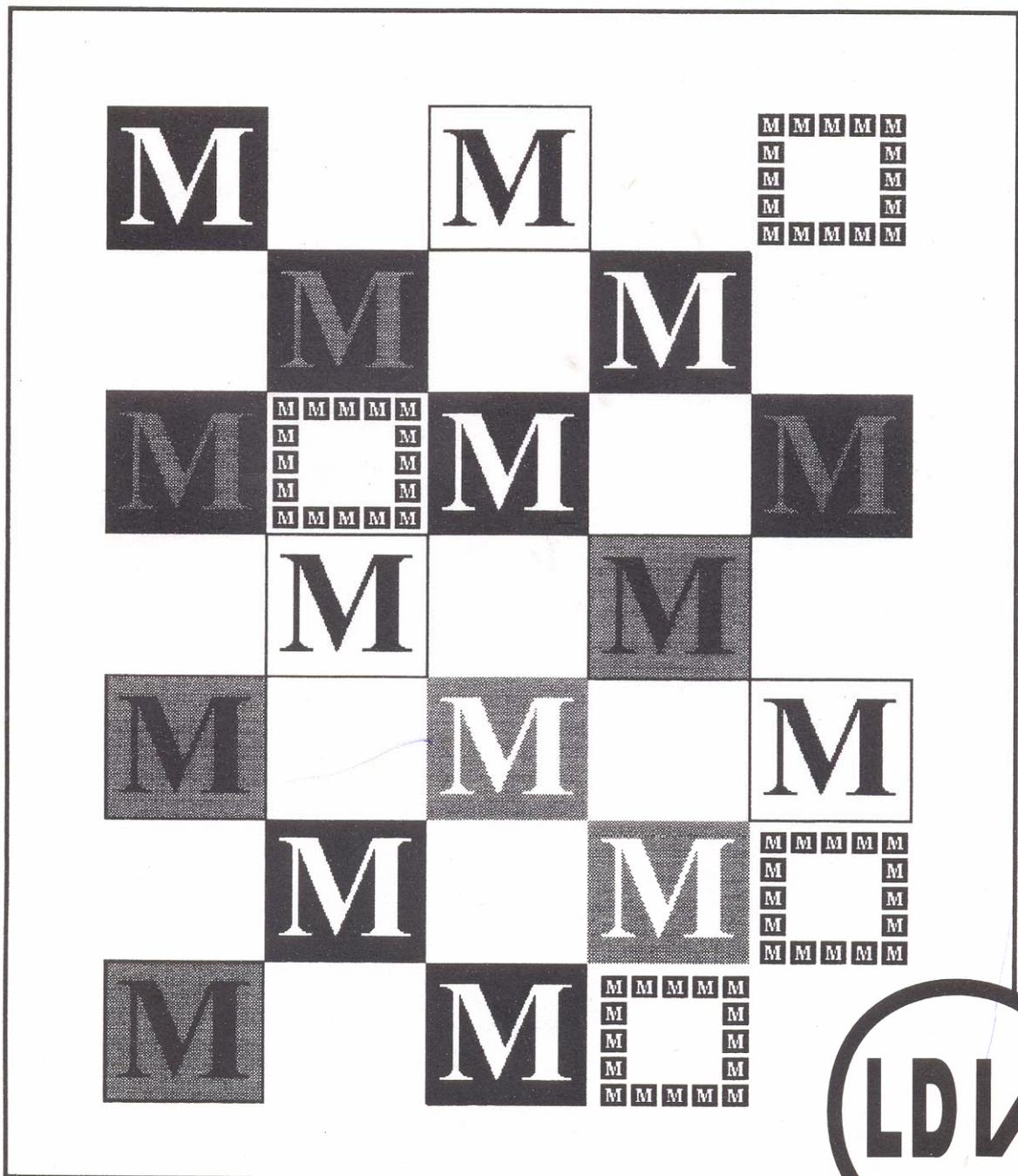


LDV-FORUM

Forum der Gesellschaft für Linguistische Datenverarbeitung GLDV



EDITORIAL	1
Redaktionelles Konzept für das LDV-Forum	3
FACHBEITRAG	
<i>Rothkegel, A.:</i> Statische und dynamische Aspekte einer Textgrammatik	5
REZENSIONEN	
<i>[Mehl, St.:]</i> S. Schierholz: Lexikologische Analysen zur Abstraktheit, Häufigkeit und Polysemie deutscher Substantive	19
Rezensionsangebote	23
NACHRICHTEN	
Götz Greiner verstorben	27
Abstrahieren, Indexieren und Klassieren	27
<i>[Voermanek, A.:]</i> WINGS beflügelt die Geistes und Sozialwissenschaften	28
<i>[Sigel, A.:]</i> Einfacher Zugang zu den Ressourcen im Internet (Gopher)	30
FOCUS COMPUTERLINGUISTIK	
<i>[Bátori, I.:]</i> Computerlinguistik und Informatik: Selbstverständnis einer Disziplin	39
VERANSTALTUNGEN	
Tagungskalender	41
Tagungsankündigungen	44
<i>[Pütz, H. P.:]</i> Bericht GLDV-Jahrestagung	49
MITTEILUNGEN AUS DER GLDV	
GLDV-Newsletter	55
GLDV-Protokolle	55
<i>[Seewald, U.:]</i> Bericht aus dem Beirat	60

LDV - FORUM

Forum der Gesellschaft für Linguistische Datenverarbeitung GLDV

LDV-Forum 10.1 (1993)

Forum der Gesellschaft für Linguistische Datenverarbeitung e.V.

Herausgeber

Prof. Dr. Gerhard Knorz; Gesellschaft für Linguistische Datenverarbeitung e.V. (GLDV)

Anschrift: Fachhochschule Darmstadt, Fachbereich Information und Dokumentation (Iud), Schöffnerstr. 1-3, D-64295 Darmstadt; Tel.: (06151)168490; Fax: (06151)16-8980

Redaktion

Gerhard Knorz, Ute Hauck

Wissenschaftlicher Beirat

Dr. Karin Haenelt, Prof. Dr. Christa Hauenschild, Prof. Dr. Gerhard Knorz, Prof. Dr. Jürgen Krause, Prof. Dr. Burghard Rieger, Dr. Dietmar Rösner, Prof. Dr. Burkhard Schäder

Erscheinungsweise

Zwei Hefte im Jahr, halbjährlich zum 30. Juni und 30. Dezember

Bezugsbedingungen

Für Mitglieder der GLDV ist der Bezugspreis des LDV-Forum im Jahresbeitrag mit eingeschlossen. Jahresabonnements können zum Preis von DM 40,- (inkl. Versand), Einzel Exemplare zum Preis von DM 20,- (zuzügl. Versand) bei der Redaktion bestellt werden.

Editorial

Sie werden es hoffentlich nicht auf den ersten Blick gemerkt haben, aber das Heft, das Sie gerade in den Händen halten, also das *LDV-Forum*, Bd. 10, Nr. 1, ist keineswegs nur das „(n+1)“te Heft der zu keiner Zeit in Routine hergestellten Reihe dieses Organs der Gesellschaft für Linguistische Datenverarbeitung.

Wenn ich mich weit, weit zurückerinnere in die graue Vorzeit (die möglicherweise bunter war als die Gegenwart?) der GLDV, die damals noch den Namen *LDV-Fittings* trug, erinnere ich mich an einen bunten Strauß zusammengelesener Presseauschnitte, angereichert mit kleineren Beiträgen und anregenden Kommentaren, nicht selten in handschriftlicher Form, der den Mitgliedern regelmäßig - fast wie eine persönliche Aufmerksamkeit - als Newsletter auf den Schreibtisch kam. Initiativ und verantwortlich war damals Brigitte Endres-Niggemeyer, zu einem späteren Zeitpunkt Vorsitzende der GLDV, vor der es im übrigen auch gegenwärtig wieder etwas zu berichten gibt (schlagen Sie doch anschließend gleich mal die Rubrik Nachrichten auf).

Die Nachfolge-Redaktion, Rostek/Schulz (GID), gab dem *LDVForum* u.a. den Namen (vorher: *Fittings-Info*) sowie ein neues Erscheinungsbild und kam bei allem Idealismus um die Erfahrung nicht herum, daß die redaktionelle Arbeit einen größeren zeitlichen Aufwand erfordert, als man so vorab veranschlagt und tatsächlich leisten kann. *Wann es denn nun erscheine, das nächste Forum*, war eine Standardfrage im kollegialen Telefongespräch, und eine solche Frage klingt im Wiederwiederholungsfall schon fast ein wenig unhöflich. Das könnte der Grund gewesen sein, wieso ich - so als Nachbar von Darmstadt zu Frankfurt - anfragte, ob ich denn vielleicht irgendwie wirkungsvoll helfen könne; nun ja, so etwas sollte man niemals fragen!

Die Redaktion verlagerte sich also nach Darmstadt und die GID leistete zunächst weiterhin ihren Beitrag, indem sie Vervielfältigung und Versand übernahm. Das Besondere am ersten Heft war das pünktliche Erscheinen, das Bemerkenswerte am zweiten die Gelassenheit des Schatzmeisters: Denn die insgesamt 163 Druckseiten machten nicht nur einen absolut gewichtigen Eindruck, sondern hoben auch die Druckkosten auf ein bisher nicht gekanntes Niveau. Es kennzeichnet im übrigen Burkhard Schaefer, daß die Brille des Schatzmeisters ihn keine Sekunde davon abhielt, die in Gang gekommene Entwicklung aktiv zu unterstützen.

Die Geschichte des *LDV-Forum* in seiner heutigen Erscheinungs-



form begann mit Heft 1986/1 als das Ergebnis mehrerer Entwicklungen: Die Kooperation mit dem Fachbereich Gestaltung der FH Darmstadt brachte ein neues Design-Konzept, dessen Konsequenz das schlanke Format, der Satzspiegel, der kohärenzstiftende und doch offene Rahmen der Titelbildgestaltung sowie die Auswahl der Grauraster für Rubrikköpfe und Titelbild sind. TEX (später LATEX) und Laserdrucker sorgten für ein Layout, das damals noch keineswegs allgegenwärtiger Standard war. Und wie erhofft war die Beziehung zwischen der Entwicklung von äußerem Erscheinungsbild und fachlicher Substanz deutlich wechselseitig und positiv.

Wengleich die Akquisition der Fach- und sonstigen Beiträge, incl. der Anzeigen durchaus permanente Aufmerksamkeit und Zeiteinsatz erforderte, spielte sich doch der eigentliche Höhepunkt des redaktionell/technischen Geschehens in der letzten Woche vor dem Abgabetermin bei der Druckerei ab. Der Ort des Geschehens verlagerte sich dabei gleitend von der Achse Großrechner-Terminal - Laserdrucker (im entfernten Nebengebäude) hin zum Schreibtisch mit Papier, Schere und Montagekleber. Daß für diese Arbeiten schwerpunktmäßig die Nächte auszunutzen waren, muß man als günstig einschätzen, denn der improvisierte Leuchttisch - ein Geschenk von Fred Zimmermann, dem Designer, dem weitgehend das Designkonzept und die Serie der Titelgestaltungen von a bis i zu verdanken sind - war tagsüber sowieso nicht leuchtstark genug und die Lösungsmittel-Konzentration bewegte sich sicher nicht im Rahmen der Arbeitsschutzverordnung.

Die Romantik dieser Zeit fand ein plötzliches Ende, als die Nächte (und nicht nur diese) nicht mehr für Texten, Kleben und Vergleichbares zur Verfügung standen, sondern weitgehend für Stefan und elf Monate danach auch für Katharina (inzwischen gut 4 und 3 Jahre alt) zu reservieren waren. Für halbherzige Entscheidungen bestand keine Realisierungschance, und so mußte die GLDV für die Zeit ab 1991 eine neue redaktionelle Lösung finden. In den letzten drei Jahren kam so das *LDV-Forum* aus Trier: Burghard Rieger - als Vorsitzender der GLDV in dieser Zeit - verantwortete redaktionell zusätzlich auch noch die Zeitschrift! Eine Konstellation, nach der sich vermutlich so schnell keiner drängt, aber von der die GLDV profitierte. Ich habe fast den Eindruck, daß das Dankeschön dafür ein wenig unauffällig ausgefallen ist. Wie dem auch sei, ich will es von dieser Stelle aus nochmal in angemessener Form nach Trier schicken!

Um an den Anfang des Editorials wieder anzuknüpfen: mit dem vorliegenden Heft liegt das Ergebnis einer neuen Konstellation vor, von der sich alle Beteiligten eine stabile Lösung für die Zukunft versprechen: Ich werde das *LDV-Forum* als Herausgeber verantworten und für eine weiterhin positive Entwicklung der Zeitschrift arbeiten. Das IAI in Saarbrücken wird es übernehmen, aus dem Rohmaterial der Beiträge die Druckvorlage für die Druckerei zu erarbeiten und die sonstigen verwaltungstechnischen Anteile abzudecken. Hans Haller und vor allem auch Ute Hauck, die sich mit viel Engagement in das Abenteuer LDV-Forum-Gestaltung gestürzt hat, sei dafür herzlich gedankt.

Titelgestaltung

Ute Hauck, Saarbrücken

Fachbeiträge

Unaufgefordert eingesandte Fachbeiträge werden vor Veröffentlichung von mindestens zwei ReferentInnen begutachtet. Manuskripte (dreifach) sollten daher möglichst frühzeitig eingereicht werden und bei Annahme zur Veröffentlichung in jedem Fall zusätzlich auch noch auf Diskette (5 1/4 bzw. 3 1/2) als

ASCII oder LATEX-Datei übermittelt werden. Formatierungshilfen (*LDVforum.sty*) werden auf Wunsch zugesandt.

Rubriken

Die namentlich gezeichneten Beiträge geben ausschließlich die Meinung der Autoren wider. Einreichungen sind - wie bei Fachbeiträgen - an die Redaktion zu übermitteln.

