

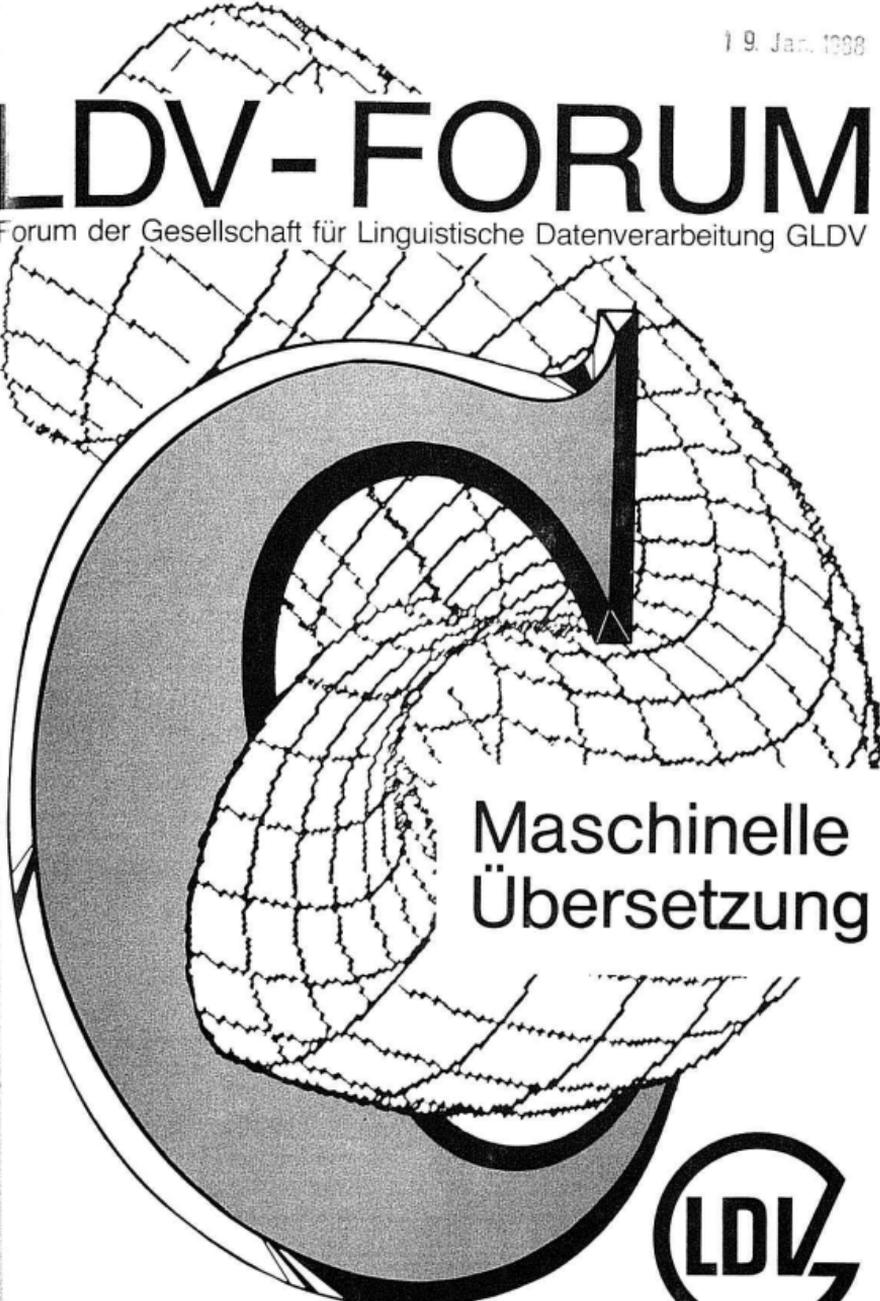
Band 5 · Nummer 1 · Juni 1987

ISSN 0172 - 9926

19. Jan. 1988

LDV-FORUM

Forum der Gesellschaft für Linguistische Datenverarbeitung GLDV



Maschinelle
Übersetzung



Editorial

LDV-FORUM

Forum der Gesellschaft für Linguistische Datenverarbeitung GLDV Band 5 - Nummer 1 - Juni 1987

LDV-FORUM 5(1987)1

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

Redaktion

Prof. Dr. G. Knorz, FH Darmstadt,
Fachbereich IuD, Schöfferstraße 8, D-
6100 Darmstadt

(Tel. THD, Sekr.: (06151) 16-2859)

FHD, Sekr.: (06151) 317881)

Netz: xid2knor@ddathd21.bitnet

Mitarbeiterin der Redaktion für Text-
fassung und TeX-Satzerstellung: Sigrid
Becker, TH Darmstadt

Wissenschaftlicher Beirat des LDV-Forum:

I. S. Batori, R. Drewek, Ch. Habel, P.
Hellwig, G. Knorz, J. Krause, R. Kuhlen,
H.D. Lutz, D. Rösner

Wissenschaftliche Zusammen- arbeit

Prof. Dr. J. Raben, *Computers and
the Humanities, SCOPE*

Erscheinungsweise:

halbjährlich, zum 30. Juni und 30. De-
zember

Bezugsbedingungen:

Der Bezug des LDV-Forum ist im Regel-
fall an die Mitgliedschaft in der GLDV
gebunden und im Jahresbeitrag einge-
schlossen. Jahresabonnements können
von Privatpersonen zum Preis von DM
30,- (incl. Versand), Einzelzeitschriften
zum Preis von DM 15,- (zuzügl. Ver-
sandkosten) bei der Redaktion bestellt
werden.

EDITORIAL

Was haben die Zeitschriften *Computer and the Humanities*, *SCOPE* und das *LDV-Forum* gemeinsam? Eine Antwort könnte lauten: einen nicht unerheblichen Teil der Interessen ihrer Leser/-innen. Daß die sich daraus ergebenden Möglichkeiten einer Zusammenarbeit tatsächlich genutzt wurden, konnten die aufmerksamen Leser/-innen bereits in den letzten Ausgaben des LDV-Forum z.T. anhand einiger Quellenangaben feststellen. Sie finden diese Kooperation nunmehr auch im Impressum der genannten Zeitschriften formal dokumentiert unter den Überschriften *Wissenschaftliche Zusammenarbeit* bzw. *Corresponding Editors*. Aus der Sicht des LDV-Forum stellt dieser Schritt für die Zukunft sicherlich einen Anspruch dar, aber er bedeutet gleichzeitig auch Motivation und schafft eine weiterhin verbesserte Arbeitssituation. Daß für den erreichten Entwicklungsstand des LDV-Forum die Kooperation mit weiteren Zeitschriften, nämlich *Infoterm/Bibliothem* und *Zeitschrift für Semiotik* bzw. Institutionen wie dem *GID-Informationszentrum (GID/IZ)* eine wichtige Rolle gespielt hat und weiter spielt, soll an dieser Stelle ausdrücklich gewürdigt werden.

Sicher haben Sie das Inhaltsverzeichnis auf dem Rückumschlag der vorliegenden Ausgabe des LDV-Forum bereits überflogen. Der Themenschwerpunkt *Maschinelle Übersetzung* kann mit der ersten größeren Veröffentlichung über EUROTRA in deutscher Sprache aufwarten (siehe dazu S. 4), und es wird über die im Februar 1987 in Saarbrücken/Dudweiler durchgeführten Veranstaltungen von EUROTRA-D und EUROTRA-D-Begleitforschung berichtet. Wenn uns die transatlantische Post keinen Streich spielt, wird auch eine einschlägige Information über *Computers and Translation (CaT)*, eine neue internationale Zeitschrift für MU, beiliegen. Der Fachbeitrag über die *formale Begriffsanalyse* wendet sich an die interessierten Leser/-innen, die bereit sind, etwas Mühe in die Erarbeitung eines ihnen (vermutlich) neuen methodischen Instrumentes zu investieren. Es dürfte für viele Leser/-innen von Interesse sein, daß die dort erläuterten Methoden und Darstellungsformen z.Z. auch für den nicht unbekanntenen Problembereich der *Systemkonfiguration* mit vielversprechenden Ergebnissen erprobt werden.

Für eine neue geplante Rubrik *Computeringuistik in Frankreich* ist die Resonanz unserer französischer Kollegen und Kolleginnen gegeben. Aufgrund zeitlicher Probleme konnte sie diesmal noch nicht starten, sie wird aber nun ab nächstem Heft 2/87 eingerichtet sein (siehe dazu S. 50).

Vergeblich werden interessierte Leser/-innen diesmal nach der Überschrift *Software - Courseware* suchen. Stattdessen bietet das LDV-Forum einen neuen Service: Ausgewählte Beiträge aus den Rubriken *Software*, *Courseware* und *Software Reviews* der Zeitschrift *SCOPE*, verwaltet in einem hierarchischen Menu-System, können auf 5¹/₄ Zoll Disketten für IBM-PC's und Kompatible mit MS/DOS zum Selbstko-



stenpreis von DM 9,- bei der Redaktion bestellt werden. Erforderlich sind eine Überweisung auf das Konto *GLDV/LDV-Forum, Sparkasse Darmstadt, BLZ 508 501 50, Kto.Nr.: 554 090*) sowie eine schriftliche Anforderung mit vorbereitetem und frankiertem Rückumschlag. Zur Weiterverbreitung der Diskette durch Kopieren wird ausdrücklich ermuntert. Wenn Kollegen und Kolleginnen konkrete Anregungen für Updates haben, sind diese herzlich willkommen!

Die Disketten-Version einer Zeitschriftenrubrik kann als ein Einstieg in neue Formen der Verbreitung aktueller Information betrachtet werden. Die GLDV hat in dieser Hinsicht einen noch viel weiterreichenden Schritt unternommen. Die Kollegen *R. Drewek* und *D. Rösner*, beide Stuttgart, bereiten einen *newsletter* vor, der via Netz und vorerst (noch) parallel auch als Papierversion verbreitet werden soll und der damit als ein besonders aktuelles Informationsmedium das LDV-Forum ergänzt.

Auch das LDV-Forum selbst wird versuchen, für Leser/-innen, Autoren und Autorinnen (und auch für Redaktion) noch attraktiver zu werden. Der Fachbereich *Information und Dokumentation* der *Fachhochschule Darmstadt*, als die neue Anschrift der Redaktion, sollte ein geeignetes Umfeld für einschlägige konzeptionelle und technische Innovationen bereitstellen. Wenn sich die Studenten/-innen des Studienganges *Medien- und Wirtschaftsinformation* im Wintersemester 87/88 im Rahmen des Projektstudiums u.a. auch für *electronisches Publizieren* entscheiden können, so können daraus schon die richtigen Schlüsse gezogen werden ...

Gerhard Knorz

- P.S. Die Dezemberausgabe 2/87 wird auf einen Themenschwerpunkt verzichten, um dafür mehr Raum für Fachbeiträge verschiedener Ausrichtung zu bieten. Vorgesehen sind insbesondere Beiträge zur Analyse von Texten für große Informationssysteme, zur maschinellen Übersetzung, zur Textgenerierung, zur Verschriftlichung von Videoprotokollen sowie zur Evaluierung von natürlichsprachigen DB-Zugangssystemen. Auch der Übersicht über die "graue Literatur" soll beim nächsten Mal mehr Platz gegeben werden.

Back-up-Exemplare bis einschließlich Ausgabe 2/85 sind zum Preis von DM 10,- (zuzügl. Porto) zu bestellen. Die Ausgabe 1/86 ist vergriffen.

Titelgestaltung:

Fred Zimmermann, *FH Darmstadt*

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 8 Seiten. Für Beitragsreihen gelten besondere Randbedingungen (vgl. LDV-Forum 4(1986)2: S. 25).

Redaktionsschluß:

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

Herstellung:

Verlagsdruckerei Hoppenstedt, Havelstraße 9, D-6100 Darmstadt

Auflage:

500 Exemplare

Anzeigen:

Media-Information kann bei der Redaktion angefordert werden.

Bankverbindung:

Sparkasse Darmstadt, BLZ 508 501 50, Kto.-Nr. 554 090

Gesellschaft für Linguistische Datenverarbeitung (GLDV)

Anschrift:

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

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

ANZEIGEN

Olms-Verlag: <i>Reihe "Linguistische Datenverarbeitung"</i>	3
ECO-Institut: <i>C-Tools</i>	14
IKS, Bonn: <i>WORD CRUNCHER</i>	20
epsilon: <i>MPROLOG</i>	22
ECO-Institut: <i>ECOINDEX</i>	30
Hoppenstedt: <i>Stellenanzeige</i>	66
Niemeyer-Verlag: <i>Reihe "Sprache und Information"</i>	Beilage
Paradigm Press: <i>Reihe "Computer and Translation"</i>	Beilage

LINGUISTISCHE DATENVERARBEITUNG

In Zusammenarbeit mit der Gesellschaft für Linguistische Datenverarbeitung (GLDV) herausgegeben von *Peter Hellwig*, Heidelberg und *Jürgen Krause*, Regensburg.

Band 1:

Trends in der Linguistischen Datenverarbeitung
Beiträge der Jahrestagung 1984 der Gesellschaft für Linguistische Datenverarbeitung e. V.
Herausgegeben von *Peter Hellwig* und *Hubert Lehmann*. Hildesheim 1986.
VIII/188 S. ISBN 3-487-07679-9 DM 29,80
Lieferbar!

Band 2:

Josef Hofmann, Breitbandiger natürlich-sprachlicher Mensch-Rechner-Dialog
Skizze eines Verarbeitungsmodells am Beispiel der Thematisierung. Hildesheim 1986. XIV/293 S.
ISBN 3-487-07737-X DM 39,80
Lieferbar!

Band 3:

Klaus Dirk Schmitz, Automatische Segmentierung natürlichsprachiger Sätze
Hildesheim 1986. X/204 S.
ISBN 3-487-07762-0 DM 29,80
Lieferbar!

Band 4:

Informationslinguistische Texterschließung
Herausgegeben von Ch. Schwarz und G. Thurmair. Hildesheim 1986. VIII/281 S.
ISBN 3-487-07763-9 DM 29,80
Lieferbar!

Band 5:

Automatische Indexierung zwischen Forschung und Anwendung
Herausgegeben von G. Lustig. Hildesheim 1986.
ISBN 3-487-07825-2 DM 29,80
Lieferbar!

Band 6:

Ioanna Andreadou, Software für den Fremdsprachenunterricht
Marktübersicht, Adressen, Bibliographie.
IV/211 S. DM 29,80

Band 7:

Computerlinguistik und Philologische Datenverarbeitung
Beiträge der Jahrestagung der Gesellschaft für Linguistische Datenverarbeitung (GLDV) 1986 in Göttingen. Herausgegeben von *Ursula Klein*, *Peter Scherber* und *Manfred Thaller*.

Band 8:

Inhalterschließung von Massendaten
Zur Wirksamkeit informationslinguistischer Verfahren am Beispiel des Deutschen Patentinformationssystems. Herausgegeben von *J. Krause*. Hildesheim 1987. XII/248 S.
ISBN 3-487-07839-2 DM 39,80
Lieferbar!

Band 9:

Analyse und Synthese gesprochener Sprache
Vorträge im Rahmen der Jahrestagung 1987 der Gesellschaft für Linguistische Datenverarbeitung e. V., Bonn, 4.-6. März 1987. Herausgegeben von *H. G. Tillmann* und *G. Willée*.
Hildesheim 1987. 179 S.
ISBN 3-487-07847-3 DM 29,80
Die Reihe wird fortgesetzt!

Außerhalb der Reihe sind erschienen:

Computer und Übersetzen

Eine Einführung. Herausgegeben von *Achim Blatt*, *Karl-Heinz Freigang*, *Klaus-Dirk Schmitz* und *Gisela Thome*. Hildesheim 1985. XVII/332 S. (HILDESHEIMER BEITRÄGE ZU DEN ERZIEHUNGS- UND SOZIALWISSENSCHAFTEN, Bd. 21)
ISBN 3-487-07661-6 DM 39,80

Sprachsynthese

Herausgegeben von *Bernd S. Müller*. Hildesheim 1985. 220 S. (GERMANISTISCHE LINGUISTIK 79-80/85) ISBN 3-487-07676-4 DM 39,80

Symposium zur Lexikographie

Hrsg. von *Karl Hyldegaard-Jensen* und *Arne Zettersten*. 228 S.
(GERMANISTISCHE LINGUISTIK 5-6/82)
ISBN 3-487-07461-3 DM 35,80

Harold Woetzel, Historisch-systematische Untersuchungen zum Komponentialismus in der linguistischen Semantik

Eine Kritik des Elementarismus. 380 S.
(GERMANISTISCHE LINGUISTIK 1-2/84)
ISBN 3-487-07525-3 DM 54,80

 Georg Olms Verlag · Hagentorwall 7 · 3200 Hildesheim

Themenschwerpunkt

THEMENSCHWERPUNKT MASCHINELLE ÜBERSETZUNG

Unter der globalen Überschrift *Maschinelle Übersetzung* für den Themenschwerpunkt des LDV-Forum werden im wesentlichen die Arbeiten der deutschen EUROTRA-D-Forschungsgruppe am Institut für Angewandte Informationsforschung IAI in Saarbrücken vorgestellt. Da die Arbeitssprache des EUROTRA-Projekts Englisch ist, liegt mit dieser Darstellung die erste größere Veröffentlichung in deutscher Sprache vor.

Das Thema *Maschinelle Übersetzung* hat in letzter Zeit soviel Aufmerksamkeit gefunden, daß man schwerlich von einer originellen oder ausgefallenen Themenwahl sprechen kann. Sicher handelt es sich aber um ein besonders interessantes und zugängliches Thema auch für diejenigen Kollegen und Kolleginnen, die nicht aktiv in diesem Bereich arbeiten. Aus diesem Grunde schien es nicht vorrangig, mit vergleichsweise allgemeinen Beiträgen in das Thema einzuführen. Das Konzept von H. Weber, der dankenswerterweise die Akquisition der Beiträge für diesen Schwerpunkt übernommen hatte, war es vielmehr gewesen, eine aktuelle Darstellung der deutschen "MÜ-scene" zu erreichen. Die Fachbeiträge in diesem Heft sind im Anschluß an das Statusseminar zu EUROTRA-D entstanden, das nach Maßgabe des Bundesministeriums für Forschung und Technologie vom 16. bis 18. Februar 1987 in Saarbrücken/Dudweiler durchgeführt wurde. Beteiligt waren neben dem Veranstalter IAI/EUROTRA-D auch die Begleitforschungsprojekte an den Universitäten Stuttgart, Berlin und Bielefeld. Sowohl über den Verlauf dieses Statusseminars (*Haller/Reyle*) als auch über den sich anschließenden *Zweiten Externen Workshop der EUROTRA-D-Begleitforschung* (*Hauenschöld/Busemann*) wird berichtet. Die Reihe der Beiträge wird eingeleitet durch eine Übersicht über den derzeitigen Stand der Forschung auf dem Gebiet der maschinellen Übersetzung weltweit. Dieser Beitrag von *J. Haller* stützt sich in der Hauptsache auf Veranstaltungen im Jahre 1986 und stellt insofern eine Aufarbeitung dieser Konferenzen dar. Die sich anschließenden Einzelbeiträge geben Beispiele für die linguistische Forschung bei EUROTRA-D. In diesem Kontext wird auch über ein Projekt an der Universität Bonn berichtet, das sich im Auftrag von EUROTRA-D mit dem Aufbau von mehrsprachigen Wörterbüchern beschäftigt.

Jetzt könnte leicht jemand auf die Idee kommen, daß mindestens zwei Dinge fehlen: Eine Darstellung der Arbeiten der EUROTRA-D-Begleitforschung und ein (gegenüber diesem redaktionellen Beitrag) etwas mehr grundsätzlicher Kommentar von H. Weber. Tatsächlich ist der vorliegende "Vorspann" zum Themenschwerpunkt nur ein "last-minute-Ersatz" für ein spezielles Editorial, das vermutlich kurz nach Drucklegung bei der Redaktion eintreffen wird. Und auch die Darstellung der in der Begleitforschung entwickelten Konzeptionen war nur deshalb nicht möglich, weil der Terminkalender zu eng war. In diesem Falle gilt: "aufgeschoben ist nicht aufgehoben"! In diesem Sinne bleibt Maschinelle Übersetzung auch ein Thema für das Dezember-Heft.

d. Red.

FORSCHUNGSSYSTEME FÜR MASCHINELLE ÜBERSETZUNG – EIN ÜBERBLICK

Johann Haller
IAI/EUROTRA-D
Martin-Luther-Straße 14
D-6600 Saarbrücken

Es wird ein Überblick über ausgewählte aktuelle Forschungssysteme zur maschinellen Übersetzung gegeben. Insbesondere werden die Lösungsansätze der behandelten Systeme den Konzepten für das EUROTRA-Projekt der Europäischen Gemeinschaft gegenübergestellt. Ausführlich wird das japanische Nationalprojekt vorgestellt, das im Rahmen des Programms zur Fünften Generation durchgeführt wird. Aus den USA werden Versuche mit "logic based machine translation" und ein auf dem Syntaxanalyse-System CRITIQUE aufsetzendes System kurz behandelt. Von den europäischen Systemen wird am ausführlichsten das französische CALLIOPE-Projekt beschrieben. Eine Reihe weiterer Forschungssysteme, sowohl aus Japan, USA und EUROPA werden grob klassifiziert und es werden Literaturhinweise gegeben.

Im Jahre 1986 fanden drei größere Veranstaltungen statt, die das Thema der Maschinellen Übersetzung als einziges oder doch als ein Hauptthema hatten: die allgemein bekannte COLING 86 in Bonn, in Verbindung damit die IAI-MT-"Messe" in Saarbrücken/Dudweiler und - fast zeitgleich - eine Veranstaltung des "European Language Service" der IBM in Kopenhagen IBM-ELS 1986.

Auf allen drei Veranstaltungen wurden im wesentlichen dieselben Systeme vorgestellt, in Saarbrücken wurden sie zum Teil auch praktisch durchgeführt.

Dabei wurden nur auf der IAI-Konferenz zur Maschinellen Übersetzung in Dudweiler auch zwei kommerzielle Systeme vorgestellt (SYSTRAN und LOGOS). Diese Systeme sollen als industrielle Entwicklungen hier nicht beschrieben werden, ebenso wenig wie das von der Universität Austin/Texas entwickelte und nunmehr von SIEMENS in Deutschland, Belgien und Spanien weiterbetriebene METAL-System. Forschungssysteme dagegen wurden auf allen drei Veranstaltungen präsentiert; sie werden im folgenden in geographischer Ordnung kurz referiert.

1.1 Asien - Japan

Der zahlenmäßige Schwerpunkt lag dabei auf Darstellungen der japanischen MÜ-Forschung, über deren wichtigste Ergebnisse im folgenden kurz berichtet wird. So gab Tsujii IBM-ELS 1986 eine detaillierte Übersicht über die Vorgehensweise des japanischen Nationalprojekts, das im Rahmen des Programms zur Fünften Generation durchgeführt wird. In der ersten Phase bis März 1986 wurde ein Prototyp für die Übersetzung Japanisch-Englisch erstellt, der mit ca. 80.000 Wörterbucheinträgen arbeitet. Die linguistischen Regeln dieses Prototyps sind in einer speziell entworfenen Regelsprache GRADE geschrieben und werden durch LISP-Programme interpretiert. Dabei wird ein "tree transducer" nach dem Vorbild des GETA-Systems (Grenoble) verwendet; weitere Prinzipien sind das Konzept eines neutralen Wörterbuchs, die starke Lexikonsteuerung sowie der Einsatz heuristischer Methoden. Die Analyse selbst geschieht in 15 kleinen Schritten:

- Postmorphologische Analyse: Erkennung von Ausdrücken
- Behandlung von Frage- und Befehlsätzen
- lokale Disambiguierungsversuche (1) (entspricht etwa der Homographenanalyse in den SUSY-Systemen)

- Erkennung von Adverbialsätzen
- Erkennung einfacher Nominalgruppen
- Erkennung von Verbalgruppen
- Koordinierte Verbalgruppen
- Koordinierte Nominalgruppen
- Appositionen
- Erkennung der gesamten Satzstruktur
- lokale Disambiguierung (2)
- Case Pattern Matching (Verwendung der Valenzinformationen)
- Umwandlung der Phrasendstruktur in eine Dependenzstruktur
- Semantische Behandlung (tense, aspect etc.)
- Prüfregeln für die Korrektheit der Gesamtanalyse.

Der Transfer wird in weiteren 5 Schritten durchgeführt:

- Erzeugung einer neutralen semantischen Struktur
- Übertragung der Prädikate
- Übertragung nominaler Konzepte
- Übertragung adverbialer Konzepte
- Übertragung pronominaler modifizierender Elemente (z.B. Adjektive)
- Erzeugung einer Dependenzstruktur in der Zielsprache.

Nagao/Tsujii (in COLING 1986) geben noch weitere Informationen zur Transferstrategie. Sie teilen den Transfer in drei Phasen: Pre-, Main- und Post-Transfer. Dabei wird eine Top-Down-Strategie verfolgt, ähnlich wie sie auch in EUROTRA vorgesehen ist; die Auswahl von lexikalischen Einheiten für höhere Knoten erzeugt Restriktionen für die Auswahl der zielsprachlichen Elemente in den abhängigen Knoten. In der Pre-Transfer-Phase werden etwa zweiteilige Konjunktionen wie "nicht nur... sondern auch..." behandelt, in der Post-Transfer-Phase werden Rotationen ganzer Gruppen gemäß den Stellungenregeln der Zielsprache durchgeführt. In der Main-Transfer-Phase werden z.B. Kombinationen aus zwei Nomina (entsprechend dt. Komposita "Holztisch" oder "Baumwachstum") in Adjektiv-Nomen-Verbindungen ("wooden desk") oder Nomen-Nomen-Verbindungen ("growth of trees") umgewandelt; dabei werden Tiefenkasusrelationen wie "Material" oder "Goal" zur Unterscheidung verwendet. Die Disambiguierung von Verben geschieht ebenfalls im Transfer, wenn es sich um Varianten der Hauptbedeutung, des "core meanings" handelt; andernfalls werden in der Analyse

verschiedene Lesarten unterschieden. EUROTRA in seiner gegenwärtigen linguistischen Konzeption legt dagegen bisher das Hauptgewicht auf die Disambiguierung in der Synthese, den von EUROTRA-D vorgeschlagenen semantischen Relationen liegt ein Konzept zugrunde, das dem obigen Vorgehen ähnlich ist.

Die Evaluierung der Ergebnisse des japanischen MU-Projekts dauert noch an; bisher soll jedoch bereits ein hoher Grad an Verständlichkeit erreicht worden sein. Eine zweite Phase, die sich bis zum Jahre 1990 erstrecken soll, hat bereits begonnen; während dieser Zeit sind folgende Maßnahmen geplant:

- Ausbau des Wörterbuchs auf 300.000 Wörter
- Neumentrierung von GRADE in PL/I zur Erreichung kürzerer Laufzeiten
- Einführung von Prädition und Einschränkung des Sprachumfangs
- Ausbau der Interaktionskomponente mit dem menschlichen Bearbeiter.

Der Schwerpunkt liegt also für die zweite Phase nicht auf einem Redesign der linguistischen Grundlagen, sondern mehr im technischen Bereich. Auffällig ist die starke heuristische Komponente, die die Abarbeitungsstrategien in den Vordergrund stellt; im Gegensatz dazu legt die EUROTRA-Konzeption zunächst größeren Wert auf die theoretische Fundierung und auf die Formulierung statischen linguistischen Wissens und möchte die Strategien erst zu einem späteren Zeitpunkt miteinbeziehen. Die Anwendungsorientierung des japanischen Projekts ist damit begründet, daß es zunächst bei einer großen Datenbank (JICST) zur Übersetzung von Titeln und Abstracts im täglichen Betrieb eingesetzt werden soll (Sakamoto/Washio in *COLING 1986*).

Einige neuere Arbeiten an japanischen Universitäten untersuchen auch Transferstrategien auf der Basis einer lexikalisch-funktionalen Grammatik, ähnlich wie die in 5.2. beschriebene EUROTRA-D-Begleitforschung (Kudo in *COLING 1986*). Schließlich existieren noch kleinere Systeme wie LUTE-AID, beschrieben von Nomura et al. in *COLING 1986* und in *IAI-MT 1986* und NARA zur Übersetzung Koreanisch-Japanisch, beschrieben von Chung/Kuni in *COLING 1986*, das eine auf GPSG basierende Syntax benutzt, sich in Semantik und Transfer aber klar zu "empirischen und heuristischen Methoden" bekennt. Für das gleiche Sprachpaar findet sich auch das System KANI in Korea (Chang Ho in *IAI-MT 86*), das in einem syntaxbasierten Verfahren die relative Ähnlichkeit der beiden agglutinierenden Sprachen auszunutzen versucht. Ein auf ARIANE (siehe unten) basierendes System für Englisch-Malaisisch beschreibt Tong in *COLING 1986*.

1.2. USA

In den USA macht McCord (*IBM-ELS 1986*) Versuche mit "logic based machine translation", die nach dem Prinzip der "modular logic grammars" Sätze aus Computermanualen vom Englischen ins Deutsche übersetzen. Auch er verwendet einen syntaxbasierten Transferansatz, der durch semantisches "type-matching" von slot-frames ergänzt wird; die semantische Klassifikation lehnt sich an Vorschläge von Dahl/McCord 1983 an. Präferenzen bilden die heuristische Komponente, indem sie eine Liste von offenen Slots in gewisser Reihenfolge abarbeiten. Das Modell, das bisher nur einen geringen Umfang hat,

soll durch die Integration eines 26.000 Elemente umfassenden Fachwörterbuchs zur Informatik für die praktische Verwendung ausgebaut werden. In einem anderen IBM-Forschungszentrum entwickelt Heidor (*IBM-ELS 1986*) ein Analyseprogramm CRITIQUE zur Überprüfung englischer Texte auf orthographische, morphologische und syntaktische Korrektheit mit einigen Stilvorgaben; dieses Programm benutzt Tsutsumi (in *COLING 1986*) als Ausgangspunkt für ein Übersetzungssystem Englisch-Japanisch, wobei (ähnlich wie das japanische Nationalprojekt) spezielle semantische Kategorien und Relationen wie MATERIAL etc. einführt, mit denen Transferprobleme zwischen diesen beiden Sprachen gelöst werden sollen.

Die beiden anderen amerikanischen Projekte (KBMT, von Tomita/Carbonell in *COLING 1986* beschrieben) und TRANSLATOR von Nirenburg (*COLING 1986*) funktionieren für minimale Bereiche und sind wissensbasiert. TRANSLATOR verwendet ein frame-Konzept, KBMT benutzt für die Syntax funktionale Grammatiken und eine "entity-oriented" Semantik, d.h. es wird immer eine gesamte Handlungseinheit betrachtet.