Redaktionsschluß

Für alle Rubriken mit Ausnahme der als Fachbeiträge eingereichten Manuskripte:

für Heft 10.2/93: 31. akt. 1993; für Heft 11.1/94: 30. Apr. 1994

Herstellung IAI,

Saarbrücken

Druck

reha GmbH, Saarbrücken

Auflage

550 Exemplare

Anzeigen

Preisliste und Informationen: Prof. Dr. Johann Haller, Institut für Angewandte Informationsforschung (IAI), Martin-Luther-Straße 14, D-66111 Saarbrücken; Tel.: (0681) 39313; Fax: (0681) 397482; Email: hans@iai.uni-sb.de

Bankverbindung LDV-Forum (Prof.

Haller): SaarLB Saarbrücken (BLZ 590 500 00) KtoNr. 20 00 21 43

GLDV-Anschrift

Prof. Dr. Winfried Lenders, Institut für Kommunikationsforschung und Phonetik (IKP), Poppelsdorfer Allee 47, D-53115 Bonn; Tel.: (0228) 735638, Fax: (0228) 735639; Email: lenders@uni-bonn.de

Redaktionelles Konzept

Wenn das Heft 93/1 nun doch nicht so ganz im Zeitplan erscheint, und auch in letzter Minute auf eigentlich zugesagte, aber (noch) nicht eingegangene Beiträge verzichtet werden mußte, dann bitte ich dies unter die hoffentlich verzeihlichen Startschwierigkeiten abzubuchen. Vielleicht mögen Sie es auch als Vorkehrung dafür interpretieren, daß für die erste Steigerung der nächsten Ausgabe einfachere Ansatzpunkte gegeben sind. Was die redaktionelle Konzeption und den Anspruch für die Zukunft betrifft, so verweise ich auf die entsprechenden Überlegungen dazu in diesem Heft.

Wenn Sie dieses Editorial mit "ausnahmsweiser" Überlänge bis hierher gelesen haben, so erscheint mir das als außerordentlich günstiges Omen für die Zukunft. Ich bedanke mich dafür bei Ihnen, den LeserInnen und allen Beitragenden - und hoffe im übrigen, daß viele LeserInnen von heute die AutorInnen von morgen sind: davon lebt eine Zeitschrift wie das LDV-Forum!

G.K.

Redaktionelles Konzept

Institutionen und Organisationen werden in der Regel gegründet, weil es eine Aufgabe für sie gibt. Wenn sie nach einer längeren Zeit immer noch existieren, ISO muß das allerdings nicht bedeuten, daß die Lösung dieser Aufgabe (oder einer an deren Stelle getretenen) tatsächlich noch wichtig ist. Es ist nur natürlich, daß eine ins Leben gerufene Organisation mit all den Menschen, zu denen sie wechselseitige Abhängigkeiten aufgebaut hat, ihre eigenen Interessen hat - insbesondere das Interesse, weiter zu bestehen.

Das LDV-Forum als das Organ der Gesellschaft für Linguistische Datenverarbeitung darf es nicht zulassen, unabhängig von einem angemessenen und anerkannten Zweck produziert zu werden. Spätestens mit dem Herausgeber- und Redaktionswechsel 1993 (der bekanntermaßen eine Wiederaufnahme der Redaktionsverantwortlichkeiten unter geänderten

Rahmenbedingungen ist) ist es an der Zeit, diesen Zweck und damit die genuinen Aufgaben des LDV-Forum zu definieren. In etwa einem Jahr könnte es dann an der Zeit sein, den damit verbundenen Anspruch kritisch mit der Realität zu konfrontieren und die richtigen Schlüsse daraus zu ziehen: Das LDV-Forum etwa zugunsten eines (elektronischen) Newsletters in Ehren der Erinnerung zu überlassen oder aber darauf werde ich hinarbeiten - es auf dem richtigen Weg weiter zu begleiten und weiterzuentwickeln.

Was sollten nun die Aufgaben des LDV-Forum sein? Was sind die Bestimmungsmerkmale, aus denen heraus ein spezifischer Zweck definiert werden kann? Das LDV-Forum ist das Organ einer wissenschaftlichen Gesellschaft, ist deutschsprachig und konkurriert mit einer Vielzahl wissenschaftlicher Periodika. Seine Aufgabe muß *Dienstleistung* sein - und zwar in verschiedenster Hinsicht und für verschiedene Zielgruppen.

in fachlicher Hinsicht:

1. Sorgfältig erarbeitete Überblicksartikel über Teilgebiete der Computerlinguistik, die es den LeserInnen ermöglichen, auch außerhalb eigener fachlicher Schwerpunktsetzung auf den aktuellen Stand zu kommen. Diese Artikel müssen geeignet sein, für eine überschaubare Zeit auch als Grundlage von bzw. als Begleitmaterial für Lehrveranstaltungen zu dienen.
Bei der Akquisition dieser Beiträge sehe ich auch die GLDV selbst in der Pflicht.
2. Eine spezielle Variante von Artikeln mit ausdrücklichem "Gebrauchswert" ist für eine ab 1/94 geplante Rubrik *Aus der Lehre für die Lehre* vorgesehen: Nach welcher Konzeption, mit welchen Materialien und mit welchen Literaturgrundlagen werden ausgewählte zentrale CL-Vorlesungen gehalten?

3. Kontroverse Stellungnahmen aus unterschiedlichen Perspektiven zu fachlichen oder fachpolitischen Thesen sollen das LDV-Forum lebendig machen und zur Meinungsbildung innerhalb der GLDV beitragen.
4. Kürzere Artikel sollten ohne lange Wartezeit es den AutorInnen ermöglichen, Forschungsergebnisse zu publizieren und der LeserIn aktuelle und spezifische Information bieten.
5. Da Rezensionen für die Zielgruppe des LDV-Forum sicherlich eine wichtige Rolle spielen, aber wegen des Aufwandes doch nicht beliebig vermehrt werden können, ist eine Ergänzung durch eine Rubrik wie *Quergelesen* sinnvoll: Besprechung von Neuerscheinungen aktuell, wertend, knapp.

in "fachsozialer" Hinsicht:

1. Nicht nur Newcomer der deutschsprachigen computerlinguistischen Szene fühlen sich überfordert, die fachlichen Schwerpunkte und Entwicklungslinien, die technische Ausstattung und institutionelle Einbindung sowie die vielleicht bekannten und die sicher weniger bekannten Personen der einschlägigen Einrichtungen zu kennen. Eine Rubrik *Focus Computerlinguistik* kann die bereits etablierte Rubrik *CL & NLP anderswo* ergänzen und den Rahmen dafür abgeben, im Laufe der Zeit ein Bild der Computerlinguistik im deutschsprachigen Raum zu zeichnen, indem sich in jedem Heft mindestens eine Forschungseinrichtung - durchaus auch etwas persönlich - vorstellt.
2. Die Landschaft der Forschungsförderung und der CL-Projekte ist ein weiteres Feld, über das zwar viele der LDVForum-LeserInnen gut Bescheid wissen mögen, das aber keineswegs für alle ein gut durchschaubares Terrain darstellt.

Auch hier sehe ich Bedarf für eine Rubrik.

in praktischer Hinsicht:

Auch ComputerlinguistInnen können Probleme haben: mit der Stellensuche, mit Softwareauswahl, mit Methodensuche, Elektronische Medien sind grundsätzlich besser als Medium für entsprechende Börsen geeignet als die gedruckte Form des LDV-Forum. Allgemein interessierende Fragen allerdings könnten vom elektronischen GLDV-Newsletter aufgegriffen und etwas ausführlicher behandelt werden.

in vereinspolitischer Hinsicht:

Arbeitskreise der GLDV führten in der Vergangenheit ein oftmals recht stilles Leben. Zu den wenigen sinnvollen Auflagen, die die GLDV einem selbstverantworteten Arbeitskreis machen kann, gehört eine Berichtspflicht für das LDV-Forum.

Ich sehe meine Rolle als Herausgeber so, daß ich mit meinem Engagement, mit dem Ideen-Produzieren und -Aufgreifen, mit Ansprechen, sanftem Drängen und stetem Nachfragen für das Umsetzen dieses redaktionellen Konzeptes sorgen will. Es muß aber jedem klar sein, daß dieses Unterfangen zum Scheitern verurteilt ist, wenn die Resonanz der Fachwelt, insbesondere die der GLDV-Mitglieder ausbleibt.

Nun, nachdem ich glaube, schon auf gute Erfahrungen in der Vergangenheit verweisen zu können, blicke ich optimistisch in die (LDV-Forum-)Zukunft. Nicht zuletzt auch, weil das IAI mit der Übernahme der technischen Herstellung einen ganz wichtigen Beitrag zum Gelingen des Projektes leistet. Und Ende 1995, also nach 2 Jahrgängen, sollten wir kritisch überprüfen, inwieweit wir das Angestrebte erreicht haben. Aber denken Sie daran: Es ist wirklich alles andere als eine leere Phrase: *Der Erfolg hängt auch und gerade von Ihnen ab!*

G.K.

STATISTISCHE UND DYNAMISCHE ASPEKTE EINER TEXTGRAMMATIK

Annely Rothkegel
Universität des Saarlandes
Saarbrücken

Vorbemerkung

Die computerlinguistische Arbeit am Text (Textanalyse, Textgenerierung, Information Retrieval innerhalb von Texten) verlangt nicht nur Festlegungen hinsichtlich der Repräsentationen und Verfahren, sondern setzt konzeptuelle Arbeit hinsichtlich des Gegenstands und der Zwecksetzung der anvisierten Computerprogramme voraus. Im weiteren scheint es im immer noch recht jungen Bereich textorientierter Forschung notwendig zu sein, die jeweilige Position im heterogenen Wald der gegenwärtigen Ansätze und Terminologien klarzulegen. Wir widmen diesem Anspruch relativ viel Raum, um allgemeine Prinzipien einer computerorientierten Textgrammatik zu erläutern. Grundlegend ist die Unterscheidung von Ansätzen, die Texteigenschaften in der Satzgrammatik behandeln, so daß vorhandene Satzgrammatiken erweitert werden können, und solchen, die dem Text eine eigenständige Struktur zuschreiben und die Effekte dieser Struktur sowohl für die gesamte sprachliche Einheit als auch für Lexik und Grammatik des Satzes untersuchen. Dies verlangt dagegen eine Neuentwicklung und die Diskussion grundsätzlicher Fragen. Unser Ansatz macht einige Vorschläge in diese Richtung. Dabei werden theoretische Modelle, die in der Textlinguistik in den letzten Jahren etabliert

worden sind, aufgegriffen und an Zwecke der automatischen Verarbeitung angepaßt. Ziel ist eine Verbindung von textexternen Faktoren mit ihren Realisierungen auf der Textoberfläche und im Einzelsatz. Dies wird erreicht in einem Schichtenmodell, das mehrere Repräsentationsebenen miteinander verbindet.

Die Arbeiten an einem entsprechenden Textmodell sind u. a. als Teil eines umfassenderen Esprit-Projekts zur Untersuchung von Textfunktion und Textrepräsentation (DANDELION: Discourse Functions and Discourse Representations) zu sehen, in dem ebenfalls die satzorientierte Forschungsrichtung eingebracht wird. Ziel dieses Projekts ist, die verschiedenen Teilmodelle als Paket in einer Werkbank für die Arbeit am Text verfügbar zu machen.

0 Fragestellung

Die Aufgabe, Texte automatisch zu erzeugen, wurde zunächst im Rahmen von Expertensystemen aufgegriffen (z.B. McKeown mit TEXT 1985). Hier ging es vor allem darum, Informationen über bestimmte Fakten (aus einer Datenbank) in Form einer kohärenten Folge von Sätzen verfügbar zu machen. Die Idee war, zusätzlich zu Problemen der Satzgenerierung textspezifische Eigenschaften zu berücksichtigen, die im Einzelsatz eine Rolle spielen

len, z.B. Pronominalisierungen oder die Wahl von definitem bzw. indefinitem Artikel. Im Vordergrund dieser Ansätze stand gemäß der KI-Orientierung die Wissensbasiierung, linguistischer Ausgangspunkt war der Satz. 'Text' bedeutete in diesem Zusammenhang hauptsächlich Linearisierung von Wissensinhalten (WHAT) und deren sprachliche Realisierung (HOW) in Kombinationen von Sätzen (vgl. Rothkegel 1989, 1992). Die letztlich unbefriedigenden Ergebnisse verlangten aber nach weiteren Perspektiven. So rückten Textprobleme als solche zunehmend in das Interesse computerlinguistischer Forschung (vgl. Sammelbände wie Kempen 1987, Zock & Sabah 1988, Dale et al. 1990, Paris et al. 1991). Sie spiegelt die beiden grundsätzlichen Vorgehensweisen, die sich auch in textlinguistischen Forschungen beobachten lassen. Die eine Richtung geht von der Satzebene aus und schaut, welche Text- und Kontexteigenschaften sich dort niederschlagen. Dies bedeutet in der Regel, daß vorhandene Satzmodelle entsprechend um eine textbezogene Komponente erweitert werden. Die andere Richtung geht vom Text als Gesamteinheit aus und schaut, welche Textstrukturen als solche und in ihrer Konsequenz auf der Satzebene beschreibbar sind. Dies bedeutet die Annahme einer autonomen Textstruktur, die - ähnlich wie die Satzstruktur - ihren eigenen spezifischen Bedingungen unterliegt. In Textgenerierungsansätzen kam ein 'Mehr' an Textstruktur schließlich über zwei Wege hinzu. Einmal durch Berücksichtigung hierarchischer Strukturen des Textinhalts (z.B. in Hierarchien von Prozessen, die u.a. in Kochrezepten (Dale 1990) oder Fehlermeldungen beim Nicht-Funktionieren von Motoren (Grishman 1990) eine Rolle spielen. Zum andern durch die Annahme rhetorischer Strukturen im Text, die adressatenbezogen aufgebaut sind und zum Teil alternative sprachliche Realisierungen steuern (Hovy 1988). Beide Möglichkeiten führten zu eigenständigen 'Textstrukturen', die

den Gesamttext als solchen erfassen.