1.3. Europa

In Europa existiert neben EUROTRA noch ein größeres nationales Projekt, CALLIOPE, das auf den Vorarbeiten der Universität Grenoble beruht und von GETA sowie französischen Softwarefirmen durchgeführt wird (Boitet in *IBM-ELS 1986*). Dabei ist der Kern des Übersetzungssystems aus ARIANE-78 hervorgegangen, dem von GETA entwickelten Prototyp, der als ARIANE-85 auch auf PC-AT verfügbar gemacht werden und schließlich als ARIANE-X einem generellen Redesign unterzogen und in LISP implementiert werden soll. Die Version CALLIOPE-Aero übersetzt Texte aus der Flugzeugtechnik vom Französischen ins Englische, eine geplante Version CALLIOPE-Comp wird Computer-Manuals vom Englischen ins Französische übersetzen. Auch dieses Projekt arbeitet mit einem Transfer-Ansatz und Überführungsregeln zwischen "dekorierten Bäumen". Das Wörterbuch ist nach Ableitungsfamilien geordnet, um strukturellen Transfer (etwa zwischen Verbalisierung und Nominalisierung) zu erleichtern und auch, um die Nähe zu einer Spielart von konventionellen Lexika zu wahren. Boitet/Nedobekine (in *COLING 1986*) betonen die Notwendigkeit einer Integration von "menschlichem" und maschinellem Lexikon, in denen der Zugriff auf Wörter über das Lemma erfolgen soll. Sie beschreiben auch das Konzept eines "Gabel-Lexikons", das zu der Information des Ausgangssprachlichen Elements die Entsprechungen in den (bisher drei) Zielsprachen enthalten soll. Die Grammatik- und Strategieregeln sind in einer "Spezialsprache für linguistische Regelsysteme" formuliert, wobei es ähnlich wie in Japan eine starke heuristische Komponente gibt. Die linguistischen Daten selbst werden in einer "Spezifizierungssprache" niedergelegt; deren Eingabe graphisch unterstützt wird; der Linguist zeichnet Bäume und füllt Dekorationen. Dabei hat er auch die Möglichkeit, "ambiguity charts" mit aufzuführen, die bei der Erstellung der Strategien mit berücksichtigt werden (Zajac in *COLING 1986*). Die Übersetzungseinheit übersteigt die Satzgrenze; normalerweise werden zwei bis drei Absätze behan-

delt. Einen in der Vorgehensweise ähnlichen Vorschlag macht auch Luckhardt (*COLING 1986*) vom bis Ende 1986 bestehenden SFB 100 in Saarbrücken mit den "Controlled Active Procedures" (CAP), wobei das Parsing-Problem in kleine Teilschritte zerlegt wird, die mit Methoden von "linguistic engineering" gelöst werden.

Die Gruppe DIMA in Italien (*Oitana 1985*) ist wiederum frame-orientiert; das Konzept folgt ähnlichen Grundsätzen wie die beiden wissenschaftsbasierenden amerikanischen Projekte, versucht aber eine gewisse Allgemeingültigkeit durch die Anreicherung der Slot-Filler mit semantischen Kategorien zu erreichen; Ambiguitäten werden teilweise durch menschliche Intervention beseitigt.

Den Autor als menschlichen Helfer will das an der Universität Manchester (UMIST) von Whitelock et al. (*COLING 86*) entwickelte System Englisch-Japanisch heranziehen.

Dieselbe Strategie ist auch in ROSETTA (Appelo in *IAI-MT 1986*) geplant, das für das Sprachenpaar Niederländisch-Englisch bis 1988 eine 3. Version fertigstellen will. Es strebt ähnliche Ziele wie EUROTRA an, wird ausdrücklich als Forschungssystem bezeichnet und verwendet keine Wissensbasen. Die wichtigsten Prinzipien sind:

- Verwendung expliziter Grammatiken (ähnlich wie die "statischen Grammatiken" des französischen Nationalprojekts, s.u.)

- Prinzip der Kompositionalität aus der Montague-Grammatik; dabei sollen für die verschiedenen Sprachen (neben Niederländisch und Englisch auch Spanisch) aneinander angepaßte Grammatiken geschrieben werden.
- Verwendung derselben Grammatik für Analyse und Synthese (wie bei den meisten Systemen)
- Prinzip der "Isomorphie", d.h. daß zwei Ausdrücke als gleichbedeutend angesehen werden, wenn sie "auf die gleiche Weise von der gleichen semantischen Basis abgeleitet sind". Bisher sind für dieses System umfangreiche Arbeiten zum Zeit-Problem (Appelo in *COLING 1986*) und zur Behandlung von Idiomen (Schenk in *COLING 1986*) geleistet worden.

In der Bundesrepublik Deutschland werden neben EUROTRA-D auch unabhängige Begleitforschungsprojekte an den Universitäten Stuttgart, Berlin und Bielefeld gefördert, die in Haller/Reyle 1987 kurz und in einigen Einzelbeiträgen des vorliegenden LDV-Forum ausführlicher beschrieben werden.

Literatur

Die Literaturangaben für die Fachsätze zum Themenschwerpunkt Maschinelle Übersetzung finden sich zusammengestellt auf Seite 21.

Gesellschaft für Linguistische Datenverarbeitung (GLDV)

Anschrift:

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

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

Vorstand:

B. Endres-Niggemeyer (1. Vorsitzende),
K.G. Schweisthal (2. Vorsitzender), B.
Schaefer (Schatzmeister), Ch. Schnei-
ter (Schriftführerin)

Arbeitskreise:

Ausbildung und Berufsperspektiven,
LDV und Nachbarn; Maschinelle Lexi-
kographie und Lexikologie; Maschinelle
Übersetzung; Spracherkennung, Sprach-
generierung und phonetische Datenban-
ken; Textanalyse

Studienführer LDV

Lutz-Hensel, M.: Studienführer Lin-
guistische Datenverarbeitung (LDV) für
die wissenschaftlichen Hochschulen der
Bundesrepublik Deutschland, 1985.
zu beziehen über: Prof. Dr. J. Krause,
Universität Regensburg, Linguistische
Informationswissenschaft, Postfach 397,
8400 Regensburg



KONTAKTADRESSEN der ARBEITSKREISE

AK "Textanalyse"

S. Goeser, Sonderforschungsbereich 129, Universität
Ulm, Am Hochstr. 8, D-7900 Ulm

AK "Spracherkennung, Sprachgenerierung und Pho- netische Datenbanken"

K.G. Schweisthal, Institut für Phonetik und sprachliche
Kommunikation der Universität München, Schel-
lingstraße 3 V6, D-8000 München 40

AK "Maschinelle Übersetzung"

Dr. Dietmar Rosner, c/o Projekt SIMSYN, Institut
für Informatik, Azenbergstr. 12, D-7000 Stuttgart 1

AK "LDV und Nachbarn"

siehe dazu den Bericht des Arbeitskreises in dieser
Rubrik

AK "Ausbildung und Berufsperspektiven"

Magdalena Lutz-Hensel, Institut für Kommunikati-
onsforschung und Phonetik, Poppelsdorfer Allee 47,
D-5300 Bonn 1

AK "Lexikographie"

Jan Brustkern, Institut für angewandte Kommunika-
tions- und Sprachforschung e.V. Bonn (IKS), Pop-
pelsdorfer Allee 47, D-5300 Bonn 1

EUROTRA – GEGENWÄRTIGER STAND UND PLANUNG 1987/88

Johann Haller
IAI/EUROTRA-D
Martin-Luther-Straße 14
D-6600 Saarbrücken

Das multinationale EUROTRA-Projekt zur maschinellen Übersetzung befindet sich gegenwärtig in der Phase II, die bis Juni 1988 abgeschlossen sein soll. In dieser Phase der linguistischen Forschungen sollen Defizite bei den kooperativen Aufgaben wie Entwicklung und formale Beschreibung linguistischer Strategien oder die Ermittlung, Definition und Spezifikation von universeller linguistischer Information durch verstärkte Anstrengungen ausgeglichen werden. Die auf die Einzelsprache bezogenen Arbeiten sollen Zuverlässigkeit und Coverage der Systemkomponenten erhöhen und die theoretische Ausarbeitung verbessern. Die Phase III soll dann bis Mitte 1990 der Systemstabilisierung und der Auswertung dienen. Der vorliegende Beitrag stellt neben dem Forschungsstand auch den Stand des Organisationsbereichs dar und unterstreicht dessen Bedeutung für den Erfolg des Forschungsprojektes.

Seit dem Beginn der Phase II des Projekts EUROTRA im Jahre 1985 sind weltweit die Aktivitäten auf dem Gebiet der Maschinellen Übersetzung und besonders auch in der sog. "Maschinengestützten Übersetzung" verstärkt gefördert worden. In Europa werden drei Systeme zur Maschinellen Übersetzung von der Industrie auf den Markt gebracht und in einigen Fällen auch in größerem Umfang angewendet; dazu kommt eine große Zahl von Software-Paketen mit integrierter Textbearbeitung und Terminologiepflege zur Unterstützung des professionellen Übersetzers.

Die kurze Übersicht in Haller 1987a zeigt, daß das Interesse an der Erstellung von MU-Systemen und der Forschung auf diesem Gebiet sehr stark gewachsen ist; allerdings haben auch die verstärkten Anstrengungen bisher noch nicht zu einem System geführt, das den bisher existierenden durch seine linguistische Konzeption, seine gründliche Ausarbeitung und benutzerfreundliche Software überlegen wäre.

Dabei gelingt es heute bereits, mit Einsatz geeigneter Software zur Sprachdatenverarbeitung den Linguisten von informatischen Fragestellungen weitgehend zu entlasten; die größeren Probleme liegen nach wie vor darin, eine für ein multilinguales System geeignete linguistische Beschreibungsebene zu finden, die auf die Belange der Übersetzung zugeschnitten ist und gleichzeitig ein Maximum an Effizienz erlaubt.

Auf dem Wege zu diesen Zielen ist EUROTRA in den letzten Jahren sowohl organisatorisch als auch technisch ein Stück vorwärtsgekommen. Die Mitgliedsländer der Europäischen Gemeinschaft haben (inklusive Spanien und Portugal) nunmehr alle Assoziationsverträge unterschrieben, in denen sie sich verpflichten, die nationalen Beiträge zur Durchführung dieses Forschungs- und Entwicklungsprogramms zu leisten; der Schwerpunkt der Organisatorischen Arbeit liegt darin, die durch die in den einzelnen Ländern unterschiedlichen Fortschritte erzielten Ergebnisse einander anzugleichen. Zu diesem Zweck hat auch das Projektteam der EG-Kommission eine erhebliche Verstärkung erfahren.

Durch den frühzeitigen vollen Einsatz einiger Mitgliedsländer (darunter auch die Bundesrepublik Deutschland) liegen nunmehr erste Erfahrungen mit den linguistischen Spezifikationen sowie der vorgesehenen Software vor, was erwarten läßt, daß die Entwicklung für die restlichen Länder beschleunigt erfolgen kann; die erste praktische Demonstration

einer Übersetzung zwischen drei Sprachen wurde am 9.2.1987 in Luxemburg durchgeführt. Die deutsche Forschungsgruppe konnte wegen ihrer günstigen Voraussetzungen dazu einen erheblichen Beitrag leisten; sie konnte auf den hier vorhandenen Ergebnissen – Material des SFB 100 in Saarbrücken sowie der EUROTRA-D-Begleitforschung – aufbauen.

Dabei war (und ist auch für die Zukunft) große Disziplin und eine intensive Kooperationsbereitschaft in der Erarbeitung und Ausführung gemeinsamer Konzepte nötig; in den mittlerweile vom Projektmanagement geschaffenen Verfahren zur konkurrierenden Forschung und zur Integration der erarbeiteten Ergebnisse kann nicht ein Staat völlig eigene Wege gehen. Es ist zu hoffen, daß sich neben dem Forschungsbereich auch der Organisationsbereich, der womöglich in gleicher Weise über Erfolg und Mißerfolg eines solchen multinationalen Forschungsprojektes entscheidet, weiter festigt.

In dem Beschluß des Rates der Europäischen Gemeinschaft vom 04.11.82 zur Schaffung eines Systems zur automatischen Übersetzung zwischen sieben Amtssprachen wurden am 26.11.86 aufgrund des Beitritts von Spanien und Portugal einige Änderungen vorgenommen. Dabei werden die Gesamtkosten des Projekts erhöht, das Projekt-Team der Kommission vergrößert sowie die Phasen II und III des Projekts um insgesamt 18 Monate verlängert. Nach diesem Zeitraum soll ein voroperationelles prototypisches System zur Verfügung stehen, das – obgleich fachspezifisch und lexikalisch eingeschränkt – als Grundlage einer industriellen Entwicklung dienen soll.

Das Ende der Phase der linguistischen Forschung ist nunmehr auf Juni 1988 terminiert, die Phase der Stabilisierung und Auswertung auf Mitte 1990.

Für die erweiterte Phase II bis Juni 1988 soll ein größerer Schwerpunkt als bisher auf die kooperativen Aufgaben gelegt werden, da sich bei diesen bisher umfangreiche Probleme gezeigt haben:

- Entwicklung moderner linguistischer Strategien
- Beschreibung der linguistischen Strategien im EUROTRA-Formalismus, Beurteilung der Adäquatheit der EUROTRA-Software sowie Vorschläge zu ihrer Erweiterung und Verbesserung (in enger Zusammenarbeit mit der ebenfalls in Saarbrücken angesiedelten Gruppe, die zur Zeit an Software-Tuning und -weiterentwicklung arbeitet, siehe Schütz 1987)

- Weitere Ausarbeitung von Vorschlägen für die linguistischen Spezifikationen und die allgemeine Methodologie
- Ermittlung und Definition universeller linguistischer Informationen (wie Zeitrelationen, Modalität, semantische Relationen und Kategorien)
- Weiterentwicklung der Transferkonzepte, auch in Zusammenarbeit mit der deutschen Begleitforschung zu EUROTRA-D
- Entwicklung und Implementierung von Transferregeln
- Erstellung von Transferlexika.

Bei den sprachspezifischen Aufgaben geht es um den Test und weiteren Ausbau der einzelnen Komponenten, die in ihrem Umfang mehr als bisher über das vorhandene Korpus hinausgehen und Repräsentativität erreichen sollen. Auch hierbei ist die theoretische Auseinandersetzung und Gegenüberstellung mit den grammatischen Modellen der Begleitforschungsprojekte wichtig. Im einzelnen handelt es sich um folgende Aufgaben:

- Test und Überarbeitung der linguistischen Spezifikationen für die deutsche Sprache

- Test und Überarbeitung der Analysemoduln, die es erlauben, bestimmte Repräsentationsebenen zu erreichen und miteinander zu verknüpfen (Morphologie, Oberflächen- und Tiefensyntax, Interface-Struktur)
- Test und Überarbeitung der Synthesemoduln, die den Übergang von einer höheren zu einer niedrigeren Repräsentationsebene ermöglichen (Interface-Struktur, Syntax, Morphologie, Stringdarstellung)
- Erstellung der Transferkomponenten zwischen den Interface-Repräsentationen der anderen Sprachen und der deutschen Interface-Repräsentation.

Über die spezifische Behandlung der Probleme des Deutschen sowie über die Vorschläge von EUROTRA-D für die semantischen Komponenten wird in den Beiträgen Schmidt 1987, Zelinsky-Wibbelt 1987a, Steiner 1987 berichtet.

Literatur

Die Literaturangaben für die Fachaufsätze zum Themenschwerpunkt Maschinelle Übersetzung finden sich zusammengestellt auf Seite 23.



DAS "LDV-FORUM" IM ABONNEMENT

Auch Nicht-Mitglieder der GLDV können jetzt das LDV-Forum regelmäßig beziehen. Dazu ist es notwendig, formlos ein Abonnement bei der Redaktion

Redaktion des LDV-Forum, Prof. Dr. G. Knorz, Fachhochschule Darmstadt, Fachbereich IuD, Schöfferstraße 8, D-6100 Darmstadt

zu bestellen und den Preis der Einzelzeitschriften jeweils für das laufende Jahr auf das Konto

Gesellschaft für Linguistische Datenverarbeitung / LDV-Forum, Kto.-Nr. 554 090, BLZ 508 501 50, Sparkasse Darmstadt

zu überweisen (Vorauskasse!). Für diese DM 30,- pro Jahr gehen Ihnen dann die beiden Hefte unmittelbar nach Erscheinen (ohne weitere Versandkosten) per Post zu. An der Tatsache, daß Sie den aktuellen Newsletter (wird gerade vorbereitet, siehe dazu das Editorial), den Rabatt auf die Reihe Linguistische Datenverarbeitung (OLMS) und vieles andere versäumen, sind Sie dann allerdings selbst schuld! Dazu müßten Sie dann doch Mitglied der GLDV werden (...was ich Ihnen in jedem Falle empfehlen kann).

Gerhard Knorz

STATUSEMINAR ZU EUROTRA-D UND BEGLEITFORSCHUNG IM FEBRUAR 1987
EUROTRA präsentiert erste Ergebnisse

Johann Haller
IAI/EUROTRA-D
Martin-Luther-Straße 14
D-6600 Saarbrücken

Uwe Reyle
Universität Stuttgart, Fachbereich Informatik
D-7000 Stuttgart

EUROTRA ist ein gemeinsames Projekt aller EG-Mitgliedsstaaten, in dem ein vorindustrieller Prototyp zur Maschinellen Übersetzung entwickelt werden soll. Alle 9 Sprachen der EG sollen sowohl Ausgangs- als auch Zielsprache sein können. In einem Statusseminar stellten die Projektgruppe EUROTRA-D am IAI, Saarbrücken sowie die von den Universitäten Stuttgart, Bielefeld und Berlin getragene Begleitforschung erste Forschungsergebnisse vor. In dem Bericht zu dieser Veranstaltung werden zunächst Konzept, Aufgaben und Abstimmung von EUROTRA-D-Forschungsgruppe und Begleitforschung skizziert. Es wird erklärt, für welche Teilaufgaben in dem EUROTRA-Stratifikationsansatz die Begleitforschung Resultate und Methoden der theoretischen Linguistik einbringen und hinsichtlich ihrer Anwendbarkeit überprüfen will. Der Verlauf des Statusseminars mit Vorträgen, Diskussion und Vorführungen wird kurz geschildert. Der Beitrag schließt mit Bemerkungen zum erreichten Stand und den Perspektiven der Maschinellen Übersetzung.

Vom 16. bis 18. Februar haben in einem gemeinsamen Statusseminar in Saarbrücken / Dudweiler die deutsche Forschungsgruppe EUROTRA-D (Universität des Saarlandes) und die Universitäten Stuttgart, Bielefeld und Berlin erste Ergebnisse ihrer Forschungen zur maschinellen Übersetzung vorgelegt und praktisch demonstriert.

EUROTRA ist ein gemeinsames Projekt aller 12 Mitgliedstaaten, das zunächst die Computerlinguistik und den Wissenstransfer innerhalb der EG fördern soll. Jedoch ist auch die Erstellung eines vorindustriellen Prototyps vorgesehen, der bis 1990 mit einem begrenzten Wortschatz Übersetzungen zu Texten, die in der Verwaltungssprache der EG abgefaßt sind, liefern soll. Dabei sollen alle 9 Sprachen der Gemeinschaft analysiert und generiert sowie Transfer in allen 72 möglichen Richtungen realisiert werden. Die Kommission der EG organisiert dieses Projekt, finanziert zentrale Forschungsarbeiten zu Linguistik und Software und beteiligt sich an den Kosten der nationalen Teilkomponenten. An der Finanzierung der deutschen Teilkomponente EUROTRA-D am Institut der Gesellschaft zur Förderung der Angewandten Informationswissenschaft ist neben dem Saarland auch das Bundesministerium für Forschung und Technologie mit 52% beteiligt.

Die Aufgabe der deutschen Forschungsgruppe in Saarbrücken ist die Erstellung der deutschen Analyse- und Syntheseprogramme und der Wörterbücher, die Lexika für den Transfer aus den anderen Amtssprachen der EG ins Deutsche werden in Zusammenarbeit mit der Universität Bonn aufgebaut. Als Grammatikformalismus ist für die deutsche Gruppe (wie für alle anderen Sprachen) ein zentral entwickeltes Konzept vorgegeben, das auf Unifikationsgrammatiken basiert und den Übersetzungsvorgang in kleine Schritte zerlegt. Zwischen Ausgangs- und Zieltext werden eine Reihe von linguistisch motivierten Ebenen angenommen, zwischen denen wieder Übersetzungen stattfinden. Für das Deutsche werden Repräsentationen auf morphologischer, konfigurationaler und relationaler Ebene erarbeitet, die jeweils auf aktuellen Forschungsergebnissen in Syntax und Seman-

tik beruhen. Gleichzeitig arbeiten deutsche Forscher an der konzeptionellen Weiterentwicklung des EUROTRA-Formalismus und an Problemen der effizienten Implementierung mit. Zur Unterstützung der EUROTRA-D-Forschung fördert das BMFT gleichzeitig unabhängige Begleitforschungs-Projekte an den Universitäten Berlin, Bielefeld und Stuttgart. Die Aufgabe dieser Projekte ist vornehmlich, die Resultate und Methoden der theoretischen Linguistik für die Anwendung auf Maschinelle Übersetzung zu überprüfen und nutzbar zu machen.

So werden z.B. an der Universität Stuttgart von theoretischen Linguisten umfangreiche Grammatiken des Deutschen und Französischen im Rahmen einer Unifikationsgrammatik, der Lexikalisch-Funktionalen Grammatik, erstellt. Parallel dazu wurden Compiler entwickelt, die diese (und beliebig andere in diesem Formalismus entwickelten) Grammatiken in Parser und Generatoren überführen. Parser erzeugen aus Sätzen syntaktisch-semantische Repräsentationen, Generatoren überführen solche Repräsentationen wiederum in sprachliche Ausdrücke. Ähnliche Forschungen werden an der Universität Berlin auf der Basis eines verwandten Grammatikformalismus, der Generalisierten Phrasenstrukturgrammatik, durchgeführt. Das Berliner Projekt versucht darüberhinaus auch Ansätze und Methoden der Künstlichen Intelligenz zu inkorporieren. An der Universität Bielefeld wurden die einzelnen Komponenten auf ihre Effizienz untersucht und entsprechend optimiert.

Für eine zweite, geplante Projektphase ist vorgesehen, entsprechende Transferautomaten zu konstruieren. Diese Transferautomaten sollen unter Verwendung von strukturellen Transferregeln und eines Transferlexikons die von den Parsern erstellten abstrakten Repräsentationen einer Quellsprache in entsprechende Repräsentationen einer Zielsprache übersetzen, aus denen ein Generator dann sprachliche Strukturen erzeugen kann. Eines der Hauptmerkmale der Begleitforschung ist eine interdisziplinäre Zusammenarbeit und klare Arbeitsteilung zwischen theoretischen Linguisten und Informatikern. Eine solche Konzeption wird vor allem ermöglicht durch die Verwendung von mathematisch

klar definierten Grammatikformalismen, die es erlauben, die verschiedenen konzeptuellen Ebenen eines sprachverarbeitenden Systems (Grammatische Beschreibung vs. Implementierung in Form von Parsern, Generatoren, Transferautomaten) zu trennen. Durch die Verwendung von einheitlichen vorgegebenen Formalismen können dabei Grammatiken und Software-Komponenten völlig unabhängig voneinander erstellt, modifiziert und optimiert werden. Ein weiteres wichtiges Merkmal dieser Forschungskonzeption ist der modulare Aufbau der Systeme. Da einzelne Module der Übersetzungssysteme (Grammatiken, Parser und Generatoren) anwendungsunabhängig sind, können sie in beliebige Sprachverarbeitende Systeme inkorporiert werden, die eine Schnittstelle zu natürlichen Sprachen benötigen.

Auf dem Statusseminar stellten sich alle Forschungsinstitutionen mit Vorträgen und Vorführungen einer kleinen Menge von Sachverständigen aus Ministerium, Universitäten und Industrie vor. Der englische Sprachwissenschaftler D. Arnold von der Universität Essex hielt den Einführungsvortrag über den EUROTRA-Formalismus, der zu lebhaften Diskussionen Anlaß gab. Der Rest des ersten Tages war dann den verschiedenen Vorführungen gewidmet, wobei die EUROTRA-D-Gruppe ein kleines Modell einer mehrsprachigen Übersetzung mit einfachen Sätzen sowie verschiedenen Analysestufen des Deutschen präsentierte; die Universität Stuttgart und die Universität Berlin zeigten Analyse und Generierung für Deutsch und Französisch, Bielefeld stellte optimierte Parser vor.

Am zweiten Tag wurden die Projekte in einer Reihe von Kurzvorträgen auch theoretisch dargestellt; ergänzend dazu hielt H. Uskoreit (Stanford/Stuttgart) einen Überblicksvortrag über die "unifikationsbasierten Grammatiken". Einige Vorträge zeigten Möglichkeiten der Zusammenar-

beit zwischen der Begleitforschung und EUROTRA-D auf; besonders auffällig war dabei ein Experiment, bei dem mit dem Programm SEMSYN zur Generierung deutscher Sprache Sätze aus abstrakten EUROTRA-Tiefenstrukturen erzeugt wurden.

In der langen und wechselvollen Geschichte der maschinellen Übersetzung ist zur Zeit also wieder eine positive Tendenz sichtbar; für alle beteiligten Forscher ist klar, daß mittelfristig keine zu großen Erwartungen an die Güte der Übersetzungen gestellt werden dürfen, und daß es noch sehr vieler Vorarbeit in Sprachwissenschaft und Informatik bedarf, um auch nur einfache technische Texte so übersetzen zu lassen, daß eine Korrektur einer solchen "Rohübersetzung" für den menschlichen Bearbeiter einfacher ist als eine direkte Übersetzung. Dabei existiert auch noch die Konkurrenz der "maschinenunterstützten Übersetzung", bei der der Übersetzer durch Textsysteme und Terminologiedatenbanken unterstützt wird, was in größeren Firmen zunehmend praktiziert wird.

All dieser Probleme waren sich die deutschen Forscher, die in Dudweiler versammelt waren, deutlich bewußt, ebenso wie der Notwendigkeit, Ergebnisse auszutauschen und gemeinsam zu arbeiten. Sie wollen doch alle auf dem Wege zu dem faszinierenden Ziel eines übersetzenden Computers vorankommen: ein Stück sprachlicher Fähigkeit des Menschen nachahmen zu können, gibt das Gefühl, einen Teil seines Geistes besser zu verstehen. Für die Gesellschaft (und die Förderer) ist der Nutzen klar: EUROTRA kann ein entscheidender Beitrag zum Abbau der Informationshindernisse im vielsprachigen Europa sein.

SYNTAKTISCHE ANALYSE DEUTSCHER SÄTZE IM EUROTRA-FORMALISMUS

Konfigurationale und relationale Struktur

Paul Schmidt
IAI/EUROTRA-D
Martin-Luther-Straße 14
D-6600 Saarbrücken

Der stratifikationale Ansatz des EUROTRA-Projektes zur maschinellen Übersetzung wird beschrieben: Zwischen Quell- und Zieltext werden verschiedene linguistisch motivierte Repräsentationsebenen angenommen, wodurch die Übersetzung in eine Abfolge kleinerer Übersetzungsteilschritte unterteilt wird. Die einzelnen Ebenen, die morphologische Struktur (EMS), die konfigurale Struktur (ECS), die relationale Struktur (ERS) und die Interface Struktur (IS) werden vorgestellt und das Konzept der Repräsentation und der Beziehungen zwischen benachbarten Ebenen werden kurz erläutert. Es wird an Beispielen gezeigt, wie die Regeln der deutschen Syntax auf der konfiguralen Ebene formuliert werden und wie ECS-Repräsentationen mittels t-Regeln auf die relationale Ebene abgebildet werden.

Probleme des Ansatzes, die in der Autonomie der ECS und in der Beschränktheit der t-Regeln liegen, werden angesprochen. Für den beschriebenen Ausschnitt der EUROTRA-D-Sprachbeschreibung liegt eine Implementierung vor.

I. Grundprinzipien

In diesem Beitrag werden die Grundprinzipien sowie einige Beispiele für syntaktische Analyserregeln eines Fragments des Deutschen gegeben. Dieses Fragment ist die Grundlage für die Implementierung der deutschen Syntaxkomponente in EUROTRA (vergleiche *Hallet/Reyle 1987*).

Die Repräsentationssprache ist der sogenannte C,A,T-Formalismus (*Arnold et al. 1986*). Er arbeitet mit den drei Elementen

Constructor
Atom
T-Rule

und ist eine formale, linguistisch motivierte Stratifikationstheorie des Übersetzungsvorganges. In der EUROTRA-Theorie werden zwei Grundannahmen gemacht:

- 1) Die Übersetzungsrelation wird nicht zwischen Quell- und Zieltext definiert, sondern muß in einzelne einfachere Übersetzungsschritte aufgeteilt werden.
- 2) Diese kleineren Übersetzungsschritte werden zwischen linguistisch motivierten Ebenen vorgenommen.

Aus diesen Annahmen folgen drei Grundeigenschaften des EUROTRA-Formalismus:

- 1) Das Übersetzungssystem hat stratifikationalen Charakter: die Übersetzungsrelation besteht zwischen $T_1 \dots R_1 \dots R_n \dots T_2$, wobei jede Ebene R eine künstliche Repräsentationssprache ist.
Zur Zeit gilt: $R_i = (EMS, ECS, ERS, IS)$
EMS = EUROTRA Morphological Structure
ECS = EUROTRA Configurational Structure
ERS = EUROTRA Relational Structure
IS = Interface structure

- 2) Die Repräsentationen sind durch Grammatiken bestimmt, die Generatoren genannt werden. Diese enthalten zwei Arten von Regeln:
b-rules bauen Repräsentationen auf
a-rules drücken Generalisierungen über Attribute aus.

a-rules werden auf von b-rules erzeugte Strukturen angewendet. Diese "Anwendung" bedeutet, daß eine Unifikation der a-rule mit der erstellten Struktur versucht wird. Wenn diese gelingt, wird die Struktur durch das Ergebnis der Unifikation ersetzt, im anderen Fall gibt es zwei Möglichkeiten:

- die Struktur überlebt ohne Änderung (etwa nach einer sogenannten "gentle"-a-rule-Anwendung)
- die Struktur wird gelöscht (nach einer "strikten" a-rule)

- 3) Die Beziehung zwischen zwei Ebenen wird durch einen "Translator" determiniert, der aus einer Menge von "t-rules" besteht. Diese t-rules haben zwei Charakteristika:

- sie sind ein-eindeutig, d.h. sie besitzen keine interne Strategie; ein Objekt der eine Ebene wird in ein einziges der anderen Ebene übersetzt
- sie sind kompositionell, d.h. die Übersetzung eines strukturierten Objekts ist eine Funktion der Übersetzung seiner Teile.

2. Formale Probleme

Die wichtigsten Probleme der EUROTRA-Theorie, die zur Zeit bestehen, sind folgende:

- die Autonomie der ECS, die zu einer erheblichen Übergenerierung führt
- die relative Inadäquatheit der t-rules für die Behandlung von Sprachen mit freier Wortstellung.