In diesem Beitrag stellen wir einen Ansatz vor, der auf der Annahme einer globalen Textstruktur gründet. Dabei verfolgen wir das Ziel, relevante Texteigenschaften konsistent und integrativ in einem zusammenhängenden Modell zu erfassen, das die Formulierung einer Textgrammatik gestattet. Als Anforderung gilt, daß die behandelten Regularitäten in deklarativer Form repräsentierbar sind (Rothkegel 1993a, 1993b).

Konsistenz wird erreicht durch eine einheitliche Perspektive der behandelten Fragestellungen. Die von uns gewählte Perspektive ist die der Textproduktion. Innerhalb dieses Rahmens gehen wir von drei Festlegungen aus.

- ▷ Dynamischer Aspekt: Text wird betrachtet als Resultat strukturbildender Prozesse, die durch TextschreiberInnen ausgeführt werden. Diese Sichtweise setzt einen kognitiven Ansatz voraus, der - hier in diesem Beitrag - textlinguistisch ausgeführt wird. Als Zusammenfassung der in diesem Bereich auftretenden Probleme gilt die Frage "was kommt als nächstes?" (Fraurud & Hellmann 1985a). Die tragende Vorstellung ist die, daß Text bzw. Textstruktur sukzessive auf der Grundlage von zwei Prinzipien aufgebaut werden: der Wiederholung von bereits Eingeführtem und der Hinzufügung von neuem Material (vgl. auch "kommunikativer Dynamismus" der Prager Schule). Die hier vorliegenden Ansätze im Rahmen der Textgenerierung unterscheiden sich darin, auf welche Einheiten die Prinzipien von Wiederholung und Wechsel angewendet werden. Im Fokusmechanismus von McKeown (1985) geht es um satzbezogene Einheiten, d.h. um die Frage, welche der Nominalphrasen in aufeinanderfolgenden Sätzen Fokus wird (mit den entsprechenden Bedingungen hinsichtlich der Selektion von bestimmtem bzw. unbestimmtem Artikel und Wortstellung). Fraurud & Hellmann (1986a) beziehen sich auf konzeptuelle Einheiten, über die Weltwissen in den Text eingebracht wird.

In dieser Sichtweise, die wir weiter verfolgen wollen, geht es um die textthematische und textfunktionale Struktur des Textes (z.B. thematische Behandlung von Sachverhalten in deskriptiven, narrativen, argumentativen, etc. Texten).

=> Wissen und Textstruktur: Textthematische und textfunktionale Strukturen stehen mit Prinzipien der Kohärenzbildung in Zusammenhang (Fritz 1982). Kohärenzbildung ist textgrammatisch interessant, weil sie geeignet ist, den Gegenstand 'Text' vom Gegenstand 'Satz' zu unterscheiden. Sie bildet neben dem Bereich der Kohäsion (vgl. Trennung von textinternen und textexternen Faktoren in de Beaugrande & Dressler 1981; textinterne Faktoren in Halliday & Hasan 1978) den spezifischen Gegenstand einer textgrammatischen Beschreibung. Hierzu noch ein paar erläuternde Worte: Kohärenz bezieht sich auf textexterne Eigenschaften wie Wissen über Sachverhalte und kommunikative Zwecke von Texten. Sie sind für unsere Darstellung insofern relevant, als sie sich im Sprachgebrauch und damit in der Selektion der sprachlichen Mittel zur Erzeugung der Textstruktur niederschlagen. Ein Zugang zu solchen textexternen Informationen eröffnet sich über sogenannte Schemata, d.h. Wissensorganisationen in bezug auf konventionelle Standards, die auf Erfahrungen bzw. Erwartungen im Rahmen sprachlicher Kommunikation eine Rolle spielen (vgl. Framekonzept ausgehend von Bartlett 1932, für KI nutzbar gemacht vor allem durch Schank & Abelson 1977, Minsky 1977 und angewandt auf textorientierte Forschung in van Dijk 1980, van Dijk & Kintsch 1983, Metzger 1980, de Beaugrande & Dressler 1981, de Beaugrande 1984, Fraurud & Hellmann 1985a, 1985b, Rothkegel 1990, Tonfoni 1990). Kohärenz wird üblicherweise beschrieben in Form von Relationen zwischen Texteinheiten. Sie lassen sich danach unterscheiden, ob sie die Gesamtstruktur des Textes in der Weise einbeziehen, daß alle Teile untereinander relationiert sind (global) oder ob sie eine einzelne Beziehung

zwischen zwei Textteilen herstellen (lokal, vgl. Diskussion von Kohärenzrelationen in NoeI1990).

=> Trägermedium und Direktionalität:

Texte gehören zu den Medien der Kommunikation, die an eine lineare Form gebunden sind. In unserem Ansatz gehen wir davon aus, daß die Bedingungen der Linearität einen Einfluß haben auf die Möglichkeiten textueller Strukturierung von Inhalten und deren sprachlicher Realisierung. Im Sinne der Dynamik der Textproduktion (siehe oben) erscheint die Linearität gerichtet. Die Hypothese ist, daß Textstrukturbildung entsprechend in zwei Richtungen festgelegt ist: zum einen von links nach rechts (LIRE), zum andern von rechts nach links (RELI). Jede Produktionsrichtung hat eine spezifische Funktion. LIRE-Produktionen sind vorwärts gerichtet. Sie zielen auf die Vervollständigung der Textstruktur und sind daher global determiniert. Globalität umfaßt dabei sowohl den Bereich des materiellen Textes, der als Ganzheit gesehen wird, als auch die Beschreibungsebene, die eine vollständige Struktur in der Weise liefert, daß die Beschreibungskategorien selbst untereinander relationiert sind. Hinsichtlich der Kohärenz geht es darum, hierarchische Strukturen der Wissens-ebene in Sequenzen der Textstruktur zu überführen. RELI-Produktionen sind dagegen rückwärts gerichtet. Sie sorgen für die Anknüpfung an bereits vorhandenes Material und sind so lokal determiniert.

Um mögliche Mißverständnisse zu vermeiden: die Annahme der Direktionalität bezieht sich strikt auf die Strukturbildung selbst. Bei der individuellen Texterzeugung kann natürlich eine andere Reihenfolge der Anwendung der Produktionen gewählt werden. D. h., ist eine Textstruktur konzeptuell bereits entwickelt, können die TextschreiberInnen an beliebigen Stellen innerhalb dieser Struktur arbeiten und auch beliebig vorwärts und rückwärts gehen. Hier handelt es sich um eine Unterscheidung der Perspektiven, die wir für wichtig halten und

deren Nichtbeachtung zu vielen Konfusionen in der Literatur geführt hat.

Kognitive Produktionsperspektive, Einbeziehung von Wissen bei der Kohärenzbildung und gerichtete Linearität des Mediums, diese drei Aspekte bilden den Rahmen für die zu entwickelnde Textgrammatik. Damit gehen wir über textorientierte Ansätze hinaus, in denen Auswirkungen des Kontextes auf Satzebene oder bei der Verbindung von zwei oder drei Sätzen behandelt werden. In unserem Ansatz versuchen wir, Kontext selbst zu modellieren und so einen direkten Zusammenhang zu seinen Auswirkungen zu repräsentieren. Dabei wird ein allgemeiner Textrahmen eingeführt, der sowohl textexterne als auch textinterne Faktoren aufnimmt. Dies ist möglich über die Berücksichtigung einer textformalen Ebene (Linearität), die die Bedingungen für beide Typen der Strukturbildung setzt. Dieser Rahmen hat zwei Komponenten: die Textliste (T-LIST) und die Konnexionskomponente (CON). TLIST repräsentiert eine ungefüllte Trägerstruktur, die die Basiseinheiten einführt. KON liefert die Informationen, die diese Basisliste zu einer sequenziellen Textstruktur macht, indem sie Art und Skopus der Verknüpfungen der Einheiten angibt (vgl. auch Conte et al. 1989, Heydrich et al. 1989). Entsprechend unterschiedlicher Textebenen (z.B. Thema, Funktion; s. u.) können verschiedene Verknüpfungsmuster formuliert werden, die als Konnexionsformulare (KONFO) repräsentiert sind. Die Textgrammatik schließlich besteht aus Spezifikationen von CON und deren Kombinationen in den KONFOs. Sie bilden eine Art Schnittstelle der Textgrammatik zur Satzgrammatik.

In den folgenden Abschnitten skizzieren wir zunächst die Textliste als abstrakte Trägerstruktur, der zwei Textebenen (die thematische und die kommunikative) zugeordnet werden. Diese Beschreibungen bilden die Voraussetzung für die Diskussion des eigentlichen Kerns der Textgrammatik,

nämlich der Komponente CON. Wir zeigen einige Spezifikationen von CON exemplarisch anhand von Textanalysen authentischer Texte. Hierbei beziehen wir uns auf Buchankündigungen, eine Textsorte mit festgelegtem Textthema.

2 Textliste (T-LIST)

Die Einführung einer Trägerstruktur ist theoretisch und praktisch motiviert. Zum einen liegt die Idee zugrunde, daß die lineare Folge von Einheiten als Struktur durch das Medium Text vorgegeben ist (Adam 1987). Diese Struktur ist eine Struktur von Leerstellen, die bei der Textherstellung gefüllt werden. Zum anderen geht es darum, einen Raum mit bestimmten Vorgaben zu schaffen, innerhalb dessen Informationen bewegt werden können. Eine von der thematisch und funktional determinierten Textstruktur getrennte Trägerstruktur bietet im weiteren den Vorteil, daß ganz unterschiedliche Textstrukturen (je nach Theorie) zugeordnet werden können. Im übrigen stellt eine solche Trägerstruktur eine Schnittstelle zwischen Textgrammatik und Satzgrammatik dar.

Ein solches Vorgehen erleichtert die Modularisierung der Textarbeit. Sie setzt andererseits voraus, daß sehr allgemeine Elemente und Relationen für diese Struktur verwendet werden. Wir gehen von drei Elementen aus: der Basiseinheit, der Liste der Basiseinheiten und den Konnexionselementen. Die Repräsentation soll für das gesamte Modell einheitlich sein. Wir wählen hierzu Listenstrukturen mit rekursiv einsetzbaren Attribut- Wert-Paaren. Sie sind geeignet, Repräsentationen übergeordneter Textrepräsentationen mit solchen der Satzebene zu verbinden (zur Modellierung formaler hierarchischer Textstrukturen vgl. Polanyi 1988).

Die Basiseinheit ist an einer propositionalen Repräsentation orientiert. Sie bildet in den meisten semantischen, textlinguistischen und auch KI-Ansätzen die Grundlage. Unabhängig von einzelnen Spezi-

fikationen geht es hier um die kognitiv-konzeptuelle Interpretation von Propositionen (p) als Zuordnung einer verbalen Einheit, die wir als Prädikat (PRED) bezeichnen, zu einer Konfiguration von Kasusrollen, die wir zunächst als Rollen (ROL) bezeichnen. Das Prädikat bestimmt Anzahl und Art der Rollen:

[1] p: PRED (ROL-1, ... ROL-n)

Alle Leerstellen, sowohl Gesamt-p als auch PRED und ROL können spezifiziert werden. Im weiteren ist jeder Leerstelle ein CON-Element zugeordnet, das wiederum spezifiziert sein kann:

[2] (p = (COR = (...))
 ((PRED = (...). (COR = (...))
 (ROL-1 = (...). (COR = (...))
 (ROL-n = (...). (COR = (...))))

Die Textliste (TS) ordnet die einzelnen Propositionen in einer bestimmten Reihenfolge:

[3] TS (p-1, p-2, p-n)

Es gehört zu den formalen Eigenschaften der linearen Folge des Mediums Text, daß es hervorgehobene Positionen (POS) gibt: Anfang (A), Mitte (M) und Ende (E). Neben der Konnexion spielt somit ebenfalls die Positionierung von Einheiten und Relationen eine Rolle für die Textstruktur:

[3'] TS = (p-1 (POS = ...), p-n (POS = ...))

Zwischen CON (Konnexionskomponente) und POS (Positionierung) treten Abhängigkeiten auf, die mit Markiertheit bzw. Unmarkiertheit von POS zusammenhängen.

Wir gehen von kanonischen Positionierungen aus, die im jeweiligen Text ikonisch nachvollzogen werden, wobei eine unmarkierte Form möglich ist. Abweichungen werden dagegen markiert. Die Zuordnung einer kanonischen Positionierung von Einheiten am Anfang, in der Mitte oder am Ende hat zu tun mit der Übertragung hierarchisch strukturierter Schemata von Inhalten/kommunikativen Zielen in die sequentielle Textstruktur. Hierbei gibt es Beschränkungen in der Weise, daß bestimmte

Konstituenten aus der Hierarchie in der Sequenz nicht getrennt werden dürfen. Dieses Prinzip führt zu sequentiellen Gruppen, die ihrerseits durch bestimmte Konnexionen charakterisiert sind (vgl. Beispiel in 4.2).

3 Textebenen

3.1 Zum Begriff

Es ist geläufig, in pragmatisch orientierten Textgenerierungssystemen, zwischen inhaltlich-thematischen und kommunikativen Aspekten im Text zu unterscheiden. Diese Unterscheidung ist leicht linguistisch zu begründen, da inhaltliche Aspekte je nach kommunikativer Einbindung durch unterschiedliche sprachliche Mittel realisiert sein können. So lassen sich Behauptungen über ein Objekt wohl von Bewertungen unterscheiden. Ein Text, der als eine Objektbeschreibung gelten soll, ist anders aufgebaut als ein Text, der eine Argumentation über die Qualität eines Objekts bringt (Beispiele in Rothkegel1993a).