In der derzeitigen Version sieht der EUROTRA-Formalismus eine autonome ECS vor, d.h. Konstituentenstrukturen werden ohne Kontrolle durch Subkategorisierungsinformation erzeugt. Diese Art von Informationen wird erst auf ERS eingeführt, wo sie zur Reduzierung der Übergenerierung führt. Da dies keine prinzipiellen linguistischen Probleme erzeugt, ist dieses Problem nicht so bedeutend; lediglich für die Testarbeiten müssen längere Laufzeiten in Kauf genommen werden. Dagegen führt die Inad-

äquation der t-rule-Komponente zu prinzipiellen linguistischen Schwierigkeiten: da die t-rules nach dem Prinzip der expliziten Aufzählung arbeiten, wird diese Komponente bei Sprachen mit relativ freier Wortstellung sehr umfangreich und unübersichtlich. Alle linken Seiten der t-rule müssen aufgezählt werden, und es muß jeweils auf der rechten Seite notiert werden, an welchen Platz der Konstituent gehen soll. Dies bedeutet beispielweise, daß für die Beispiele in (1) zwei t-rules existieren müssen, eine für das topikalisierte Subjekt und eine für das topikalisierte direkte Objekt:



Dies bedeutet allgemein, daß alle möglichen Wortstellungskombinationen aufgezählt werden müssen, was eine große Anzahl von t-rules zur Folge hat. Dabei gehen alle Generalisierungen wieder verloren, die auf ECS bereits festgelegt sind; außerdem ist keine (ökonomische) Behandlung von long-distance-Phänomenen möglich. Zur Zeit werden vom EUROTRA-Zentralteam verschiedene Abänderungen am Formalismus diskutiert, die diese Probleme beheben sollen.

3. Konfigurationale Struktur (ECS)

Auf dieser Ebene selbst ist es möglich, Regeln zur deutschen Syntaxanalyse zu formulieren, die den aktuellen Forschungsstand widerspiegeln. Die Regeln folgen der einschlägigen Literatur (M. Reis, H. Haider, H. den Besten etc.): von einer kanonischen Wortstellung (Endstellung des finiten Verbs) werden alle Varianten durch "movement-rules" abgeleitet.



Der Verbalplex wird auf die folgende Weise behandelt:



Diese Regel erlaubt eine leichte Berechnung von Zeit und Diathese sowie eine Tilgung der Hilfsverben beim Übergang zur relationalen Ebene.

Dadurch ergibt sich der derzeitige Stand der Gesamtgrammatik wie folgt:



Den Variationsmöglichkeiten der freien Ergänzungen wird durch das optionale Auftreten von adverbialen oder propositionalen Elementen zwischen den Komplementen Rechnung getragen. Obwohl diese Lösung nur vorläufigen Charakter hat (siehe oben), können damit die meisten Strukturen des EUROTRA-Korpus, eines Textes aus der EG-Verwaltungssprache, bearbeitet werden.

4. Relationale Struktur (ERS)

Die relationale oder Dependenz-Struktur, die in etwa den f-Strukturen der Unifikationsgrammatiken entspricht, wird durch die Eigenschaft der lexikalischen Einheiten definiert, andere Elemente zu binden. Diese Eigenschaft wird "Valenz" genannt. Die Definition dieser Ebene für das Deutsche basiert auf Arbeiten des IDS Mannheim und ist detailliert in Schmidt 1986 und 1987 beschrieben.

Für die Beschreibung der relationalen Struktur eines Satzes wird im Prinzip eine einzige Regel benötigt:



Ein Satz besteht aus dem regierenden Verb und einem bestimmten Satz an Komplementen, die optional einmal vorhanden sind, sowie einer beliebigen Zahl von freien Ergänzungen, was wieder durch den Kleene-Star ausgedrückt wird. Welche Features im aktuellen Fall mit Werten gefüllt werden, bestimmt der Eintrag für das Verb:

- (5) gov, (cat = vrb, lu = kommen,
comp0 = (cat = np, case = nom)),
comp4 = (cat = pp, prop = von))

Die Werte im gov-Argument werden durch Unifikation automatisch in die comp-Features des Satzes übertragen. Sogenannte "gentle rules" führen einen Default-Wert "no" ein für alle anderen comp-Features. Auf diese Weise wird sichergestellt, daß bei der Überführung auf die relationale Ebene nur die Syntagmen in die Slots des Verbs eingesetzt werden, die die entsprechenden Eigenschaften aufweisen.

5. Die Beziehung zwischen ECS und ERS

Wie bereits erwähnt, sind die Überführungsregeln zwischen diesen beiden Ebenen, die ja für Analyse und Synthese in gleicher Weise verwendet werden, ein komplexes Problem. Dies gilt speziell für Sprachen mit relativ freier Wortstellung, in denen keine konfigurationale Behandlung von Komplementen möglich ist. Auch für die Verbalgruppe sind vier Arten von T-Regeln nötig, die durch zwei paarweise Merkmale charakterisiert sind:

- die Stellung des Verbs (1,2, final)
- das Vollverb bzw. Hilfsverb

Die vier Regelarten sind also:

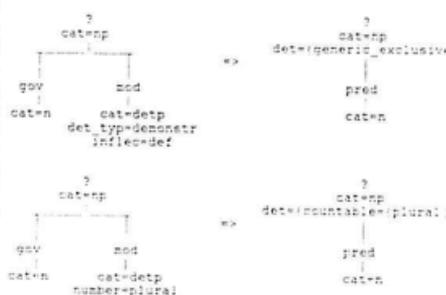
- V/1,2 Vollverb
- V/1,2 Hilfsverb
- V/final Vollverb
- V/final Hilfsverb

Die A-Regeln werden auf die gebauten Strukturen angewandt.

Im Übersetzungsprozess von der syntaktischen Ebene in die semantische Ebene werden Generalisierungen über Attribute beim Übersetzen von Werten ausgedrückt. Partikel (valenzgebundene Präpositionen und Konjunktionen) werden im Übersetzungsprozess auf die semantische Ebene getilgt, wobei ihr Wortlaut als Wert des LU-Attributs in die Dekoration des Komplements - der NP bzw. des Satzes - kopiert wird und zusätzlich in semantische Merkmale übersetzt wird. Mit diesen beiden Informationen repräsentieren wir zur Zeit die Bedeutung der Präposition und der Konjunktion. Wir verdeutlichen diesen Übersetzungsprozess an den folgenden Beispielen (in vereinfachter Form):



Auch die Determinanten und Quantifikatoren werden bei der Übersetzung in die Interface Struktur getilgt und in semantische Merkmale übersetzt:



Literatur

Die Literaturangaben für die Fachaufsätze zum Themenschwerpunkt Maschinelle Übersetzung finden sich zusammengefaßt auf Seite 23.

GRUNDPRINZIPIEN UND VORSCHLÄGE FÜR EINE SEMANTISCHE BESCHREIBUNG VON ARGUMENTSTRUKTUREN Überlegungen zur EUROTRA Interface-Struktur

Erich Steiner

IAI/EUROTRA-D
Martin-Luther-Straße 14
D-6600 Saarbrücken

Zwischen Quell- und Zielsprache werden in der Konzeption von EUROTRA, dem europäischen Projekt zur maschinellen Übersetzung für alle EG-Sprachpaare, verschiedene linguistische Ebenen angenommen. Damit wird der gesamte Übersetzungsprozess in eine Folge von Teilübersetzungen zerlegt. Zwischen Analyse und Generierungsphase findet der Transfer auf der Ebene der *Interface-Struktur* (IS) statt. Diese IS ist keine *interlingua*, sondern eine Repräsentation, die möglichst wenig Transferoperationen erfordert. Der vorliegende Beitrag diskutiert die *semantischen Relationen* (Prädikat-Argument-Beziehungen) als eine der wichtigsten Teiltheorien der IS. Für diesen Bereich gibt es noch keinen von allen EUROTRA-Gruppen akzeptierten Ansatz. Zunächst werden Anforderungen an und Funktionen von semantischen Relationen im Rahmen eines Übersetzungssystems umrissen. Anschließend wird das Konzept der EUROTRA-D-Forschungsgruppe, das sich auf die Systemic Functional Grammar abstützt, vorgestellt und begründet. Der Stand der Realisierung und die weitere Entwicklung werden skizziert.

1. Grundprinzipien der IS

Semantische Relationen (SRs) sind die semantischen Beziehungen zwischen den "governors" (govs) einer Konstruktion und den von ihnen abhängigen Gliedern, den Komplementen (comps). Semantisch interpretiert handelt es sich um die Beziehungen zwischen Prädikaten und Argumenten (preds und args). Inwieweit hierher auch die semantischen Relationen zwischen govs und modifiers gehören, ist eine Frage, mit der wir uns hier nicht beschäftigen wollen. Die verschiedenen Standardtheorien aus dem Bereich der Linguistik, die hier zugrundeliegen, sind recht gut bekannt und sollen auch hier nicht weiter diskutiert werden (Fillmore, Halliday, Longacre, Starosta, Gruber, Jackendoff, Bresnan). Der im folgenden beschriebene EUROTRA-D-Ansatz ist u.a. dokumentiert in Steiner 1985, Steiner 1986, Steiner 1987a, Steiner 1987b, Steiner et al. 1987.

2. Die Funktion von SRs in einem multilingualen MT-Projekt

Semantische Relationen haben eine Reihe von Funktionen, die sie in einem multilingualen MT-Projekt als wesentlich erscheinen lassen:

3.1. Die Repräsentation von Ambiguitäten

Innerhalb von EUROTRA ist die *Interface Struktur* definiert als Ebene minimalen Transfers zwischen einer Ausgangssprache und einer Zielsprache. Die *Interface Struktur* (IS) ist somit keine *Interlingua* - eine grundlegende Festlegung, die immer wieder in der Diskussion um EUROTRA übersehen wird.

Die Theorie der IS umfaßt eine Reihe von Teiltheorien: MODALITY, TIME, SEMANTIC FEATURES, SEMANTIC RELATIONS und möglicherweise weitere. Wir werden uns hier mit der Theorie der Semantischen Relationen beschäftigen.

2. Semantische Relationen in der theoretischen Linguistik

Durch semantische Relationen lassen sich grundlegende Klassen von lexikalischen Ambiguitäten darstellen, die mit syntaktischen Repräsentationen alleine nicht darstellbar sind. Da solche Fälle in der Literatur hinreichend dokumentiert sind, wollen wir uns hier das Anführen von Beispielen ersparen.

3.2. Vermeidung von komplexem strukturellem Transfer

Eine Repräsentation, die SRs beinhaltet, ist ein wesentliches Mittel, klassische Arten von komplexem strukturellem Transfer zu vermeiden. Es handelt sich dabei um Erscheinungen wie Thematisierung, Passivierung im besonderen, Reflexivierung, Kausativierung, verschiedene Arten von "raising". Da aus offensichtlichen Gründen komplexer struktureller Transfer in einem multilingualen MT-System sehr arbeits- und kostenintensiv ist, handelt es sich hier bei um einen wichtigen Gesichtspunkt.

3.3. Steuerung des lexikalischen Transfers

SRs können bis zu einem gewissen Grad zur Steuerung von automatischem lexikalischem Transfer eingesetzt werden. Untersuchungen zum Transfer Deutsch-Französisch im Rahmen der Zusammenarbeit zwischen EUROTRA-D Saarbrücken und der EUROTRA-D-Begleitforschung Stuttgart sind vorgelegt worden u.a. in Eckert/Heud 1987.

3.4. Behandlung anaphorischer Referenz

SRs sind im Rahmen einer entsprechenden Konzeption Bestandteil der logischen Form von Sätzen und können in diesem Rahmen auch eingesetzt werden bei der Behandlung anaphorischer Referenz.

4. Der Ansatz von EUROTRA-D

Der Ansatz von EUROTRA-D berücksichtigt zunächst einmal die in Abschnitt 2 genannten Funktionen von SRs. Des weiteren ist davon auszugehen, daß das Ziel von EUROTRA für IS eine Ebene minimalen Transfers ist, nicht aber eine Form von *Interlingua*. Als linguistische Modelle kommen von daher eher funktionalistische Theorien in Frage als solche, die a priori einen starken Anspruch auf Universalität erheben. Es erscheint ebenfalls geboten, sich weitgehend auf relativ langlebige und gut ausgearbeitete Ansätze zu stützen. Kurzlebige Theorien werden rezipiert, können aber in einem Projekt von einer Größenordnung wie EUROTRA nicht die Grundlage in irgendeinem Bereich bilden. Letztendlich muß ein SR-System für ein Projekt wie EUROTRA größtmögliche Konsistenz unter einer großen Anzahl von Kodierern/-innen verschiedener Nationalität und mit verschiedenem Hintergrund gewährle-

sten, ähnlich wie das zu entwickelnde System für die Kodierung semantischer Merkmale von lexikalischen Einheiten, das in *Zelinsky-Wibbelt 1987b* beschrieben ist. Aus den geschilderten Überlegungen heraus entschloß man sich in EUROTRA-D, sich in der SR-Theorie weitgehend, wenn natürlich auch nicht ausschließlich, auf die "Systemic Functional Grammar" zu stützen (vgl. *Halliday 1985, Fawcett 1980, Winograd 1983, Butler 1985, Steiner 1987*).

Der Ansatz arbeitet mit einer kleinen Anzahl von Prozeltypen (4 Grundtypen, die in 17 Untertypen unterteilt sind), die durch eine sie charakterisierende semantische Rolle festgelegt sind. Diese Schlüsselrollen sind durch ein eindeutiges syntaktisches Realisierungspotential gekennzeichnet. Jeder der 4 Grundtypen ist als Eingangsbeschränkung für eine Untermenge von semantischen Rollen definiert. Die Rollensätze der 4 Grundtypen sind fast vollständig disjunkt. Bei der Lexikonkodierung wird so ein hohes Maß an Konsistenz zwischen verschiedenen Kodierern/-innen gewährleistet. Gegenwartig sind die SFs in verschiedenen Versionen eines lauffähigen Systems implementiert, wobei systematisch bisher SRs für Verben sowie SRs für Nomen implementiert wurden.

5. Die weitere Entwicklung

Der geschilderte Ansatz für SRs ist innerhalb von EUROTRA in der Diskussion, ist jedoch zur Zeit nicht der in allen Sprachgruppen vertretene Ansatz. Es ist die Auffassung von EUROTRA-D, daß unabhängig von wissenschaftlichen Richtungskämpfer die in Abschnitt 2 angesprochenen Grundanforderungen an jede SR-Theorie von jedem vorgeschlagene Ansatz eingehend werden müssen. Diese Grundanforderungen bilden eine Art von gemeinsamem Bezugspunkt für die Diskussion innerhalb des Gesamtprojektes. Weiterhin müssen in der Zukunft die verschiedenen Arten der Interaktion der SR-Theorie mit der Theorie der "Semantic Features" und den Theorien für Modality und Time geklärt werden. Damit sind unweigerlich Beiträge zu leisten zur Praxis der Implementierung solcher Systeme, aber auch zur aktuellen Diskussion innerhalb der beteiligten wissenschaftlichen Disziplinen, wobei natürlich in erster Linie die Linguistik angesprochen ist.

Literatur

Die Literaturangaben finden sich für alle Fachbeiträge zum Themenschwerpunkt *Maschinelle Übersetzung* zusammengefaßt auf Seite 23.

PROJEKT SEMSYN NEW PUBLICATIONS

Walter Klett

GLITEX Ein System zur Verbalisierung geometrischer Konstruktionen

(GLITEX - a system for verbalizing geometric constructions, in German)

GLITEX is an application of the SEMTEX text generator for German. The text generator is combined with a tool for interactively creating geometric constructions. The latter offers formal commands for manipulating (or creating, naming and deleting) existing geometric objects of Euclidean geometry. The generator is used to produce descriptive texts in German, related to the geometric construction, descriptions of the geometric objects involved, descriptions of the sequence of steps during a construction. SEMTEX context handling mechanisms have been extended for GLITEX.

English words longer restricted to adjectives. For repetitive operations verb and subject will be added in subsequent sentences.

The distinction between known information and new one is exploited to decide on consistent ordering. The constituent referring to the known object is topicalized, i.e. put in front of the sentence.

The system allows for more ways to refer to objects introduced to the text: pronouns, textual deixis using demonstrative pronouns, names. The choice between these variants is done automatically.

GLITEX is implemented in RETALIB and runs on IBMPC/compat-ibp machines.

Thomas Künze

The generation system of the SEMSYN project

Towards a task independent generator for German

We report on our experiences from the implementation of the SEMSYN generator, a system generating German texts from semantic representations, and its application to a variety of different areas: input structures and generation tasks. In its initial version, the SEMSYN generator was used within a Japanese-German MT project, where it produced German equivalents to Japanese titles from scientific papers. Being carefully designed in abstract oriented style (and implemented with the HAZON system) the system proved to be easily adaptable to other semantic representations, e.g. output from a MU's Universal Parser, and extendible to other generation tasks, generating German news stories, generating descriptive texts to geometric constructions.

Thomas Künze

The automated news agency: SEMTEX - a text generator for German

As a by-product of the Japanese-German machine translation project SEMSYN the SEMTEX text generator for German has been implemented (in RETALIB for SYMBALIOS, see Nakano). SEMTEX's first application has been to generate newspaper stories about submarine development. Starting point for the newspaper application are just the data from the monthly submarine report (lots of unprocessed open jobs ...). Automatically, text planner takes three days and does a relevant previous months checks for changes and significant developments, simulates possible arguments of various political speakers on these developments and finally creates a representation for the intended text as an ordered list of frame descriptions. SEMTEX then converts this list into a newspaper story in German using an extended version of the generator of the SEMSYN project. The extensions for SEMTEX include:

- Building up a representation for the content during the utterance of successive sentences that allows for avoiding repetitions in wording
- avoiding redundancy in information still valid
- nominalization and other types of references
- Grammatical tense is dynamically derived by checking the temporal information from the conceptual representations and relating it to the time of speech and the time period focussed by the story.
- When simulating arguments the text planner uses abstract rhetorical schemata. The generator is enriched with knowledge about various ways to express such rhetorical structures as German text texts.

For copies of these papers please write to: Frau Maria ...
Projekt SEMSYN, Institut für Informatik, Azenbergstr. 12, D-702
Stuttgart 2. (West Germany) or e-mail to: aemayn@inf.fh-stuip.

LEXIKALISCHE DISAMBIGUIERUNG MIT SEMANTISCHEN MERKMALEN Überlegungen zur EUROTRA Interface-Struktur

Cornelia Zelinsky-Wibbelt

IAI/EUROTRA-D
Martin-Luther-Straße 14
D-6600 Saarbrücken

Die Interface-Struktur ist diejenige linguistische Ebene des EUROTRA-Ansatzes, auf der der Transfer von der maschinell zu übersetzenden Quellsprache in die Zielsprache stattfindet. Der vorliegende Beitrag skizziert ein Konzept zur semantischen Vereindeutigung lexikalischer Einheiten im Lexikon mittels kognitiv motivierter *semantischer Merkmale*. Der vorgestellte Ansatz basiert auf den *Cognitive Grammars* und soll ein Beitrag zu dem für maschinelle Übersetzungssysteme wesentlichen Problem der Konsistenz von Einträgen in großen Wörterbüchern sein. Die Metasprache, in die die semantischen Merkmale eingebettet sind, wird an einem Beispiel illustriert.

Auf der bereits in *Zelinsky-Wibbelt 1987a* beschriebenen Ebene der Interface-Struktur (IS) sollen Sätze möglichst "euroversell" repräsentiert werden, was bedeutet, daß die Transferschritte zwischen den Interface-Strukturen der verschiedenen Sprachen möglichst klein werden sollen. Dabei werden unter anderem verschiedene Lesarten einer lexikalischen Einheit durch "euroverselle" semantische Merkmale identifiziert. Für die semantische Vereindeutigung lexikalischer Einheiten entwickeln wir ein vom Lexikographen interaktiv zu nutzendes Generierungssystem. Mit diesem soll für jede Lesart diejenige Liste von semantischen Merkmalen erzeugt werden, die diese Lesart von anderen Lesarten abgrenzt. Diese Liste ist Teil der lexikalischen Information.

Die semantischen Merkmale basieren auf **grundlegenden kognitiven Prinzipien**, mit denen der Mensch seine Umgebung logisch im Hinblick auf allgemeine Übereinstimmungen und Unterschiede strukturiert. Diese semantischen Merkmale sind die Terme einer semantischen Metasprache. Diese verfolgt folgende Ziele im Hinblick auf eine weitestgehende Operationalisierung des semantischen Raumes, der tatsächlich ein Kontinuum ist:

- (1) Das Ziel der Objektivität
- (2) Das Ziel der psychologischen Realität
- (3) Das Ziel der Generalisierung

Die drei Ziele bedingen sich in der gegebenen Reihenfolge. Sie bilden auch das Leitmotiv der *Cognitive Grammar* (vgl. *Langacker 1982*, S. 23-25,35). Unter Grundprinzipien, also auch die Entwicklung der Definitionen der Merkmale, basieren auf diesem Ansatz, wir gehen davon aus, daß eine Erfüllung dieser Bedingungen zu einem System führt, bei dem eine genügend hohe Konsistenz in der Kodierung lexikalischer Einträge erreicht werden kann, was bei allen existierenden Systemen als größtes Problem beim Aufbau umfangreicher Wörterbücher berichtet wird.

Das Formulieren von Generalisierungen über die Kompetenz der Sprachbenutzer impliziert die Hypothese von der "Euroversalität" dieser Metasprache.

Als Teilziel von (3) bleibt hier noch das Verfahren zu nennen, mit dem Generalisierungen ermöglicht werden. Auch hierin können wir uns auf die *Cognitive Grammar* beziehen. Für diese sind hierarchische Schemata unter anderem ein adäquates Erklärungsmittel dafür, daß Ähnlichkeit zwischen Konzepten abhängig ist von der Hierarchiestufe, auf der sie verglichen werden. Diese Erklärungskraft muß auch unsere semantische Metasprache besitzen, um die Bedeutungen lexikalischer Einheiten in analoger Weise zu vergleichen.

Zur Beschreibung der vier Wortklassen erhalten wir vier Inventare semantischer Merkmale: "Situations" für Verben, "Entities" für Substantive, "Properties" für Adjektive und "Specifiers" für Adverbien. In jedem Set bilden jeweils Teilmenge wohlgeformte Ausdrücke. Ihre Wohlgeformtheit ist durch ein semantisches Regelsystem definiert, in dem die Merkmale durch die Relationen der Hierarchie, der Disjunktion und der Konjunktion verbunden sind. Zusätzliche Koinkurrenzbeschränkungen gelten für linguistisch nicht realisierte konjunktiv verbundene Merkmale (für eine genaue Erläuterung des Grammatikformalismus siehe *Zelinsky-Wibbelt 1988*).

Mit unserem semantischen Regelsystem können wir zum Beispiel für vier Lesarten der lexikalischen Einheit **Absatz** die folgenden hierarchischen Strukturen semantischer Merkmale generieren (die eckige Klammer entspricht in der Horizontalen der Relation der Hierarchie, in der Vertikalen verbindet sie Merkmale, die in konjunktiver Relation zueinander stehen).

1. Absatz als *Entity* (Substantiv, Nomen)

[ABSATZ [EINHEIT [MERKAL [DISJUNKTION [SITUATION [PROPERTIE [SPECIFIER]]]]]]
[SITUATION]
[PROPERTIE]
[SPECIFIER]

2. Absatz als *Property* (Adjektiv)

[ABSATZ [EINHEIT [MERKAL [DISJUNKTION [SITUATION [ENTITÄT]]]]]]
[ENTITÄT]
[PROPERTIE]
[SPECIFIER]

3. Der Absatz als *Situation* (Verb)

[ABSATZ [EINHEIT [MERKAL [DISJUNKTION [SITUATION [SITUATION]]]]]]
[SITUATION]
[PROPERTIE]
[SPECIFIER]

4. Der Absatz als *Specification* (Adverb)

[ABSATZ [EINHEIT [MERKAL [DISJUNKTION [SITUATION [ENTITÄT]]]]]]
[ENTITÄT]
[PROPERTIE]
[SPECIFIER]

Die Merkmale "individutive" und "partitive" sind Teil eines Merkmalssets, das auch für die funktionale Representation eines bestimmten Typs der Determination notwendig ist, der von *Lyber 1986* als "classifier construction" bezeichnet

wird. Oberflächenstrukturell wird dieser Typ folgendermaßen ausgedrückt:

[ZENTRIEREN] [RP + DF + defNP]

Diesem Ausdruck entspricht das folgende semantische Schema:

[ZENTRIEREN] [x[+]₁][+](x[+]₁)[+]
x[+]₂[+]

Diese Merkmale, die zu den Merkmalen gehören, die "classifier constructions" strukturell identifizieren sollen, sind Teil eines umfassenderen

Merkmalsets, das zur Zeit im Rahmen eines Transfer-Experiments zur Determination entworfen und getestet wird.

Literatur

Die Literaturangaben finden sich für alle Fachbeiträge zum Themenschwerpunkt *Maschinelle Übersetzung* zusammengefaßt auf Seite 23.

ANZEIGE

WordCruncher

FEATURES OF WORDCRUNCHER *

Preindexing.

- With WordCruncher's *IndexETC*, a preindex is created for a specific text. It is this electronic "road map" that makes possible sophisticated search combinations of large documents with amazing speed.

- Existing electronic texts require very little preparation before preindexing with *IndexETC*. Texts not already stored in electronic form may be entered via the keyboard or an optical scanner.

- WordCruncher allows the preindexing of documents in many foreign languages. Users are able to define the proper sorting sequences of characters, such as umlauts in German, accents in French, and the "ch" in Spanish. Texts may be shown from right to left for languages such as Hebrew. Preindexed files can be linked together to create a larger file.

Retrieving

- With WordCruncher's *ViewETC*, text material can be retrieved according to a set of user-defined parameters from very large documents.

- References can be looked up using a word or phrase, a list of words, two or more words in a defined context, a substring (such as a suffix or stem), or any logical combination of these. Supplementary dictionaries and thesauruses, supplied by the user and preindexed by WordCruncher, offer additional retrieval clout.

- Selected references can be displayed within windows of modifiable size which show the reference in its context. The Expanded Window displays an entire screen of text surrounding the reference, and you may page forward or backward through the document.

- To make it all easy, help displays are available throughout.

Manipulating.

- Once the desired data has been retrieved from a text, it may be manipulated in any number of ways. For instance, *ViewETC* can create printable book-style indexes, as well as keyword-in-context (KWIC) concordances of words or phrases.

- Retrieved information can be printed immediately, saved as a DOS text file, or placed in a temporary file such as WordPerfect Library's clipboard function provides. WordPerfect Library allows the user to transfer text from *ViewETC* to WordPerfect without exiting either program.

Marketing Representative

Institut für angewandte
Kommunikations- und
Sprachforschung e.V.

Poppelsdorfer Allee 47
D-5300 BONN 1
Tel.: 02 28 73-56 45 73-56 20
EARN: DBNRHRZ1.UPK000

*Formerly known as BYU Concordance

EUROTRA-Software-Entwicklung

Jörg Schütz

IAI/EUROTRA-D
Martin-Luther-Straße 14
D-6600 Saarbrücken

Zur Formulierung linguistischen Wissens steht in dem europäischen Projekt zur maschinellen Übersetzung EUROTRA die Schnittstelle einer virtuellen Maschine zur Verfügung, die in zwei Varianten, jeweils in PROLOG implementiert, vorliegt: Die *RS-Implementierung* ("runnable specifications") dient als Instrument der Weiterentwicklung der EUROTRA-Theorie. Die *TI-Implementierung* ("Turbo Implementierung") soll eine effiziente und benutzerfreundliche Entwicklungsumgebung für den Linguisten bereitstellen. Es wird beschrieben, inwieweit beide Implementierungen über den gegenwärtigen Stand der ausformulierten EUROTRA-Theorie hinausgehen, um den Prozess der maschinellen Übersetzung vollständig beschreibbar zu machen. Die bisherigen und geplanten Software-Entwicklungsarbeiten (und der bestehende Bedarf an konzeptionellen Überlegungen) sowie erste Ergebnisse der Verwendung verschiedener kompilierter PROLOG-Versionen werden dargestellt.

1. Gegenwärtiger Stand

Die EUROTRA (ET) μ_2 Theorie, das $\langle C, A \rangle, T$ -Framework (Arnold 1986), beschreibt die Semantik der virtuellen ET-Maschine (VM), die auf einem Rechner realisiert werden muß. Zu dieser Implementierung gehört dann auch die Syntax für die Sprache der VM (vergleichbar einer Programmiersprache, allerdings mit dem Anspruch, linguistisches Wissen adäquat handbar zu machen). Zur Zeit ist diese Sprache identisch mit der Benutzersprache (die Sprache, in der der Linguist sein Wissen ausdrücken soll).

Die VM existiert gegenwärtig in zwei Implementierungsvarianten, die sich aus den ersten prototypischen Implementierungen (V0 bis V5) herausgeschält haben:

- "Runnable specifications" (RS)
- "Turbo-Implementierung" (TI)

Bei den RS handelt es sich um Programmfragmente, mit denen gezeigt werden soll, daß die in der μ_2 -Theorie definierte Semantik implementierbar ist (d.h. auf einem Rechner realisierbar). Das Laufzeitverhalten spielt dabei eine untergeordnete Rolle, ebenso eine Benutzeroberfläche für den Benutzer. Die RS waren (und sind) notwendig, da die μ_2 -Theorie ständigen Veränderungen unterlag (und noch unterliegt), die auf Implementierbarkeit hin untersucht werden müssen. Änderungen sind einerseits durch die Tests und Evaluierungen auf μ_1 -Ebene (Linguistik) erforderlich, andererseits aus μ_2 -theoretischen Überlegungen heraus.

Die Semantik der TI ist identisch mit der der RS. Diese Implementierung soll ein günstigeres Laufzeitverhalten zeigen und eine benutzerfreundliche Benutzerumgebung bereitstellen (insbesondere für die Arbeiten auf μ_1 und μ_0 -Ebene), d.h. ein Syntaxchecker für die Benutzersprache mit aussagekräftigen Fehlerdiagnosen, eine Kommandosprache für den Dialog mit der VM, Trace-Mechanismen für die Testphasen der Grammatikimplementierungen und die Möglichkeit einer günstigen (platzeffizienten) Konservierung erzeugter Objekte.

Als Programmiersprache wurde Prolog (hier: G-Prolog) gewählt; die ET-Maschinen sind Micro-Computer (Vax, Micro-Vax, Sun-Familie und Bull) mit dem Betriebssystem Unix 4.2 bsd (resp. System V). Die Implementierungen (RS und TI) entsprechen zur Zeit aus technischen Gründen nicht exakt der μ_2 -Theorie, d.h. sie besitzen Eigenschaften, die dort nicht vorgesehen resp. definiert sind. Für die

konfigurale Ebene (ECS) wird ein Earley-ähnlicher Chart-Parser und ein Lexikon eingesetzt. Ohne diese Komponenten wäre eine Satz(Text-)analyse unmöglich.

Im Lexikon werden in einem Deklarationsstil Attribute und deren mögliche Werte, Abhängigkeiten von Attributen, sowie komplexe Attribute definiert. Komplexe Attribute dienen der Beschreibung linguistisch sinnvoller Attributzusammengehörigkeiten (z.B. ein Attribut "agreement={gender, number, person}"), andererseits erhöhen sie aber auch die formale Mächtigkeit des Formalismus (die Definition rekursiver Attribute erlaubt die Analysekontextsensitiver Sprachen). Komplexe Attribute sind ebenfalls nicht in der μ_2 -Theorie vorgesehen, erscheinen jedoch notwendig zur Analyse bestimmter sprachlicher Phänomene, die nicht kontextfrei formulierbar sind. Auf diesen lexikalischen Deklarationsstil folgt ein Datenteil, in dem die Merkmalsbeschreibungen zu Wort-Vollformen (Schlüssel für das Lexikon) verzeichnet sind (es erscheint zur Zeit unmöglich, eine Morphologie innerhalb der μ_2 -Theorie zu formulieren, siehe dazu die Diskussion zum sogenannten ET base level - EBL). Die restlichen Ebenen (ERS, IS) werden gemäß der μ_2 -Theorie mittels Translatoren und Generatoren aufgebaut.