In unserem Modell plädieren wir für eine methodische Trennung der thematischen und funktionalen Beschreibungsebene. Dies vermeidet Probleme, wie sie z.B. bei Anwendung der RST (Rhetorical Structure Theory, Mann & Thompson 1987, Thompson & Mann 1987) auftreten (zu einer unterscheidenden Klärung vgl. Matthiessen & Thompson 1989). So darf hier z.B. nur jeweils eine Relation zugeordnet werden, im Inventar der Relationen sind aber inhaltlich und funktional orientierte Relationen gemischt (z.B. "elaboration" (inhaltlich) und "motivation" (funktional)).

Unser Vorschlag zielt in eine andere Richtung. Wir betrachten die Gesamtheit des Textes als Realisierung einer einheitlichen Struktur (globale Struktur). Als Bezugspunkt wählen wir konventionell bestimmte Schemata, die Wissen über Objekte oder Sachverhalte einerseits und über kommunikative Ziele andererseits zusammenfassen. Die Schematheorie stammt aus der

Wahrnehmungspsychologie (Bartlett 1932) und wurde hauptsächlich durch Schank & Abelson (1977) sowie Minsky (1977) für sprachverstehende Systeme nutzbar gemacht. Schemata sind konzeptuelle Strukturen von Wissenseinheiten, die auf Standards von Erfahrungen bzw. Erwartungen der Sprachteilnehmer beruhen. Ihnen wird eine steuernde Funktion beim Verstehen und Produzieren zusammenhängender Spracheinheiten zugeschrieben. Insofern gelten sie ebenfalls als relevant für die Kohärenzbildung von Texten (van Dijk 1980, Rosenberg 1980, de Beaugrande & Dressler 1981, de Beaugrande 1984, Fraurud & Hellmann 1985b, Tonfoni 1990). Wir verwenden das Schemakzept im folgenden, um das globale Textthema und die globale intentionale Struktur von Texten zu spezifizieren.

3.2 Textthematische Ebene

Der Begriff des Themas wird in der Linguistik so unterschiedlich gebraucht (vgl. auch Lotscher 1987), daß er fast mehr zur Konfusion beiträgt als zur Klärung und im Grunde untauglich geworden ist. Dennoch gibt es keine nützliche Alternative, außer vielleicht daß man vom Textthema spricht, wenn globale Textstrukturen gemeint sind, was am ehesten dem alltagssprachlichen Gebrauch des Begriffs nahekommt. Im weiteren ist zu berücksichtigen, daß Zusammenhänge bestehen zwischen dem so verstandenen Textthema und thematischen Strukturen im Satz (z.B. Thema-Rhema-Strukturen).

Unser Interesse bei der Behandlung des Textthemas ist vor allem durch die Frage bestimmt, auf welche Weise die beiden oben genannten Prinzipien der Kohärenzbildung, Wiederholung (Rekurrenz) und Wechsel, auf der Inhaltsebene des Textes festzumachen sind. Dies ist deswegen interessant, als hier die Konnexionskomponente zum Tragen kommt. Doch stellt sich die Frage, was jeweils verknüpft wird. D.h., es geht zunächst darum, dieses "was" näher zu spezifizieren. Wir versuchen dies über das

Textthema, das wir als konzeptuelle Struktur einer Textfrage (TQ) einführen. Dieses Vorgehen stellt eine Erweiterung des Quaestio-Ansatzes dar, wie er in Klein & von Stutterheim (1990) und von Stutterheim 1992 modelliert worden ist (zum Thema in Form von Fragen vgl. auch Hellwig 1984).

Textfragen im Quaestio-Ansatz entsprechen rhetorischen Standards, bekannt als journalistische W-Fragen. Der Text entspricht dann den Antworten auf solche Fragen. Da Texte einen Zusammenhang darstellen, beziehen sich die Fragen auf einen solchen Zusammenhang. So wäre es z.B. im Hinblick auf die Buchankündigungen sinnlos, als Textthema so etwas wie "Buch" anzunehmen. Was Produzenten und Leser interessiert, ist ein bestimmter Zusammenhang von Informationen über den Gegenstand. Auch liegt relativ stark fest, zu welcher Domäne diese Informationen gehören. So interessiert nicht die Geschichte der Buchherstellung oder die Verkaufszahlen, sondern Autor, Adressaten, Gegenstand ("Thema"), Präsentationsweise, usw. Mögliche Textfragen wären entsprechend "was bringt das Buch für die Leser?", "wie ist das Buch aufgebaut?",

"warum wurde das Buch geschrieben?", usw. Es handelt sich um Fragestandards, die in den Kommunikationssituationen relevant sind, für die der Buchankündigungstext geschrieben wurde. Man kann davon ausgehen, daß sowohl Produzenten als auch Leser wissen, welcher Art Informationen erwünscht sind. Und sie können davon ausgehen, daß es sich um gemeinsames Wissen handelt. Auf dieser Basis können wir die Textfrage im Zusammenhang mit Schemata sehen, in denen die thematisierten Konzepte geordnet für die Sprachteilnehmer verfügbar sind. Diese Konzepte sind es, deren Wiederholung bzw. Wechsel für die textthematische Struktur verantwortlich sind. Wie sieht dies nun formal aus?

Wir haben drei Repräsentationsebenen, die durch fortschreitende Selektionen mit

einander verbunden sind:

Wissensebene, Ebene der Textfrage und Ebene der propositionalen Repräsentation. Die Wissensebene liefert - in Form eines Gefüges von Leerstellen innerhalb ausgewählter Frames - die Kandidaten, die für die Ausführung der Textfrage relevant sind. Die Textfrage wiederum bietet Kriterien für die Selektion von Prädikaten einerseits und von Leerstellen aus der Menge der Kandidaten andererseits. Wir bezeichnen sie entsprechend als textuelle Prädikate (TPRED) und Text-Rollen (TROL). TPREDs sind also Konzepte von Zuständen, Ereignissen und Handlungen, die zur Domäne gehören. TROLS sind Individuumkonzepte aus der Domäne. Im Text werden beide lexikalisch gefüllt. Hierbei sind die Füller der Textrollen abhängig von den lexikalischen Füllern der Textprädikate (Beispiel unten).

Die Basiseinheit der propositionalen Struktur erhält nun durch die Einführung des Textthemas in Form von Textfragen eine zusätzliche Interpretation, die durch die Zuordnung von Leerstellen aus bestimmten Frames zur propositionalen Struktur zustandekommt:

```
[4] P (pos=...)
      (TPRED = (COI (...))
      (   (TROL-1 (COI (...)))
          (TROL-2 (COI (...)))
          ...
          (TROL-n (COI (...)))
      ))
```

Die Berücksichtigung des Textthemas macht also folgende Schritte erforderlich:

- Festlegung der Domäne (n) als Frame F mit einer Reihe von Leerstellen (L):

```
F: <L1, L2, L3, ... , Ln>
```

- Festlegung der Textfrage(n) (TQ) mit einer Reihe von Textprädikaten und der Selektion von Leerstellen, die dadurch zu Textrollen werden:

```
TQ: < TPRED1, ..., TPREDn >
     < TROL-1/L1, ..., TROL-n/Lj >
```

Das Textthema selbst kann komplex und hierarchisch strukturiert sein (Brinker

1992), z.B. in ein Hyperthema und Teilthemen. Dies bedeutet dann, daß TQ zerlegt werden kann in untergeordnete Textfragen:

```
TQ-HYPER -> TQ-TEIL1, TQ-TEIL2, ..., TQ-TEILn
```

Im Hinblick auf die Positionierung gilt die kanonische Linearisierung in der Weise, daß Elemente von TQ-HYPER die Anfangs- bzw. Endposition besetzt, während Elemente der Teilthemen die Mittelposition besetzen:

```
TQ-HYPER: < (POS = A), (POS = E) > <
TQ-TEIL: (POS = M)>
```

In Realisierungen kann HYPER sich auf die Anfangs- oder Endposition beziehen bzw. ganz fehlen (vgl. Beispiel in 4.2).

3.3 Kommunikative Ebene

Textziele (TZ), die durch den Text als Kommunikationsmittel erreicht werden sollen, werden unterschiedlich modelliert. Es lassen sich grob vier Richtungen unterscheiden, zwei sprechaktororientierte, eine rhetorische und eine an Textsorten orientierte Richtung (natürlich gibt es auch Mischungen):

- Die intentionale Struktur wird in Form von Dominanz-Subsidiaritäts-Relationen dargestellt (Grosz & Sidner 1986). Hierbei geht es um die strukturellen Beziehungen von Sprachhandlungen, ohne daß die Sprachhandlungen selbst eingeführt werden.
- Ziele und Teilziele bilden eine hierarchische Struktur des Gesamttextes, die in Form von Textilokutionen, d.h. in spezifizierten Sprachhandlungen, ausgeführt wird (Brandt & Rosengren 1991).
- Ziele werden in Form von rhetorischen Relationen (MOTIVATION, ENABLEMENT, etc) formuliert, die jeweils zwei Texteinheiten miteinander verbinden, wobei diese Verbindung wiederum Teil einer übergeordneten rhetorischen Relation sein kann. Auf diese Weise wird der Gesamttext als Hierarchie einzelner Relationen beschrieben (RST, Mann & Thompson 1987).

- Die Diskussion kommunikativer Ziele findet sich u. a. auch bei der Behandlung von Textsorten, sofern sie bestimmten Kommunikationstypen zugeordnet werden (Gülich 1986, 1988).

In unserem Modell umfaßt die kommunikative Ebene solche Aspekte der globalen Textstruktur, die die Präsentation des Textthemas betreffen. Wir gehen davon aus, daß Inhalte für die Kommunikation aufbereitet und je nach kommunikativen Zielen unterschiedlich präsentiert werden. Ein Text, der als eine Objektbeschreibung gelten soll, ist anders aufgebaut und sprachlich realisiert als ein Text, der eine Argumentation über die Qualität dieses Objekts liefert. Entsprechend unterscheiden wir **DESKRIPTION**, **ARGUMENTATION**, **NARRATION**, **REPORT**, usw. Hierbei gehen wir davon aus, daß es ebenfalls in diesem Bereich Standards gibt, die sich als Schemata darstellen lassen. So z.B. Deskriptionsschemata mit Relationen wie **TEIL-VON**, **FUNKTIONVON**, **EIGENSCHAFT-VON**, während Argumentationsschemata Relationen enthalten wie **KONKLUSION-VON**, **ARGUMENT-VON** sowie **PRO-ARGUMENTATION**, **CONTRA-ARGUMENTATION** (vgl. auch Superstrukturen in van Dijk 1980). (Zur lokalen Ebene der kommunikativen Struktur und ihrer Beschreibung durch Einzelsprachhandlungen vgl. Rothkegel 1993a, 1993b). Auf der Schemaebene sind kommunikative Strukturen hierarchisch geordnet als Textziele (TZ) mit Gesamtzielen und Teilzielen. Sie müssen im Text - ebenfalls wie das Textthema - in eine sequentielle Struktur überführt werden:

```
TZ: < ZIEL-GES1MT (POS = 1. pos = E
>
  < ((ZIEL-TEIL1, ..., ZIEL-TEILn)
  (POS = K)>>
```

Die globale Schemastruktur hat Auswirkungen auf die Positionierung der Einzelpropositionen. Ist das Textthema seinerseits komplex und strukturiert, erfolgt in der Regel eine parallele Positionierung von Teilen der kommunikativen und der thematischen Struktur.

4 Konnexionskomponente (CON)

4.1 Produktionen im Textraum

Die Konstitution von Text und Textstruktur kann auf der Grundlage der beiden strukturbildenden Strategien Wiederholung und Wechsel beschrieben werden. Wiederholung sorgt für Stabilität der Konstruktion, Wechsel für Informativität. Im Hinblick auf die Beschreibung stellt sich die Frage nach den Bezugseinheiten für Wiederholung und Wechsel.

In dynamischen Ansätzen der Textlinguistik bzw. entsprechenden Computermodellierungen gibt es hierzu einige Vorschläge. Fraurud & Hellmann (1985a) verfolgen einen wissensorientierten Ansatz, den sie auf der Basis von Experimenten erarbeiten. Versuchspersonen wird ein Satz vorgelegt, an den sie einen weiteren Satz anfügen sollen, an diesen wiederum einen neuen Satz, usw. Insgesamt soll ein kohärenter Text hergestellt werden. Die Ergebnisse zeigen deutlich, daß die Kohärenz über Wissensschemata zustandekommt, die jeweils durch bestimmte lexikalische Einheiten aktiviert werden (vgl. auch die Modellierung solcher Frameaktivierungen in Tonfoni 1990). Wiederholung bezieht sich hier auf Konzepte, die bestimmten Frames angehören. Wechsel bedeutet, daß ein anderer Frame aufgerufen wird.

Im Modell der referentiellen Bewegung (Klein & von Stutterheim 1991, von Stutterheim 1992) gehören die Bezugseinheiten zu Positionen, die im Rahmen einer ereigniszentrierten Semantik beschrieben werden. Wiederholung und Wechsel sind dabei auf referentielle Ausdrücke beschränkt. Im Fokusmechanismus von McKeown (1985) geht es ebenfalls um Beibehaltung bzw. Wechsel solcher referentieller Einheiten, die im Fokus (Aufmerksamkeitszentrum) des jeweilig produzierten Satzes stehen. Dieser Ansatz findet sich - mit der Terminologie

der Prager Schule - ebenfalls in Hajicova (1991), wobei Wiederholung des Fokus einer thematischen Wiederaufnahme und der Wechsel einer rhematischen Beziehung entspricht.

In unserem Ansatz ist Wiederholung und Wechsel auf Kategorien bezogen, die der propositionalen Trägerstruktur zugeordnet werden, so z.B. die Gesamtproposition *p*, die Textprädikate TPRED und die Textrollen TROL. Die beiden letzteren bedeuten Konzepte, die mit der Struktur des Textthemas im Zusammenhang stehen. Wiederholung erzeugt insofern eine Stabilität der textthematischen Konstruktion. Wechsel führt eher zur Instabilität, fördert aber die Informativität, z.B. durch Vergleiche, Kommentare, Nebenthemen, Einführung potentieller Handlungsebenen, usw. Damit ist ein Forschungsbereich angesprochen, der noch viele weiße Flecken hat. Hier soll nun an je einem Beispiel gezeigt werden, wie sich die beiden Strategien von Wiederholung und Wechsel in der Konnexionskomponente sprachlich manifestieren.

4.2 Beispiel für globale Produktion: Kettenbildung

Kettenbildung gründet sich auf globale Textstrukturen des Textthemas bzw. der kommunikativen Komponente. Hierbei entsteht ein Profil in der sequentiellen Struktur, das durch Wiederholung der Konnexionskategorie zustandekommt. Wir erläutern dies an folgendem Beispiel einer authentischen, aber für unsere Darstellung gekürzten Buchankündigung. Die die Konnexion markierenden Ausdrücke sind fett gedruckt: Text 1 mit fortlaufend nummerierten Texteinheiten (Ti):

T1 *In diesem Buch wird erstmals umfassend untersucht, wie diese Daten... geschützt werden können.*

T2 *Nach einer Problemanalyse werden die... Grundverfahren dargestellt.*

T3 *Anschließend wird gezeigt, wie...*

T4 *Überlegungen zur Netzbetreiberschaft ... werden hinzugefügt.*

T5 *Abschließend gibt der Autor einen Ausblick auf die Nutzung solcher Kommunikationsnetze.*

Der Text bietet ein Beispiel für Kettenbildung in bezug auf Teilthemen des im Buch behandelten Gesamtthemas, wobei das kommunikative Schema des REPORTS eine temporal markierte Sequenzierung der entsprechenden Passagen determiniert. Die Elemente der Kette (KETTE) gehören unterschiedlichen Wortklassen an (Konjunktion, Partizipform von Verben, Verbformen): KETTE: < *nach, anschließend, hinzufügen, abschließend* >

Die Gerichtetheit mit der Dimension 'LINKS-RECHTS' ist daran erkennbar, daß eine veränderte Reihenfolge der Konnektoren nicht möglich ist (mit Ausnahme des Mittelteils von t3 und t4, die eventuell umgetauscht werden können). Im folgenden leiten wir die entsprechende Repräsentation ab.

(a) Das globale Schema des REPORTS (kommunikative Struktur) liefert die Spezifikationen von CON. Die in diesem Beispiel auftretende VOR-Relation und die Bildung einer Kette aus einer Folge von VOR-Relationen entspricht einem Typ der Verkettung, der aus der kommunikativen Struktur abgeleitet ist (möglich sind auch Ketten, die aus der thematischen Struktur abgeleitet sind, vgl. Rothkegel1993a).

Das Textgesamtziel ist die Präsentation der in der Propositionsliste repräsentierten Ereignisse (PRÄSENT (PI, ..., Pn)). Das Textziel ist aufgeteilt in mehrere Teilziele (ZIEL-TEILi), die einer bestimmten temporalen Relation stehen (auch Moens & Steedman 1987). Im Beispiel ist es die Relation der Vorzeitigkeit (VOR):

(ZIEL-GESAMT	=	PRAESEIT (p1, ..., pn)
(ZIEL-TEIL1	=	PRAESEIT (p1 (t1»)
(ZIEL-TEIL2	=	PRAESEIT (p2 (t2»)
(ZIEL-TEILn	=	PRAESEIT (pn (tn»)

wobei

REL (pi, p2) = VOR (ti, ti+i)

In der Konnexionskomponente werden folgende Spezifikationen vorgenommen: die Art des Schemas (SCHEMA), Typ der das Schema bestimmenden Relation (REL) und lexikalische Kandidaten für den Füller von CON (connec). CON für p1 als erste Einheit einer Kette stellt sich wie folgt dar (die übrigen analog):

```
pi: (COI =
      ( (SCHEM1 = REPORT)
        (REL = VOR (~i,
                    ~i+1)
          (connec = {nach, danach, ...}
        ))
    )
```

(b) Das globale Schema für die textthematische Struktur liefert Angaben zur Positionierung und Gruppenbildung im Hinblick auf den Gesamttext und Angaben zu Textprädikaten bzw. Textrollen im Hinblick auf die Repräsentation der Einzelpropositionen.

Gehen wir von einer Textfrage (TQ) wie folgt aus: TQ: "Was tut der Autor?" bzw. (AKTION (AUTOR)?).

Die entsprechenden Textprädikate (TPRED) gehören zu einem Aktionskonzept, das u. a. die verwendeten Verben umfaßt:

(AKTION = {darstellen, einen Ausblick geben, hinzufügen, untersuchen, zeigen, ...}) Die Textrollen bestimmen sich nach einem Frame, der mögliche Kandidaten für Konzepte auflistet, die in der betreffenden Kommunikationssituation im Hinblick auf Bücher als Produkte üblicherweise eine Rolle spielen (vgl. Tannen 1979, Rosner 1987). Dazu gehören u. a.

```
(FRAME-BUCH:
  (LUTOR = ... )
  (LESER = ... )
  (BUCH = .. )
  (THEM1 (des Buches) = ... )
  (ZIEL = ... )
  (MOTIVLTIOTI = (PRIESEITLTIOTI
  = ... ) usw.) »
```

Die oben genannte Textfrage (AKTION (AUTOR)?) führt im Textbeispiel zu einer

Selektion von Leerstellen wie BUCH (zu beschreibendes Objekt), AUTOR (Produzent des Objekts) und THEMA (Inhalt des Objekts). Hinsichtlich der propositionalen Basisstruktur ergibt sich folgende Modifikation:

```
(Aktion =
  (
    (TROL-1 = BUCH) (TROL-2 =
    LUTOR) (TROL-3 =
    THEM1 )))
```

Im Hinblick auf die Positionierung wird das Textthema des Buchankündigungstextes nach dem Schema von HYPERTEIL (siehe oben Abschnitt 3.2) entwickelt. Dies bedeutet, daß in der Sequenzierung das HYPER-Segment die Anfangsposition einnimmt. Die Endposition wäre auch möglich, wurde aber ein anderes Kettenelement erforderlich machen. Nicht möglich ist dagegen eine Mittelposition. Dies wirkt sich wie folgt auf die Textliste (Liste der Einzelpropositionen entsprechend dem Beispiel) aus:

```
p1 (POS = 1)
  ( (T-PRED = (LKTIOI = untersuchen) )
    (TROL-i = (BUCH = dieses Buch)
    (TROL-3 = (THEM1-HYPER = vie ... ) »
p2 (POS = M)
  ( (T-PRED = (AKTION = darstellen) )
    (TROL -3 = (THEM1-TEILi =
Grundverfahren » ) p2 (POS = M)
  ( (T-PRED = (AKTION = zeigen) )
    (TROL-3 = (THEM1-TEIL2 = vie .. . » )
p3 (POS = M)
  ( (T-PRED = (AKTION = hinzufügen »
    (TROL-3 = (THEM1-TEIL3 =
Ueberlegungen » ) p4 (POS = M)
  ( (T-PRED = (AKTION = Ausblick geben) )
    (TROL-2 = (AUTOR = Autor »
    (TROL-3 = (THEM1-TEIL4 = Nutzung ... » )
```

Die Konnexionskomponente CON verbindet die kommunikative und textthematische Struktur in der Weise, daß die Kette des REPORT- Typs die Mittelposition abdeckt, die textthematisch motiviert ist. Hierbei geht es um die Wiederholung der THEMA-TEIL-Rolle, die kommunikativ markiert ist. Der Zusammenhang wird im weiteren über die Einführung aufeinander beziehbarer Konnexionsformulare organisiert (siehe unten KONFO).

4.3 Beispiel für lokale Produktion: Paarbildung

Paarbildung war als lokal und durch RECHTS-LINKS-Gerichtetheit bestimmt worden. Die Idee ist hier, daß Strukturbildung durch Anknüpfung an bereits vorhandenes Material geschieht. Dieses Phänomen findet sich - im Gegensatz zu globalen Strukturen - in der Literatur relativ stark beschrieben, wenn auch ohne die dynamische Dimension der Gerichtetheit (vgl. anaphorische Beziehungen, rhetorische Relationen, inferentielle Relationen, Relationen der traditionellen Semantik wie Kausal-, Temporal-, Final-, usw.- Relationen). Mit Gerichtetheit war die Perspektive der Produktion gemeint, nicht die unterschiedliche semantische Relation von Referent und Proform (Anapher und Katapher). So würde bei der Anapher die dynamische Anknüpfung der Proform zum Referenten, bei der Katapher die Anknüpfung des Referenten an die Proform erfolgen. Diese Dinge sollen aber hier nicht weiter ausgeführt werden. Wir beschränken uns auf den spezifischen Beitrag lokaler Anknüpfung auf der Basis der strukturbildenden Prinzipien von Wiederholung und Wechsel. Wie oben im Beispiel gezeigt, basiert Verkettung auf dem Prinzip der Wiederholung. Wir gehen davon aus, daß Paarbildung auf dem Prinzip des Wechsels beruht. Hierzu folgendes Beispiel (mit fett gedruckten Konnektoren).

Text 2:

T1 *Dieser Grundriß der Musikpädagogik setzt nicht die Reihe der musikdidaktischen Konzeptionen fort.*

T2 *Er ist vielmehr ein Versuch} das umfassendere Gebiet der Musikpädagogik zu umreißen.*

Der Wechsel bezieht sich auf den Inhalt der Gesamtpropositionen. Die Anknüpfungsstruktur ist zu sehen als eine Relation, die die Verbindung der beiden Textsegmente kohärent macht. Im vorliegenden Beispiel

handelt es sich um eine Kontrastrelation (KONTRAST), die durch das Konnektorenpaar 'nicht - vielmehr' ausgedrückt wird. Anknüpfung rückwärts (RECHTS-LINKS) bedeutet, daß eine Struktur gebildet wird, in der das zweite Element einen Wechsel in bezug auf das erste Element einführt. Dieser Wechsel beruht nun nicht auf dem Gebrauch der beteiligten Textrollen selbst (diese werden im Beispiel wiederholt), sondern auf der Einordnung der Textrollen in unterschiedliche Modi der zugrundeliegenden Textfrage. Die Textfrage "was tut der Autor?" bzw. stellvertretend "was tut das Buch?" wird ergänzt durch die Klasse potentieller Aktionen. Die Form (AKTION (BUCH)?) ist entsprechend zu modifizieren in (POT-AKTION (BUCH)?). Zwischen t1 und t2 findet also ein Wechsel im Hinblick auf die Modi der Textfrage statt. In der Konnexionskomponente stellt sich dies wie folgt dar:

(p1 (CON =
(SCBEK. = [+]POT-.ITIOI)
(REL = KONTRAST)
(connec = { (nicht/vielmehr),
[+]POT = nicht } »)
(p2 (CON =
(SCHEMA. = [-]POT-AKTION)
(REL = KONTRAST)
(connec = { (nicht/vielmehr),
[-]POT = vielmehr } »)

Das Paar könnte auch in umgekehrter Reihenfolge realisiert werden. In diesem Fall müßten ([+]POT-AKTION) und ([-]POTAKTION) vertauscht werden, was Konsequenzen hinsichtlich der Selektion der einzelnen Konnektoren hat. V gl. dazu:

t1 *Dieser Grundriß ist ein Versuch} das umfassendere Gebiet der Musikpädagogik zu umreißen.*

t2 *Er setzt nicht (keineswegs/ auf keinen Fall/ ...) die Reihe der musikdidaktischen Konzeptionen fort.*

Erscheint ([-]POT-AKTION) in einem solchen Paar an erster Stelle, wird auf einen Konnektor für den ersten Teil des Paares verzichtet. Dies bedeutet, daß in der Textgrammatik Paare zweifach zu berücksichtigen sind.

5 Konnexionsformulare (KONFOs)

Es ist davon auszugehen, daß es eine Reihe unterschiedlicher Typen der Ketten- und Paarbildung gibt. Im weiteren, daß sie beschreibbar sind. Hierbei spielen zwei Arten von Informationen eine Rolle: Eigenschaften der kommunikativen, textthematischen und formalen Textstruktur werden in Form von Mustern beschrieben, die sich auf Spezifikationen von CON und POS beziehen. Die davon abhängigen lexikalischen Selektionen werden über lexikalische Inventare (INV) organisiert, die alternative Angebote zur sprachlichen Realisierung liefern. Muster und Inventare sind in den hier eingeführten Konnexionsformularen (KONFO) aufeinander bezogen. Dies bedeutet, daß vertikal und horizontal bestimmte Textstrukturen in einem Schichtenmodell einander zugeordnet werden. Vertikal sind Zuordnungen über verschiedene Beschreibungsebenen hinweg, z.B. über die textthematische, kommunikative und lexikalische Ebene. Horizontal sind Zuordnungen, die die Linearität des Textes abbilden, z.B. Konnexionen, die sowohl explizit (Beispiel in diesem Beitrag) als auch implizit sein können.