Da TI auf V1 bis V5 gestützt ist, mußten zunächst Fehler aufgedeckt und eliminiert werden (auch als "streamlining" bezeichnet). Die rudimentär vorhandene Benutzeroberfläche wurde wesentlich verbessert, insbesondere durch die Einführung von Fehlerdiagnosen auf allen Ebenen. Ein Syntax-Checker ist implementiert, allerdings konnte eine volle Mächtigkeit nicht erreicht werden, da Restriktionen durch Prolog vorhanden sind. Es werden u.a. zyklische Regeln und ϵ (epsilon)-Regeln (leere rechte Seite) erkannt, und dem Benutzer gemeldet. Der Anwendbarkeitstest von a-Regeln und t-a-Regeln wurde effizienter realisiert. Die graphische Darstellung der einzelnen Objekte wurde geringfügig modifiziert (die lu-Werte erscheinen in den Blättern). Die Turbo-Arbeiten wurden im wesentlichen von IAI-Mitarbeitern durchgeführt, unterstützt wurden die Arbeiten zusätzlich durch jeweils einen Informatiker aus Dänemark und Griechenland (je für etwa 10 Tage).

Außer den genannten Arbeiten wurden Tests mit kompilierten Prologs durchgeführt (IF, Mu, Quintus und YAP in der Turbo-Gruppe selbst, BIM in Leuven). ECRC-Prolog konnte nur theoretisch eva-

liert werden (keine Auslieferung vor Sommer 87, da neuer Release). Die geleisteten Implementierungsarbeiten konnten in der kurzen Zeitspanne nicht ganz den gewünschten Erfolg, eine wesentliche Reduzierung der Laufzeiten, erbringen. Die Implementierung eines BUP-ähnlichen Parsers (Matsumoto) ist noch nicht abgeschlossen, von dessen Realisierung erwarten wir uns eine starke Herabsetzung der Laufzeit (erste Tests zeigen den Faktor 10). Allerdings haben die diversen Testzyklen auch gezeigt, daß das Laufzeitverhalten durch die hohe Komplexität der Sprachkonstrukte forciert wird, z.B. weniger Optionalität in den Regeln. Dies geht aber wieder zu Lasten größerer Regelanzahl. Wir sind zudem der Meinung, daß die RS in ihrer derzeitigen Konzeption nicht ausreichen; das Aufzeigen von Implementierbarkeit alleine sagt noch nichts über die Komplexität der μ_2 -Konstrukte aus. Eine Komplexitätsanalyse erscheint uns unabdingbar, diese soll auch die sog. "Computational Limits" aufzeigen.

2. Die weitere ET-Software-Entwicklung

Die "Liaison Group" hat in ihrer Januar-Sitzung beschlossen, andere Wege der Definition der μ_2 -Theorie und ihrer Implementierung zu beschreiten. Es wurde eine sog. "Engineering Group" gebildet, die auf der Basis der bisherigen μ_2 -Theorie (Ref.Man. 2.0) diese um fehlende Konstrukte erweitert und deren Implementierungsarchitektur definiert. Eine neue Implementierungsgruppe auf der Basis der "Turbo-Gruppe" soll gebildet werden.

Zur Zeit ist die Turbo-Gruppe in die Engineering-Gruppe integriert. Ein erstes Konzept des neuen Framework ist gerade abgeschlossen. Insbesondere wird nun den Generatoren, die die Ebenen ECS, ERS und IS erzeugen, mehr Mächtigkeit zugebracht (wir unterscheiden unterschiedliche Anwendbarkeitsmodi), wogegen die Translatoren zwischen den Ebenen "einfacher" werden (keine komplexe Strukturen).

Ein Vorschlag für die Verwertung des lexikalischen und grammatikalischen Wissens ist die Integration eines kommerziellen Datenbanksystems in die Gesamtsystemarchitektur.

Von EUROTRA-D werden Untersuchungen zu Datenstrukturen und Algorithmen (basierend auf der neuen Framework-Definition) durchgeführt. In Kooperation mit Belgien und Dänemark werden Konzepte zum ET-Lexikon und zur Benutzersprache ausgearbeitet (Varile 1987).

Literatur

Die Literaturangaben für die Fachaufsätze zum Themenschwerpunkt Maschinelle Übersetzung finden sich zusammengefaßt auf Seite 23.

ANZEIGE

MProlog

Der sichere Weg zur intelligenten Problemlösung.

Beim MProlog werden hochentwickelte Algorithmen der Expertentechnik auf den Rechner übertragen, um die Investition zu bewahren.

In weltweit über 800 Installationen wird das MProlog-System für die Intelligenz von Wert eingesetzt.

MProlog hat den Kopf der Intelligenz für die hochentwickelte Problemlösung der Expertentechnik. Die Aufrechterhaltung dieses Wertes ist ein Ziel der MProlog-Entwickler.

Das MProlog-System ist ein hochentwickeltes System zur Lösung von Problemen. Es ist ein System, das die Intelligenz von Wert in den Rechner überträgt. Es ist ein System, das die Intelligenz von Wert in den Rechner überträgt.

Wir sind stolz auf die MProlog-Systeme, die die Intelligenz von Wert in den Rechner überträgt. Es ist ein System, das die Intelligenz von Wert in den Rechner überträgt.

kostengünstige
Trennung von
Entwicklungs- und
Einsatzumgebung

für PC bis
Großrechner



bequeme Wartung
durch modularen
Programmaufbau

Fehlersuche mit
interaktivem
Debug-System

über 140 eingebaute
Prädikate

Programmierung
umgebung mit
sprachorientiertem
Editor

Anschluß an
Fortran, Pascal,
C, Assembler

MProlog bietet
die besten Algorithmen für
VAX IBM 310
Siemens 7000
Nixdorf 8800 90
Nixdorf Target 31 35
SUN Apollo
AT&T ST IBM PC

ε
epsilon

epsilon
Gesellschaft für
Software-Entwicklung und
Systementwicklung
mbH
Kurfürstendamm 188
1000 Berlin 19
Tel. 030-882 92 91

Literatur

- Arnold, D. et al. 1986: The EUROTRA Reference Manual 2.0, Utrecht.
- Batori, I. / Weber, H. J. (Hrsg.) 1986: Neue Ansätze in maschineller Sprachübersetzung: Wissensrepräsentation und Textbezug. Tübingen.
- Bergmann, H. et al. 1986: Beratung und natürlichsprachlicher Dialog - eine Evaluation von Systemen der Künstlichen Intelligenz. WISBER-Bericht Nr. 8, Universität Hamburg.
- Block, H.-U. et al. 1985: Neuere Grammatiktheorien und Grammatikformalismen. WISBER-Bericht Nr. 1, Siemens AG München.
- Bresnan, J. (Ed.) 1983: The Mental Representation of Grammatical Relations. MIT Press, Cambridge.
- Butler, Chr. 1985: Systemic Linguistics. Theory and Application. London: Batsford.
- Chomsky, N. 1981: Lectures on Government and Binding. Dordrecht.
- COLING 1986: Proceedings of the 11th International Conference on Computational Linguistics, Bonn, August, 25th to 29th, 1986.
- Dahl, V. / McCord, M.C. 1983: Treating Coordination in Logic Grammars. American Journal of Computational Linguistics, Bd. 9, S. 69-91.
- Eckert, U., Heid, U. 1987: Methoden zur teilautomatischen Erstellung von Transferwörterbüchern. In: IAI Working Papers, Bd. 3.
- Fawcett, R.P. 1980: Cognitive Linguistics and Social Interaction. Heidelberg: Groos.
- Gazdar, G. et al. 1985: Generalized Phrase Structure Grammar. Oxford (Blackwell).
- Haller, J. 1987a: Forschungsergebnisse für Maschinelle Übersetzung - Ein Überblick. LDV-Forum, Bd. 5, Nr. 1, S. 5-7.
- Haller, J. 1987b: EUROTRA - gegenwärtiger Stand und Planung 1987/88. LDV-Forum, Bd. 5, Nr. 1, S. 8-9.
- Haller, J. / Reyle, U. 1987: Statusseminar zu EUROTRA-D und Begleitforschung im Februar 1987. LDV-Forum, Bd. 5, Nr. 1, S. 10-11.
- Halliday, M.A.K. 1985: An introduction to functional grammar. London: Edward Arnold.
- Hauenschild, Ch. 1986: KIT/NASEV oder die Problematik des Transfers bei der maschinellen Übersetzung. In: Batori/Weber 1986.
- IAI-MT 1986: Proceedings of the 1. International Conference on the State of the Art in Machine Translation in America, Asia and Europe, 20. - 22. August 1986, Dudweiler.
- IBM-ELS 1986: Proceedings of the ELS Conference on Translation Mechanization in Cooperation with the University of Copenhagen, August 20-22, 1986, Kopenhagen.
- Karttunen, L. 1985: Working with HUG. SRI International and CSLI, Stanford.
- Kay, M. 1984: Functional Unification Grammar: A Formalism for Machine Translation. In: Proceedings of the Tenth International Conference on Computational Linguistics, Stanford.
- Kilbury, J. 1984: A Modification of the Earley-Shieber Algorithm for Direct Parsing of ID/LP Grammars, in: Laubsch, J. (Hrsg.): GWA1-84, Berlin.
- Kindermann, J. 1985: Implementierungsvorschläge für Grammatiken zum Bottom-Up-Parsing der Sprachen L1, L11, L12 und L13. Arbeitspapier Universität Bielefeld.
- Kindermann, J. / Meier, J. 1985: Bottom-Up-Parsing und Generierung in Prolog, ein Vergleich mit Definite-Clause Grammars Bielefeld, August 1985.
- Langacker, R. W. 1982: Space Grammar, Analyzability, and the English Passive. Language 58, S. 22-80.
- Langendonck, W. Van 1980: Infinites, Exemplars and Kinds. In: Van der Auwera: The Semantics of Determiners. London, Croom Helm.
- Lavorel, B. 1982: Experience in English-French Post-Editing. In: Lawson, V. (Ed.): Practical Experience in Machine Translation. North-Holland ASLIB 1982.
- Licher, V. et al. 1986: FUSL - Eine funktionale Sprache zur Repräsentation linguistischen Wissens und linguistischer Strategien. Linguistische Arbeiten des SFB 100. Neue Folge Heft 16, Universität des Saarlandes, Saarbrücken.
- Oitana, C. 1985: A Man-Aided Machine Translation System. Olivetti Res. & Tech. Review Nr. 4/1985, S. 99-110.
- Pause, P. E. 1986: Zur Modellierung des Übersetzungsprozesses. In: Batori/Weber 1986.
- Rohrer, C. 1985: Maschinelle Übersetzung mit Unifikationsgrammatiken. In: Batori/Weber 1986.
- Schmidt, P. 1986: Valency Theory in a Stratificational MT System. Coling Proceedings 1986.
- Schmidt, P. 1987a: A Syntactic Description of a Fragment of German in the EUROTRA Framework, in: Fawcett, forthcoming.
- Schmidt, P. 1987b: Syntaktische Analyse deutscher Ansätze im EUROTRA-Formalismus. LDV-Forum, Bd. 5, Nr. 1, S. 12-14.
- Schütz, J. 1986: Testing and Evaluation of Compiled Protocols. Internes Arbeitspapier des IAI, Saarbrücken.
- Schütz, J. 1987: EUROTRA-Software-Entwicklung. LDV-Forum, Bd. 5, Nr. 1, S. 21-23.
- Sells, P. 1985: Lectures in Contemporary Syntactic Theories: An Introduction to Government-Binding Theory, Generalized Phrase Structure Grammar, and Lexical Functional Grammar CSLI, Stanford. Nr. 85-73221.
- Steiner, E. 1983: Die Entwicklung des Bräutchen Kontextualismus Heidelberg, Groos.
- Steiner, E. 1985: Working with Transitivity system networks in semantic-grammatical descriptions. In: Henson/Greaves eds. 1985. Systemic perspectives on discourse. Norwood, N.J.: Ablex, Vol. 1.
- Steiner, E. 1986: Generating semantic structures in EUROTRA-D. In: COLING Proceedings 1986 (Universität Bonn).
- Steiner, E. 1987a: Semantic Relations in LFG and EUROTRA-D: a comparison. In: IAI Working Papers, Nr. 3.
- Steiner, E. 1987b: Zur Zuweisung satzsemantischer Rollen im Rahmen des MT-Projektes EUROTRA-D. In: Klent/Scherer (eds.), forthcoming. Ausgewählte Beiträge der GLDV-Jahrestagung 1986. Hildesheim: Olms.
- Steiner, E. 1987c: Grundprinzipien und Vorschläge für eine semantische Beschreibung von Argumentstrukturen. LDV-Forum, Bd. 5, Nr. 1, S. 17-18.
- Steiner/Eckert/Weck/Winter 1987: The development of the EUROTRA-D system of semantic relations. In: IAI Working Papers, Nr. 2. Erschienen ebenfalls als: The application of Systemic Functional Grammar to Machine Translation within the MT-Project EUROTRA-D. Dursburg: LAUD Papers in Linguistics.
- Uszkoreit, H. 1984: Word Order and Constituent Structure in German. Doctoral Dissertation, University of Texas at Austin.
- Varile, N. et al. (in preparation): The New EF-Framework.
- Wehrli, E. 1984: A Government-Binding Parser for French. ISSCO Working Paper No. 48.
- Winograd, T. 1983: Language as a cognitive process. Bd. 1: Syntax. New York: Wiley.
- Zelinsky-Wibbelt, C. 1987a: Die Überführung von der relationalen Strukturbeschreibung in die Interface-Struktur. LDV-Forum, Bd. 5, Nr. 1, S. 15-16.
- Zelinsky-Wibbelt, C. 1987b: Lexikalische Disambiguierung mit semantischen Merkmalen. LDV-Forum, Bd. 5, Nr. 1, S. 19-20.
- Zelinsky-Wibbelt, C. 1988: The Semantic Representation of sentences by Means of Semantic Features. In: Fawcett R. forthcoming. London: Pinter.

DER AUFBAU MEHRSPRACHIGER LEXIKA MIT DEM LEXIKONSYSTEM "MULI"

EUROTRA-D-Teilprojekt am Institut für Kommunikationsforschung und Phonetik (IKP), Bonn

Seit 1985 wird vom Bundesministerium für Forschung und Technologie (BMFT) das EG-Projekt zur Entwicklung eines maschinellen Übersetzungssystems (EUROTRA) gefördert. Als Teilprojekt des EUROTRA-D Projekts in Saarbrücken werden am Institut für Kommunikationsforschung und Phonetik (IKP Bonn) die Transferlexika zu den Sprachen: Dänisch, Griechisch, Niederländisch und Italienisch aufgebaut. Damit verbunden ist die Aufgabe, Software zu entwickeln und einzusetzen, die eine weitreichende Computerunterstützung der lexikographischen Arbeit leistet. Der folgende Beitrag soll die bei der lexikographischen Arbeit eingesetzte Software vorstellen und die Vorgehensweise bei der Lexikonerstellung erläutern.