KONFOs liefern die Grundstruktur für Textgrammatik und Textrepräsentation. Die Idee ist, daß für jeden Typ globaler und lokaler Strukturbildung ein KONFO etabliert wird, so daß eine bestimmte Kombination ausgewählter KONFOs den jeweiligen Text repräsentiert. Ein KONFO ist organisiert als Liste mit Attribut-Wert-Paaren, wobei die Attribute zwei Funktionen kennzeichnen: die Identifikation (ID) einer Einheit aus der propositionalen Struktur und die Spezifikationen von POS und CON. Sie werden je einer Proposition zugeordnet. Globale KONFOs umfassen mehr als zwei Einheiten, lokale KONFOs sind auf zwei Einheiten beschränkt. Die allgemeine Form ist wie folgt:

(KONFO_i =

```

1( (ID = (p = (....)
    ((TPRED = ... )
      (TROL1 =      )
      ...
      (TROLn =      ))
    ))
  (POS = ( ...
  (CON = (TYP      = <GLOBAL, LOKAL> )
        (SCHEMA   = ...           )
        (REL       = ...           )
        (CONNEX   = INVi         ) ) )
  ...
n( (ID = (p = (....)
    ((TPRED = ... )
      (TROL1 =      )
      ...
      (TROLn =      ))
    ))
  (POS = ( ...
  (CON = (TYP      = <GLOBAL, LOKAL> )
        (SCHEMA   = ...           )
        (REL       = ...           )
        (CONNEX   = INVj         ) ) ) )

```

Die weitere Perspektive in der Forschungsarbeit ist die empirische Untersuchung von unterschiedlichen Konnexionsmustern nach dem KONFO-Schema sowie von darauf basierenden Kombinationsregeln für die Auswahl zusammenpassender KONFOs in authentischen Textrealisierungen. Zusammenpassende KONFOs werden auf die Trägerstruktur projiziert, so daß eine Textrepräsentation in verschiedenen Schichtungen entsteht. Diese geschichtete, aber lineare Textrepräsentation wiederum kann als Ausgangspunkt für grafische Visualisierungen verwendet werden, die ihrerseits als Eingabemedium für ein computerunterstütztes System zum Textschreiben fungieren können.

6 Zusammenfassung

Der skizzierte Ansatz weist in eine Richtung, in der noch etliche Forschungsarbeit zu leisten ist. Andererseits gibt es eine Reihe von Vorarbeiten, u.a. ein implementiertes Analysemodell, das ebenfalls kommunikative und textthematische Aspekte einbezieht (vgl. TEXAN in Rothkegel 1986). Die Vorgehensweise ist komplementär zu Textgenerierungsmodellen, in denen auf vorhandenen Satzgrammatiken aufgebaut wird. Die Idee ist vielmehr, eine eigenständige Textgrammatik zu entwickeln, die über eine Schnittstelle mit vorhandenen Satzgrammatiken verbindbar ist.

Die Textgrammatik selbst integriert textexterne Faktoren, die zur Textstrukturbildung beitragen und textinterne Faktoren, die aus Eigenschaften des Trägermediums abgeleitet werden. Die wesentlichen Punkte lassen sich wie folgt zusammenfassen:

=> Das Konzept der Textgrammatik ist am Sprachgebrauch orientiert. Dies entspricht der Annahme, daß ein Text das Resultat kommunikativer Handlungen ist. Theoretische Grundlagen dazu finden sich in der textpragmatischen Forschung. Diese Sichtweise wird ergänzt durch textformale Eigenschaften des Trägermediums.

=> Die Computermodellierung liefert Repräsentationen der Textstrukturbildung, die textexterne und textinterne Faktoren mit sprachlichen Realisierungsmöglichkeiten verbindet. KONFOs sind die Repräsentationseinheiten, die einzelne spezifische Konnexionsmuster erfassen und so die elementaren Einheiten einer Textgrammatik darstellen.

=> Aufgrund einer separaten linearen Trägerstruktur ist das Modell offen für die Integration anderer globaler und lokaler Relationierungen, sofern sie sich in die propositionale Basisstruktur einordnen lassen. Im weiteren gestattet es die Formulierung einer Schnittstelle für die Anbindung von Satzgrammatiken.

Bibliographie

- Adam, J.-H. (1987)**, Textualité et sequentialité: l'exemple de la description. In: Langue française 74 : 51 ff.
- Bartlett, F.C. (1932)**, Remembering. Cambridge University Press, Cambridge.
- de Beaugrande, Robert, A. (1984), Text production. Toward a science of composition. Ablex, Norwood.
- de Beaugrande, R.-A. & Dressler W. (1981)**, Einführung in die Textlinguistik. Niemeyer, Tübingen.
- Brandt, Margareta & Rosengren, Inger (1991)**, Handlungsstruktur und Informationsstruktur - zwei Seiten einer Münze. In: Sprache und Pragmatik. Arbeitsberichte 24: 120-139, Lund.

- Brinker, Klaus (1992)**, Linguistische Textanalyse. Eine Einführung in Grundbegriffe und Methoden. E. Schmidt-Verlag, Berlin.
- Conte, Maria-Elisabeth & Petöfi, Janos S. & Sozer, Emel (eds) (1989)**, Text and discourse connectedness. J. Benjamins, Amsterdam.
- Dale, Robert (1990)**, Generating recipes: An overview of EPICURE. In: Dale et al. (eds), 229-255.
- Dale, Robert & Mellish, Chris & Zock, Michael, (eds), (1990)**, Current research in natural language generation. Academic Press, London.
- Dijk, Teun A. van (1980)**, Textwissenschaft. Eine interdisziplinäre Einführung. dtv, München.
- van Dijk, T.A. & Kintsch, W. (1983)**, Strategies of discourse comprehension and production. Academic Press, New York.
- Fraurud, Kari & Hellman, Christina (1985a)**, What's next in the text. A method for investigating discourse processing, Papers from the Institute of linguistics, 54, University of Stockholm.
- Fraurud, Kari & Hellmann, Christina (1985b)**, In quest of coherence. Approaches to discourse comprehension. Papers from the Institute of Linguistics, 53, University of Stockholm.
- Fritz, Gerd (1982)**, Kohärenz. Grundfragen der linguistischen Kommunikationsanalyse. Narr, Tübingen.
- Grishman, Ralph (1990)**, Domain modeling for language analysis. In: Schmitz et al. (eds), 41-57.
- Grosz, B. J. & Sidner, C. L. (1986)**, Attention, intentions, and the structure of discourse. In: Computational Linguistics 12: 175-204.
- Gülich, Elisabeth (1986)**, Textsorten in der Kommunikationspraxis. In: Kallmeyer (ed), 15-46.
- Gülich, Elisabeth (1988)**, Handlungsschema und Formierungsstruktur. In: Sprache und Pragmatik, Arbeitsberichte 8: 43-66, Lund.
- Hajicova, Eva (1991)**, Topic-focus articulation and coreference in models of discourse production. Journal of Pragmatics 16(2), 157-166.
- Halliday, M.A.K. & Hasan, R. (1976)**, Cohesion in English. Longman, London.
- Hellwig, Peter (1984)**, Grundzüge einer Theorie des Textzusammenhangs. In: Rothkegel, A. 8<. Sandig, B., (eds), Text - Textsorten Semantik. Linguistische Modelle und maschinelle Verfahren, 51-79. Buske, Hamburg

- Heydrich, Wolfgang & Neubauer, Fritz & Petöfi, Janos S. & Sözer, Emel (eds) (1989)**, Connexity and coherence: analysis of text and discourse. de Gruyter, Berlin usw.
- Hovy, Eduard H. (1988)**, Generating natural language under pragmatic constraints. Lawrence Erlbaum Ass., Hillsdale.
- Kallmeyer, Werner (ed) (1986)**, Kommunikationstypologie. Handlungsmuster, Textsorten, Situationstypen. Schwann, Düsseldorf.
- Kempen, Gerard (ed) (1987)**, Natural language generation. New results in artificial intelligence, psychology and linguistics. Martinus Nijhoff Publishers, Dordrecht.
- Klein, Wolfgang & Stutterheim, Christiane von (1991)**, Text structure and referential movement. In: Sprache und Pragmatik. Arbeitsberichte, 22. Lund.
- Krings, Hans P. & Antos, Gerd (eds) (1992)**, Textproduktion. Neue Wege der Forschung. Wissenschaftlicher Verlag, Trier.
- Lotscher, Andreas (1987)**, Text und Thema. Studien zur thematischen Konsistenz von Texten. Niemeyer, Tübingen.
- Mann, William & Thompson, Sandra A. (1987)**, Rhetorical structure theory: description and construction of text structures. In: Kempen (ed), 85-95.
- McKeown, Kathleen R. (1985)**, Text generation. Using discourse strategies and focus constraints to generate natural language text. Cambridge University Press, Cambridge.
- Metzing, Dieter (ed) (1980)**, Frame conceptions and text understanding. Berlin.
- Minsky, Marvin (1977)**, Frame system theory. In: Johnson-Laird, P.N. & Wason, P.C. (eds), Thinking: Readings in Cognitive Science, 355-377. Cambridge University Press, Cambridge.
- Moens, M. & Steedman, M. (1987)**, Temporal ontology in natural language. In: Computational Linguistics 14(2):1-7.
- Noel, Dirk (1990)**, Coherence relations and mental models. Linguistica Antverpiensia, Nr. XXIV, Universität Antwerpen.
- Paris, Cecile L. & Swartout, William R. & Mann, William C. (eds) (1991)**, Natural language generation in artificial intelligence and computational linguistics. Kluwer, Dordrecht.
- Polanyi, L. (1988)**, A formal model of the structure of discourse. In: Journal of pragmatics 12,5/6: 601 ff.
- Rösner, Dietmar (1987)**, The automated news agency: SEMTEX - a text generator for German. In: Kempen (ed), 133-148.
- Rosenberg, St. J. (1980)**, Frame-based text processing. In: Metzing (ed) 1980: 96-119.
- Rothkegel, Annelly (1986)**, TEXAN. Eine semantisch-pragmatische Textanalyse. Arbeitsbericht des Projekts A3, TEMA, Nr. 1 (Textstrukturen und maschinelle Übersetzung), SFB 100, Universität Saarbrücken.
- Rothkegel, Annelly (1989)**, Textualisierung von Wissen. Einige Forschungsfragen zum Umgang mit Wissen im Rahmen computerorientierter Textproduktion. LDV-Forum, Bd. 6, Nr 1: 3-13.
- Rothkegel, Annelly (1990)**, Knowledge representation and text processing. In: Schmitz et al. (eds), Linguistic approaches to artificial intelligence. 331-352. Lang, Frankfurt etc.
- Rothkegel, Annelly (1992)**, Textstruktur und Computermodelle der Textgenerierung. In: Krings & Antos (eds), 339-353.
- Rothkegel, Annelly (1993a)**, Textualisieren. Theorie und Computermodell der Textproduktion. Lang, Frankfurt.
- Rothkegel, Annelly (1993b)**, Text knowledge and object knowledge. Pinter, London.
- Schank, Roger C. & Abelson, Robert P. (1977)**, Scripts, plans, goals and understanding. Erlbaum, Hillsdale.
- Schmitz, Ulrich & Schütz, Rüdiger & Kunz, Andreas (eds) (1990)**, Linguistic approach es to artificial intelligence. Lang, Frankfurt usw.
- Stutterheim, Christiane von (1992)**, Quaestio und Textstruktur. In: Krings, H.P., Antos, G. (eds), 159-171.
- Tannen, Deborah (1979)**, What's in a frame? Surface evidence for underlying expectations. In: Freedle, Roy O. (ed), New directions in discourse processing, 137-181. Ablex, Norwood.
- Tonfoni, Graziella (1990)**, Text representation systems. Eurographica, Helsinki.
- Zock, Michael & Sabah, Gerard (eds) (1988)**, Advances in natural language generation. An interdisciplinary perspective. Vol. I, 11, Pinter Publishers, London.

COMPUTERLINGUISTIK UND INFORMATIK SELBSTVERSTÄNDNIS EINER DISZIPLIN*

Prof. Dr. István Bátori

Universität Koblenz-Landau, Abteilung Koblenz

Institut für Computerlinguistik

CL und Informatik sind untereinander auf mehreren Ebenen verbunden:

1. Für beide Disziplinen sind Sprachen, sprachliche Formulierungen, Sprachverstehen, sprachliche Repräsentationen, die Benutzung der Sprachen in Kommunikation, der sprachliche Verarbeitungsprozess von zentraler Bedeutung. Sprachen und Computer, genauer: die von-Neumann-Typ-Rechner sind symbolverarbeitende Einrichtungen: Wesentlich ist dabei nicht die physikalische Fluktuation der Zustände, sondern die vermittelten Inhalte. An dieser wesenhaften Ähnlichkeit ändert auch der Umstand nichts, daß die natürlichen Sprachen einen anderen (viel höheren) Komplexitätsgrad besitzen, als die (künstlichen) Sprachen, die im Umgang mit den Rechnern benutzt werden.

2. Informatik und CL sind auch methodisch eng verbunden. In der CL werden die Konzepte, Beschreibungstechniken und die Software-Instrumente der Informatik gezielt aufgegriffen

* Eröffnungspanel auf der Jahrestagung der GLDV, in Kiel, 3.3.1993, nach der Diskussion überarbeitete Fassung

und benutzt. Mit Hilfe der Informatik-Instrumente strebt die CL eine hohe Beschreibungspräzision an. Da die Komplexität der natürlichen Sprachen, die der Programmiersprachen übersteigt, reichen die entlehnten formalen Mittel für die adäquate Erfassung der natürlichen Sprachen nicht aus. Daher müssen die in der CL entwickelten Systeme über die Informatik-Vorbilder hinausgehen. Beispiel: Compiler-Compiler YACC und die split stack parser von Tomita oder die Entwicklung von relationalen Datenbanken zu Systemen mit nichtmonotonen Vererbungstechniken im lexikalischen Bereich.

3. Informatik und CL sind zweifelsohne auch durch die professionelle Computerbenutzung verbunden. Allerdings ist diese Verbindung nicht einmalig. Zeitgenössische Wissenschaft arbeitet generell mit professioneller Rechnerunterstützung. Computerunterstützung allein könnte die einmalig enge Verbindung zwischen Informatik und CL nicht erklären. Die instrumentale Nutzung der Computer für die CL als Wissenschaft ist nicht konstitutiv.

Die Beziehung zwischen CL und Informatik ist nicht symmetrisch: von einem Infor-

matiker wird nicht erwartet, daß er über linguistische oder CL-Kenntnisse verfügt; ein Computerlinguist muß dahingegen die Verarbeitungstechniken der Informatik für Sprachen kennen. Verarbeitungstechniken für Sprachen sind allerdings nicht - wie oben bereits angedeutet - eine isolierbare Teilmenge der Informatik, sondern Informatik schlechthin. Von einem (idealen) Computerlinguisten wird eine Doppelkompetenz (Informatik und Linguistik) erwartet. Es wäre unsinnig, dieses oder jenes Teilgebiet der Informatik für die CL als irrelevant oder nutzlos zu erklären. Je umfangreicher das Informatikwissen ist, umso größer ist die Katalisationsfläche für innovative Entwicklungen.

An dieser Stelle soll hervorgehoben werden, daß die CL eine forschungsorientierte Disziplin ist:

1. CL ist noch im Aufbau, wir stehen hier eher am Anfang einer Entwicklung und nicht mitten in einem saturierten Software-Feld. Die heutigen CL-Systeme sind Prototypen und funktionieren unbefriedigend.
2. Die CL-Absolventen finden typischerweise Stellen in Forschungs- und Entwicklungsprojekten.
3. Auch die CL-Ausbildung ist forschungsorientiert.

Meine Beobachtungen stützen sich auf die Daten unserer Absolventen in Koblenz. Im Prinzip kann ein Universtätsabsolvent nach dem Studium drei Wege einschlagen: 1. den Weg der Praxis, 2. den der Forschung und 3. den der Lehre. Die CL bietet z. Zt. zwei aus den obigen drei Möglichkeiten:

1. eine Stelle in der Forschung (genauer R&D) oder
2. eine Nachwuchsstelle in der Lehre.

Im Gegensatz zu Tiermedizin, BWL oder Jura, existiert die Praxis als dritte Karrieremöglichkeit nicht. Unsere Absolventen, insoweit sie computerlinguistisch "im Beruf" bleiben, haben typischerweise Stellen in R&D-Abteilungen, auch wenn sie in der Wirtschaft tätig sind.

Wenn unsere Koblenzer Informatikstudenten mit Anwendungsfach CL fertig sind, gibt es für sie drei Möglichkeiten:

1. Sie nehmen eine (lukrative) Stelle in der Wirtschaft oder in der Verwaltung an und hängen CL an den Nagel.
2. Sie bekommen eine Stelle in einem CL-Projekt und machen Forschung full time.
3. Sie bleiben in der Lehre (und machen Forschung nebenbei).

Schließlich möchte ich darauf hinweisen, daß Forschung in der CL und Forschung in der Linguistik stark konvergieren. Es gibt eine wachsende Anzahl von Modellen, die sowohl in der CL als auch in der Linguistik benutzt werden. LFG, GPSG und HPSG sind die prominentesten. Daher ist es nicht zweckmäßig, eine scharfe Trennlinie zwischen echter linguistischer Forschung und CL-orientierter Grundlagenforschung zu ziehen.

Computerlinguistik ist die moderne Form der linguistischen Forschung. Die Konvergenz der beiden Felder (CL und Linguistik) ist positiv zu bewerten und es ist auch angebracht, dieser Sachlage Rechnung zu tragen.

TAGUNGSBERICHT

JAHRESTAGUNG 1993 DER GLDV IN KIEL

"SPRACHTECHNOLOGIE: METHODEN, WERKZEUGE, PERSPEKTIVEN"

3.3. - 5.3.1993

Die Tagung wurde begonnen mit einem Einleitungspanel unter der Leitung von W. Lenders (Bonn), das in diesem Band dokumentiert wird. Das Thema der Tagung wurde in 4 Sektionen (Quantitative Linguistik: R. Köhler, Fuzzy Linguistik: B. Rieger, Maschinelle Übersetzung: J. Haller, Maschinelle Korpora: W. Lenders) sowie in einigen sektionsfreien Vorträgen abgehandelt.

Sektion Quantitative Linguistik (Leitung: Reinhard Köhler): R. Köhler und M. Galle (Trier) verdeutlichten in ihrem Beitrag "Dynamische Eigenschaften von Textmaßen" die Problematik des Ausdrucks von Texteigenschaften mit traditionellen statistischen Methoden. Weder die Fixierung solcher Eigenschaften in einer einzigen Zahl noch Darstellungen von Häufigkeits- oder Rangverteilungen ermöglichten es bislang, die textuelle Dynamik angemessen auszudrücken; die exemplarisch vorgeführten Verlaufskurven zur Handlungsintensität mit ihrem flachen bzw. auffällig gleichmäßigen Verlauf verdeutlichten den Mangel an Ausdrucksfähigkeit dieser Methodik eindrucksvoll. Köhler / Galle schlagen daher auf der Grundlage eines Modells von G. Altmann eine differenzierte Vorgehensweise vor, wobei sie in Abhängigkeit von der Position des jeweils ausgewählten Textausschnitts Korrekturglieder in die statistische Berechnungsformel des Verhältnisses von Adjektiven zu Verben einführen.

Ergebnis dieser Modifikation sind wesentlich ausdrucksfähigere Aktionsquotientenkurven, anders gesagt: erst nach dieser Veränderung sind die Kurven überhaupt aussagekräftig.

- P. Schmidt (Trier) stellte in seiner Untersuchung zu "Metrisierungsmöglichkeiten in der Morphosyntax" Möglichkeiten einer vergleichenden Analyse des relativen Gewichts und der Interaktion" verschiedener "Faktoren bei der (menschlichen oder maschinellen) syntaktischen Analyse und ihrer typologischen Beziehungen" vor. Der Vortrag war in seiner Dokumentation von Breite und Tiefe des Wissens beeindruckend, dem Nicht-Spezialisten (so z.B. dem Berichterstatter) wurde damit der Nachvollzug des Vorgetragenen allerdings nur schwer möglich. Der Referent legte Wert auf die Feststellung, daß "die . . . vorgestellten Maße offensichtlich ohne Schwierigkeiten für die Messung analoger Eigenschaften in anderen sprachlichen Subsystemen, wie der Flexionsmorphologie oder dem Lexikon, adaptiert werden" können.

- Die weiteren in dieser Sektion angekündigten Vorträge fielen wegen Erkrankungen der Referenten aus.

Am Nachmittag des ersten Tages fanden dann in parallelen Sitzungen "freie Vorträge" statt. So berichtete U. Seewald (Hannover) über "Objektorientierte Programmierung als Werkzeug der Linguistischen

Datenverarbeitung - Überlegungen anhand von Lexikon- und Datenbankdesign". Die Referentin benutzt die Programmiersprache Objectworks\ Smalltalk von Parc Place Systems. Die darin zur Verfügung stehende Einheit eines "Objekts" kann sich aus dem jeweils aktuellen Zustand, den Eigenschaften und Operationen des zu beschreibenden Objekts zusammensetzen. Auf die Werte der Variablen dieses Objekts kann durch Nachrichten an das Objekt zugegriffen werden. Am Beispiel eines Morphemlexikons wurde gezeigt, wie eine hierarchische Ordnung der zu beschreibenden Einheit geschaffen werden kann, wobei die übergeordneten Einheiten als Klassen jeweils ihre Eigenschaften an ihre Unterklassen vererben, soweit sie dort nicht durch aktuelle Werte überschrieben werden; so lassen sich ohne Schwierigkeiten auch Ausnahmen von Default-Werten realisieren. An sich unterstützt die gewählte Programmiersprache keine Mehrfachklassifikation (polyhierarchische Vererbung), wie sie jedoch gerade in der Lexikographie oft benötigt wird (z.B. *Leber: ORGAN und SPEISE*). Über die Definition von Abhängigkeitsrelationen hat die Referentin jedoch eine Struktur aufgebaut, die einem polyhierarchisch geordneten System nahekommt. Sie demonstrierte die Möglichkeiten dieser Programmierung an einer Datenbank zu Morphemen des Französischen. Die Abhängigkeits(baum)struktur konnte sie durch hintereinander angeordnete Teilfenster der Lexikonoberfläche sichtbar machen.

- Hans-Dieter Lutz zeigte in seinem Vortrag über "Software-Ergonomie für Sprachsoftware" zunächst die Gründe für eine grundlegende Veränderung des Verhältnisses von Software-Entwicklung und Software-Ergonomie auf. Resultat dieser Entwicklung ist, daß heute Software generell "von außen nach innen" zu entwickeln ist. Im Zentrum des Interesses stehen die Funktionalitäts-, die Korrektheits- und die Schnittstellen-Ergonomie. Lutz differen-

zierte anschließend den Begriff der Sprachsoftware und versuchte eine Beschreibung der Benutzerklassen dieser Produkte. Auf der Grundlage von EG-Richtlinien zur ergonomischen Gestaltung von Bildschirmarbeitsplätzen betrachtete Lutz abschließend die Konsequenzen für 3 Typen von Sprachsoftware: Textbe- und -verarbeitung, Software auf natürlichsprachlicher Kommunikationsbasis und Sprachlernprogramme. Dabei zeigte sich, daß die Computerlinguistik (CL) auch zukünftig noch wichtige Aufgaben zu erfüllen hat, daß sie aber auch das Spektrum ihres Grundlagenforschungsansatzes um Fragebereiche der Sprachverwendung zu erweitern haben wird.

- Jutta Marx (Regensburg) referierte über den "Einsatz natürlichsprachlicher Komponenten in einer multimodalen Benutzerschnittstelle für Werkstoffdatenbanken" und berichtete von Erfahrungen mit unterschiedlichen Datenbankabfragekonzepten allgemein und Kombinationen dieser Grundtypen bei Tests anlässlich der Entwicklung einer Werkstoff-Datenbank (Projekt WING). Ein Schwerpunkt lag dabei auf der Evaluierung der Einsatzmöglichkeit natürlicher Sprache, die in der aktuellen Realisierung des Projekts bei der Zustandsanzeige, im Korrekturmodus und im Hilfesystem verwendet wird. Weitere Tests, die auch ein aktualisiertes Retrieval-Modul einschließen sollen, sind für Anfang 1994 vorgesehen.

- I. Batori (Koblenz) plädierte in einem gemeinsam mit M. Volk (ebd.) erarbeiteten Beitrag dafür, "Workbenches auch für die Forscher in der theoretischen Linguistik anzubieten". Batori führte die verschiedenen Phasen des Linguistic Engineering vor (Problemanalyse, Problemdefinition, Systemspezifikation, Implementierung) und beschrieb die Beziehungen zwischen CL und Sprachforschung. Kernstück des Referats war jedoch eine Kritik an der Verfahrensweise linguistischer Forschung, die - jeweils veranlaßt durch die Erkenntnis der Unzulänglichkeit einer Grammatik

- ständig neue Grammatikmodelle entwirft, welche zwar die zuvor entdeckten Probleme zu lösen versprechen, ohne daß jedoch die Deckungsgleichheit mit der Reichweite des vorhergehenden Modells überprüft und ohne daß das im früheren System enthaltene korrekte Wissen weiterverwendet wird. Es wurde gezeigt, daß die CL hingegen "die Vererbung des linguistischen Wissens im Falle eines Modellwechsels zu gewährleisten" vermag.

- Marc Domenig als Vertreter einer Arbeitsgruppe (Basel) beschrieb "Werkzeuge zur Akquisition und Verwaltung von morphologischem und phrasealem Wissen" einschließlich Navigationsmöglichkeiten im Gesamtwissensbestand eines in Basel entwickelten Datenbanksystems. Dieses ist auf der Basis eines Client/Server-Modells konstruiert und erlaubt dem Benutzer Eingriffe in Detailbereiche, ohne daß er sich um die Position oder Editionsformate der benutzten Subdateien zu kümmern braucht. Die verwendeten (Fenster-)Editoren "kennen" die jeweiligen Spezifikationen der Dateien, und ggf. vorgenommene Eingabeänderungen werden vom Compiler automatisch entdeckt und verarbeitet. Dieses hochspezifizierte Werkzeug eignet sich entsprechend der Vorführung während der Tagung vorzüglich für moderne Lexikonarbeiten.

- H. Elsen und J. Hartmann (Bonn) berichteten in einem kurzfristig noch eingefügten Vortrag über "Satzbandanalyse und Aufbereitung des Definitionswortschatzes eines deutschen Wörterbuchs" über die Vorarbeiten zur Extraktion von Informationen aus einem maschinell lesbaren Teil des 'DUDEN - Deutsches Universal Wörterbuch' sowie die erforderlichen Arbeiten zur Überführung dieser Daten in ein SGML-Format. Sie bedienen sich bei dieser Arbeit teilweise der auf dem Satzband vorgefundenen internen Markierungen und entwickelten eigene Programme, die sie zusammen mit kommerzieller Software (FLEX, BISON und den in Bonn bereits vorhandenen lexikalischen Ressour-

cen (Bonner Wortdatenbank, Wortanalytisches Wörterbuch) zur Segmentierung lexikalischer Einheiten auf dem Magnetband einsetzen konnten.

- Im Zusammenhang des gleichen Bonner Projekts führte N. Weber die Beschreibung der (computer-)linguistischen Probleme einer "Computergestützte(n) Analyse von Definitionstexten in einem deutschen Wörterbuch" weiter aus. Ziel dieser Arbeiten ist eine formalisierte Repräsentation von Lexembedeutungen, die aus den maschinenlesbaren Lexikoneinträgen des DUDEN nach umfangreichen Vorarbeiten gewonnen werden konnten, damit diese dann anderen sprachverarbeitenden Systemen zur Weiterverarbeitung bereitgestellt werden können.

Sektion Fuzzy Linguistik (Leitung: B. Rieger): Arbeitsteilig berichteten B. Rieger ("LLAMA - ein Pilotsystem zum referentiellen Sprachlernen mit unscharfem Bedeutungserwerb"), B. Badry ("Sprachliche Unschärfe und ihre experimentelle Modellierung in LLAMA") und M. Reichert ("Cluster-Strukturen der Zwischenrepräsentationen in LLAMA") - alle aus Trier - über Teilaspekte des gleichen großen Projekts "Language Learning And Meaning Acquisition". Am Anfang standen grundsätzliche Überlegungen zu Grundlagen und Problemen des Erwerbs und der Verarbeitung von Wissen in kognitiven Systemen, zu denen auch der Mensch gehört. Der Forschungsansatz des Projekts beruht variierend auf Erkenntnissen des International Computer Science Institute, Berkeley, in dessen MLA-Projekt ("Miniature Language Acquisition"). Das in Trier entwickelte System besitzt weder ein Regelwissen noch eine Kenntnis der Bedeutungen der zu verarbeitenden Zeicheneinheiten, vermag aber nach Verarbeitung der eingegebenen Texte und kurzer Lernphase, sich in einer 'Situation' zu orientieren und z.B. die Lage bestimmter Objekte mit Hilfe vorgegebener Richtungsparameter zu bestimmen.

- M. Galle (ebenfalls Trier) beschrieb die

Erfahrungen, die man im Rahmen eines studentischen Studienprojekts mit der Aufbereitung von sehr großen Zeitungskorpora (dpa, taz) gemacht hat. Dabei ging es nicht nur um die Eliminierung von störenden Satzsteuerzeichen oder die einheitliche Umsetzung von Umlauten und anderen Sonderzeichen; unbedingt Ziel war es vor allem, die im Originalkorpus enthaltenen Informationen auch in der aufbereiteten Korpusversion möglichst vollständig beizubehalten. Zu den zu bewältigenden Problemen zählte u. a. der Aufbau von Routinen, welche die enormen Datenbestände auch Forschern zugänglich machen, die von der textinternen und speichertechnischen Struktur der Daten keine Kenntnis haben.

Sektion Maschinelle Übersetzung

(Leitung: J. Haller): Aufgrund von Absagen der METAL- und LOGOS- Referenten sowie einer weiteren kurzfristigen Absage reduzierte sich das Programm der Sektion beträchtlich. Anstelle der LOGOS-Referentin lieferte R. Nübel (Saarbrücken) einen Erfahrungsbericht über "Möglichkeiten zur Evaluierung eines kommerziellen MÜ-Systems". In der Vortragsfassung berichtete die Referentin zunächst über die frustrierenden Einsichten in die Schwierigkeiten einer Verständigung zwischen Systementwickler (konkret: LOGOS) - Bewahrung von Entwicklungs- /Firmengeheimnissen - und wissenschaftlichen Evaluieren. Im zum Druck vorliegenden Beitrag beschränkt sich die Autorin auf grundsätzliche Fragestellungen im Rahmen von Evaluierungsaufgaben bei kommerziellen MÜ -Systemen, die ohne Einblick in die Entwicklungsinterna des Systems erfolgen müssen (black box-Evaluierung). Dabei werden ausführlicher drei Typen von Evaluierungsstrategien behandelt: Testen der linguistic coverage, der text type coverage und der textsortenspezifischen (Übersetzungs-)Ergebnis-Adäquatheit. Abschließend machte die Autorin deutlich, daß eine sinnvoll-produktive Zusammenarbeit zwischen Systementwickler

und externen (Computer-)Linguisten den Zugang zu Systeminterna zur unbedingten Voraussetzung hat.

- J. Haller (Saarbrücken) referierte anschließend über Verarbeitungsstrategien und linguistisches Konzept von CAT2 (" Vom Forschungssystem zum präindustriellen Prototyp"). Im Unterschied zur EUROTRA-Entwicklung wird bei CAT2 ein SICSTUS-Prolog-Compiler anstelle eines YAP-Compilers verwendet. Die auf drei Ebenen repräsentierte Grammatik ist stark lexikonorientiert. Ein besonderes Charakteristikum ist auf der Interface-Ebene ein großer sprachunabhängiger Teil, der eine relativ problemlose Integration weiterer Zielsprachen zu ermöglichen verspricht. Auf der Software-Seite werden eine auf Sun-View aufbauende Dialogversion, eine (für Testzwecke geeignete) Batchversion sowie einen für die linguistische Evaluierung wichtige Debug-Version angeboten. Die geplante Auslagerung des Lexikons in eine externe Datenbank soll den weiteren Ausbau des Lexikons erleichtern. Der Referent kündigte an, daß das System von einem industriellen Pilotanwender im Medizinbereich sowie im Rahmen von EUROLANG bei der anvisierten Entwicklung eines modernen Übersetzerarbeitsplatzes eingesetzt werden wird. Die Tagungsteilnehmer hatten mehrfach Gelegenheit, sich das System ausführlich vorführen zu lassen.

- K. Schubert (Flensburg) sprach über die Beziehung "Zwischen Benutzerschulung und Wissenschaft. Sprachtechnologie in der Übersetzerausbildung. "Er beschrieb dabei zunächst den Wandel in der berufsorientierten und daher notwendigerweise praxisnahen Übersetzerausbildung. Hervorgehoben wurde, daß sich eine verantwortungsbewußte zukunftsorientierte Ausbildung nicht mit der Schulung an jeweils aktuell verfügbaren Systemen der Textbearbeitung! mit Datenbankaufbau und maschinellen Übersetzungshilfsmitteln begnügen darf. Die Tendenz zur Automatisierung und die potentielle Vielfalt im Übersetzer-

arbeitsbereich (z.B. als Berater und Evaluierer bei Modernisierungsentscheidungen durch das Firmenmanagement) zwingen zu einer auch systemunabhängigen Ausbildung, welche die Grundlagen und Arbeitsweisen moderner automatisierter Sprachverarbeitungssysteme verstehen läßt.

In der Sektion Maschinelle Korpora (Leitung: W. Lenders) wurden vier Vorträge angeboten (die oben bereits erwähnten Referate von H. Elsen/ J. Hartmann und N. Weber ließen sich hier allerdings ebenfalls einordnen). Den Anfang machte K. Wothke (Heidelberg) mit einem Vortrag über "Statistisch basiertes Wortklassentagging von deutschen Textkorpora. Einige Experimente." Wothke berichtete von Experimenten des Wissenschaftlichen Zentrums der Fa. IBM Deutschland GmbH im Rahmen der projektierten Entwicklung einer deutschsprachigen Version des Spracherkennungssystems TANGO RA. Ziel war dabei zunächst die Entwicklung einer statistisch orientierten Prozedur, die mit hoher Treffsicherheit deutsche Textkorpora mit Wortklassenangaben zu versehen vermag (Tagging). Längerfristig soll mit dem so entwickelten Algorithmus ein großes Textkorpus teilautomatisch annotiert und aus dem Ergebnis Auftretenswahrscheinlichkeiten von Wortklassen abgeleitet werden. Diese Erkenntnisse wiederum sollen der zukünftig besseren Differenzierung von akustisch ähnlichen Wörtern durch TANGORA dienen. Die vorgelegten Prozentzahlen über die bei kleinen Textmengen erreichte Trefferquote zeigten ein erstaunlich positives Bild, bei größeren Textmengen allerdings war das Ergebnis deutlich schlechter. Als Ursache vermutete Wothke einen unzureichenden Umfang des zugrundeliegenden Trainingskorpus, wodurch die Auftretenswahrscheinlichkeiten nicht hinreichend realitätskonform berechnet werden konnten.

- B. Schröder (Bonn) trug anschließend seine Überlegungen zu "Fragen der Repräsentativität linguistischer Kor-

pora" vor. Zunächst beschrieb er die grundsätzliche Problematik der gegenseitigen Abhängigkeit von Empirie und Theorie. Bei der näheren Betrachtung von Korpora als linguistischen Datensammlungen ging Schröder dann auf Fragen der Bestimmung von Datentyp, Datenbereich und Auswahlprinzipien im Verhältnis zur angestrebten Repräsentativität ein, hierbei insbesondere auf unterschiedliche Verfahren der Zufallsauswahl und der Schichtenbildung. Abschließende Betrachtungen galten der Anlage sehr großer Korpora, die für unterschiedliche Zwecke verwendbar sein sollen und daher auch eines besonderen Augenmerks bei der Gewichtung der diversen Auswahlparameter bedürfen.

- Auf der Grundlage eines bereits vorhandenen Korpus wandte sich G. Willee (Bonn) dann den "Erfahrungen mit morphologischem Tagging am Beispiel des LIMAS-Korpus" zu. Er stellte fest, daß zwar die Annotation lexikalischer Elemente in großen Korpora als Desiderat gilt und Voraussetzung für die Nutzung der Korpus-Daten in linguistischer Forschung ist, daß jedoch über die Art und Weise der Annotation noch keineswegs Einigkeit herrscht. Für seinen Test (zum Zwecke der Feststellung des benötigten Zeitaufwands) hat Willee das am IKP /Bonn zur Verfügung stehende Lemmatisierungsprogramm LEMMA2 benutzt. Für die Verarbeitung der ca. 1.1 Millionen laufenden Wortformen des LIMAS-Korpus wurde ein Zeitaufwand von insgesamt ca. 2050 Stunden errechnet, d.h. in etwa ein Jahr Arbeitszeit für einen einzelnen Bearbeiter. Bei einem Korpusumfang wie dem des Trierer dpa/taz-Korpus wären somit ca. 70 Jahre (bei einem einzigen Bearbeiter) erforderlich.

- In einem Vortrag zu "Tagging - Formen und Tools" beschäftigte sich W. Lenders (Bonn) mit unterschiedlichen Formen sowie mit einigen vorhandenen Taggingtools. Nach einer knappen Analyse der Begrifflichkeit und der Tagging-Typologie cha-

rakterisierte er eingehend entwicklungsge-
 schichtlich wichtige Taggingssysteme je nach ihrer
 Zielanwendung (Wortklassentagging,
 Syntaxtagging) und ihrer Methodik (Lexikonbasis,
 Umgebungsanalyse, morphologische Analyse,
 Vorkommenswahrscheinlichkeit). Zur
 Veröffentlichung des Referats wird ein dreiteiliger
 Anhang gehören, worin ein Überblick geboten
 wird über die typologische Einordnung
 existierender Taggingssysteme, über
 Detailcharakteristika dieser Tagger sowie über
 zwei Tagsets für das Englische (Penn Treebank
 und SUSANNA formtags).

Daß neben den genannten Vorträgen eine
 Begrüßung der Teilnehmer durch den GLDV -
 Vorsitzenden sowie die Rektorin der Universität
 Kiel anlässlich eines kleinen Empfangs, eine
 Mitgliederversammlung, ein Konzert und ein
 gemeinsames (und trotz der Eiseskälte draußen
 gemütliches) Abendessen stattfand, sei nur am
 Rande und der Vollständigkeit halber erwähnt.
 Schade nur, daß die Teilnehmerzahl nicht sehr
 groß war, die Referentinnen und Referenten hätten
 ein zahlreicheres Publikum verdient gehabt.

Horst P. Pütz (Kiel)

S+

Sprache und Information

Dagmar Schmauks **Deixis in der Mensch-Maschine- Interaktion**

Multimediale Referentenidentifikation
 durch natürliche und simulierte Zeigegesten
 1991. XII, 172 Seiten. Kart. DM 84.-. ISBN 3-484-31923-2
 (Band 23)

Wolfgang Menzel **Modellbasierte Fehlerdiagnose in Sprachlehrsystemen**

1993. IX, 230 Seiten. Kart. 98.-. ISBN 3-484-31924-0
 (Band 24)
 Vor dem Hintergrund einer Anwendung in computergestütz-
 ten Sprachlehrsystemen werden die Möglichkeiten zur ma-
 schinellen Diagnose grammatischer Fehler untersucht, wo-
 bei auf die bisher dominierende Antizipation bestimmter
 Fehler durch den Übungsautor vollständig verzichtet werden
 kann. Das Diagnoseproblem wird zunächst am Beispiel der
 Kongruenzfehler im Deutschen untersucht und die
 Ergebnisse anschließend auf weitere Teilbereiche der
 Grammatik übertragen. Die Diagnoseresultate sind als
 Ausgangspunkt für die Erzeugung von Fehlererklärungen
 und Korrekturvorschlägen sehr gut geeignet.

Hooshang Mehrjerdian **Automatische Übersetzung englischer Fachtexte ins Persische**

IX, 171 Seiten. Kart. ca. DM 92.-. ISBN 3-484-31925-9
 (Band 25)

Die vorliegende Arbeit gehört zum Themenbereich Maschi-
 nelle Übersetzung (MÜ). In der Informatik, insbesondere im
 Bereich der Künstlichen Intelligenz, beschäftigt man sich
 schon seit längerem mit der Idee, Computer in der Überset-
 zung natürlicher Sprache einzusetzen. Die Übersetzung wird
 dabei häufig auf spezielle Gebiete, z. B. wissenschaftlich-
 technische Unterlagen eingeschränkt, da solche Texte u. a.
 weniger Mehrdeutigkeiten enthalten.
 Das hier vorgestellte Übersetzungssystem, ATSTEP 1,
 wurde für die Übersetzung englischer Fachtexte ins
 Persische entwickelt. Die Komponenten des Systems,
 Analyse-, Transfer- und Synthesephase, sind unabhängig
 voneinander implementiert. Die Analysephase bildet aus
 dem eingegebenen englischen Text mit Hilfe englischer
 Wörterbücher und englischer Grammatik eine
 Zwischenrepräsentation. Die Transferphase überführt diese
 mit einem bilingualen Wörterbuch in die zielsprachliche
 Darstellung. Daran anschließend wird in der Synthesephase
 die erwünschte persische Textausgabe erzeugt.

Max Niemeyer Verlag GmbH & Co. KG
 Postfach 2140. D-72011 Tübingen

Niemeyer