Im Projektzeitraum (1.9.85-30.4.87) wurden die Transferlexika griechisch-deutsch, dänisch-deutsch und niederländisch-deutsch in einem Umfang von ca. 2.500 Worteinträgen bearbeitet. Als Ausgangsbasis diente hier der in alle EG-Sprachen übersetzte Text des ESPRIT-Programms der EG (kurz ESPRIT-Text). Für die Lexikonbearbeitung wurde das Lexikonsystem MULI (Multi LIngual Dictionary System) entwickelt und zu einem rechnergestützten Arbeitsplatz für den Lexikographen ausgebaut.

Das Lexikonsystem MULI erlaubt es, die bilingualen Lexika in EUROTRA zu erstellen, zu vergleichen und zu verwalten (Abb. 1). Neben den Grundfunktionen Erfassen, Korrigieren und Löschen von Lexikoneinträgen zu einem ausgewählten Sprachpaar unterstützt MULI den Zugriff auf Lexikoneinträge anderer Sprachpaare und ggf. die Übernahme von Teilinformationen in das gerade bearbeitete Lexikon. Das Importieren und Exportieren lexikalischer Daten (von einer Datenbankdatei bzw. in eine ASCII-Datei) wird durch zusätzliche Module realisiert.

Die Erstellung verschiedener bilingualer Lexika mit gemeinsamen Sprachkomponenten wirft das Problem einer konsistenten Übersetzung der verschiedenen Lesarten eines Wortes auf. Die durch Lesartennummern gekennzeichneten Lesarten sollten im Hinblick auf die gemeinsame Nutzung der Lexika in einem Gesamtsystem durchgängig in der gleichen Wortbedeutung übersetzt sein. Das heißt z.B., daß [Bezug 1] im Dänischen nicht in der Wortbedeutung [Bezug einer Zeitung] und im Niederländischen im Sinn von [Bezug der Wohnung] übersetzt sein darf. Das Lexikonsystem bietet durch die Strukturierung der Datenbank und ihre Verknüpfung die Möglichkeit zur Konsistenzüberprüfung und -

absicherung. Jeder Worteintrag einer Sprache wird nur einmal in der Datenbank des Lexikonsystems geführt. Übersetzungsäquivalente werden durch eine 'Datenbankrelation' miteinander verknüpft. Insbesondere werden importierte lexikalische Daten zunächst auf Inkonsistenzen mit den bereits vorhandenen Lexikoneinträgen überprüft.

Die Erweiterung des Lexikonsystems in Richtung auf einen lexikographischen Arbeitsplatz ist vor allem durch die Verbindung mit einem Konkordanzsystem erreicht worden. Das Konkordanzsystem erlaubt das direkte Aufsuchen von Textstellen mit bestimmten Charakteristika (Suchbedingungen). So kann z.B. in einem Text, wie dem ESPRIT-Korpus, ganz gezielt nach dem Vorkommen bestimmter Wortformen und Wortverbindungen gesucht werden oder auch nur ein schneller Überblick darüber gewonnen werden, welche Lesarten eines Wortes im Text auftreten. Verwendet wird das Konkordanzsystem BYU, bei dem es sich um ein Retrieval System mit umfangreichen Funktionen zur Textanalyse handelt. Da das Konkordanzsystem seinerseits auch den Zugriff auf ein angeschlossenes Lexikon erlaubt, kann bei der Analyse einer Textstelle ein direkter Vergleich mit den betroffenen Lexikoneinträgen vorgenommen werden.

Für den computerunterstützten Lexikonaufbau stellt das BYU-System schon bei der Materialerstellung eine große Hilfe dar. Die Lexikonerstellung ist rechnergestützt in den folgenden Schritten erfolgt.

1. Mit dem BYU-System wird ein Textindex aufgebaut, der sämtliche Wortformen eines Textes mit Angabe der Fundstellen, der absoluten Häufigkeit und einem Kontextausschnitt enthält. Eine solche Wortliste, generiert aus dem dänischen, niederländischen und griechischen ESPRIT-Text, ist die ideale Ausgangsbasis für die weitere lexikographische Bearbeitung. Der Lexikograph kann sich so auf die mitgegebene Kontextangabe stützen oder falls notwendig die gesamte Textpassage herausuchen.
2. In einer ersten manuellen Bearbeitungsphase wird die Wortliste aller 'types' lemmatisiert und so die Grundlage für das zum Text gehörende Grundformenlexikon geschaffen. Für das Deutsche ist inzwischen das System zur automatischen Lemmatisierung von G. Willée auch auf Personal Computer lauffähig.
3. In der weiteren Bearbeitung werden Informationseinheiten wie Wortartenspezifizierung und die deutsche Übersetzung ergäzt. Durch diese Vorgehensweise kann der Lexikograph zumindest zu einem Teil von der Erfassungsarbeit befreit werden.

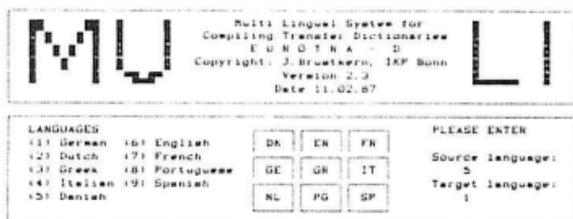


Abb. 1: Auswahl eines Sprachpaars im Lexikonsystem MULI

Source Language: DANISH		Target Language: GERMAN	
Word entry aftagelae	Rnr 1	Word entry Bezug	Rnr 2
geometry : condition: SEMF # * COMMERCIAL SL-Phrase: vid aftagelae af 1000 stykker TL-Phrase: bei Bezug von 1000 Stück Comment: IDzu:dtak		TRANSFERLEXIKON	

SIDE	ENTRY	PAGE	MODE	ACTION	SELECT	Dictionary	UPDATE
Left	Next	Up	Index/Transfer	Quit	Entry	W/Change	Add
Right	Previous	Down	Byu-Concordance	Help	Move/View	all	Get rid Show xt

Abb. 2: Benutzeroberfläche zur Lexikonbearbeitung

4. Die so gewonnenen lexikalischen Grunddaten werden in das Lexikonsystem MULL übernommen. Für Bearbeitung, Zugriff und Vergleich der lexikalischen Daten stehen damit zahlreiche Funktionen zur Verfügung (Abb. 2).

In einer späteren Phase sollen die Transferlexika inhaltlich ergänzt werden. Dies betrifft vor allem die Aufnahme weiterer Lesarten und semantischer Spezifikationen. Die lexikalische Arbeit wird außerdem auf das Italienische ausgedehnt. Die

im Projekt aufgebaute Softwareumgebung für den computergestützten lexikographischen Arbeitsplatz wird weiter entwickelt. Für diese Arbeiten ist eine Verlängerung des Projekts bis zum 30.6.1988 beantragt.

Jan Brustkern, EUROTRA-D, Institut für Kommunikationsforschung und Phonetik, Bonn, Universität Bonn

"Semantik ... UND ... Generierung"

↓

and(semantics+, generation+)

Bericht über den Zweiten Externen Workshop der EUROTRA-D-Begleitforschung

Es ist vermutlich ungewöhnlich, daß die Organisatoren eines Workshops selbst die Berichterstattung übernehmen – man könnte Mangel an Objektivität vermuten; aber Objektivität ist ja auch kein Wert an sich, und außerdem ist sie auch für "normale" Teilnehmer nur schwer zu erreichen. Das hier gewählte Verfahren hat zumindest den Vorteil, daß auch über Voraussetzungen berichtet werden kann, die nur den Organisatoren bekannt sind.

Eigentlich sollte der Workshop unter dem Titel "Transfer und Generierung" laufen (so beschlossen beim Bielefelder internen Arbeitstreffen der Begleitforschung im Juli 1986). Dann stellte sich aber heraus, daß es sehr schwierig sein würde, zum Thema "Transfer" eine genügende Anzahl von interessanten Vorträgen zu bekommen – das Theorie- (und wohl auch Praxis-) Defizit ist in diesem Bereich offensichtlich noch größer, als von uns zunächst angenommen. Im Gegensatz zu anderen defizitären Theoriebereichen der Maschinellen Übersetzung gibt es zum Thema "Transfer" anscheinend auch keine brauchbaren Vorschläge aus der Linguistik oder ihren Nachbardisziplinen. Daher wurde im November in Stuttgart (beim nächsten internen Arbeitstreffen) beschlossen, sich für den Workshop solchen Themen zuzuwenden, die zwar im Bereich der MÜ Desiderata sind, für die es aber in anderen Bereichen hinreichend vielversprechende Ansätze gibt: "Koordination", "semantische Komponenten (im Anschluß an syntaktische)" und "Generierung". Es wurde dabei in Kauf genommen, daß der 2. externe Workshop der EURATO-D-Begleitforschung dadurch thematisch (noch) etwas heterogener wurde als der erste externe Workshop, der unter dem Titel "Word Order and Parsing in Unification Grammars" im April 1986 in Friedenweiler (Schwarzwald) stattfand. Die beiden Workshops sind auch in anderer als thematischer Hinsicht komplementär zueinander; während in Friedenweiler die Arbeitssprache ausschließlich Englisch war, wie sie in Dudweiler vorwiegend Deutsch (mit interessanten polyglotten Einsprengeln, wo ein englischer Muttersprachler auf Deutsch einem Englisch fragenden deutschen Muttersprachler auf Deutsch antwortete – eigentlich hatten wir uns die Sache umgekehrt gedacht, nämlich so, daß jeder in seiner Muttersprache reden (können) sollte. Die Frage der Arbeitssprachen hängt natürlich immer eng mit dem jeweils angesprochenen Teilnehmerkreis zusammen: in dieser Hinsicht war der zweite Workshop eine Nummer kleiner als der erste: es waren in Dudweiler ca. 30 Teilnehmer (mit punktueller Verstärkung durch ortsansässige Saarbrücker), also etwa die Hälfte im

Vergleich zu Friedenweiler – die Teilnahme aus dem europäischen Ausland war in Dudweiler wegen kurzfristigen Absagen bedauerlich gering. Allerdings ist es unsere ganz persönliche Meinung, daß es auch Vorteile hat, wenn die deutschen Kolleg/inn/en diffezile Themen tatsächlich auf Deutsch diskutieren können (was sich besonders bei der Gesamtdiskussion über Fragen der Generierung positiv bemerkbar machte).

Übrigens gab es auch mit der alternativen Thematik noch Probleme, Koordination scheint ein in jeder Hinsicht schwieriges Gebiet zu sein (sowohl auf Objekt- als auch auf der Meta-Ebene!); jedenfalls ist die Koordinations-Sektion auf eine einzige Klammer – immerhin mit einem öffnenden und einem schließenden Teil – zusammengeschumpft. Die Semantik-Sektion litt besonders unter kurzfristigen Absagen von geplanten Vorträgen aus dem europäischen Ausland, was allerdings den Syntaktikern die Chance bot, ihre traditionelle starke Rolle in gewissem Umfang zu behaupten (präziser: einige von ihnen sind dankenswerterweise kurzfristig eingesprungen). Allein das Generierungs-Thema erwies sich als zuverlässige Basis für eine ordentliche Sektion mit einer von vielen Teilnehmern als sehr fruchtbar eingeschätzten Gesamtdiskussion.

Zu den Beiträgen im einzelnen (die Zusammenfassungen stützen sich z.T. auf Vorlagen der Vortragenden):

Manfred Pinkal (Stuttgart) öffnete die Koordinationsklammer mit einem Beitrag zum Thema "Koordination in Komparativen". Der Vortrag setzte an beim auffälligen semantischen Verhalten der Konstruktion NP oder NP in Komparativ-Komplementen, z.B.:

- (1) Maria ist größer als Anna oder Sarah
- (2) Maria ist größer als Anna oder als Sarah
- (3) Maria ist größer als Anna und als Sarah

Satz (1) ist mehrdeutig zwischen den Lesarten (2) und (3), und (3) scheint die präferierte Lesart zu sein. Dies und verwandte Phänomene haben zu komplexen semantischen Theorien des Komparativ geführt, bei denen das gesamte Komparativkomplement sematisches Argument des Komparativoperators ist ('Enge-Skopus-Analyse'). Im Vortrag wurden verschiedene Argumente für eine einheitliche einfache Semantik des Komparativ als transparente zweistellige Relation angeführt ('Weite-Skopus-Analyse'). Die Problematik von (1) wurde darauf zurückgeführt, daß oder eine zweite, Nicht-Standard-Lesart besitzt, die auch außerhalb von Komparativkonstruktionen verbreitet ist ('free-choice-oder'). Diese Analyse soll nicht nur für phrasale Komparativkomplemente, sondern auch für Satzkomplemente gelten. Im zweiten Teil des Vortrags wurden die Eigenschaften von free-choice-Interpretationen diskutiert und mit anderen Arten von Nicht-Standard-Lesarten bei NPs (z.B. generische Lesart bei Indefiniten) in Beziehung gesetzt.

Da der zweite Vortrag zum Thema "Koordination" aus technischen Gründen erst am letzten Tag des Workshops stattfinden konnte, hatten am Mittwochnachmittag dann die Parsing-Spezialisten das Wort. Sven Naumann (Göttingen) berichtete über seinen "einfachen Parser für generalisierte Phrasenstrukturgrammatiken".

Im Mittelpunkt des Vortrags standen die Differenzen zwischen dem Parser zugrundeliegenden Annahmen und der Standard-Version der generalisierten Phrasenstrukturgrammatiken.

Dem Parser liegt folgendes Organisationsmodell zugrunde: einerseits werden zunächst kontextfreie Regeln erzeugt, mit denen dann der Parser arbeitet; andererseits aber bilden die Instanzierungsprinzipien einen integralen Bestandteil des Parser, d.h. sie wirken bei der Generierung von Strukturen als spezielle Filter, die die Erzeugung von Strukturen verhindern, die eines der drei Instanzierungsprinzipien verletzen.

Das Programm besteht aus drei Teilen: einem Pre-Compiler, dem Parser und einem Post-Filter. Die Aufgabe des Pre-Compilers besteht darin, aus den ID-Regeln, den LP-Regeln und den Metaregeln eine (endliche) Menge von kontextfreien Regeln mit komplexen Symbolen zu erzeugen. Der Post-Filter operiert auf den vom Parser gebildeten Strukturen; er ergänzt Kategorien um weitere Merkmale und filtert die Strukturen aus, die Kategorien enthalten, die nicht den FCRs genügen. Der Parser basiert auf einem modifizierten 'active-chart'-Parser für kontextfreie Syntaxen.

Die zentrale Annahme, auf der das Programm basiert, ist die, daß es möglich ist, die in der Standard-Version der GPSG postulierte 'freie' Instanzierung von Merkmalen durch eine lexikalisch determinierte Instanzierung zu ersetzen, ohne daß die Beschreibungsadäquatheit von GPSG darunter leidet. Dem Lexikon kommt damit eine wichtige Rolle zu: es muß relativ umfangreiche Informationen bereitstellen, die beim Parsen zunächst an die lexikalischen Kategorien weitergegeben und dann durch die Instanzierungsprinzipien über die erzeugten Strukturen verteilt werden. Die FSDs und FCRs verlieren so erheblich an Bedeutung; die FSDs instanzieren, wo möglich, weitere Merkmale in Kategorien und die FCRs testen, ob die erzeugten Strukturen akzeptabel sind.

Im Anschluß daran stellte Jörg Kindermann (Bielefeld) den im Bielefelder Teilprojekt der Begleitforschung entwickelten erweiterten LR-Parser für Lexikalisch-Funktional Grammatiken vor. Bei der Entwicklung wurde von den folgenden Kritiken ausgegangen:

- Effizienz des Parsing-Prozesses, d.h. Grad an Determinismus sowie Rechenzeit und Speicherplatz im Verhältnis zur Länge des Inputs,
- Restriktionen für die Grammatik, d.h. Beschränkungen bezüglich des zulässigen Typs von Grammatikregeln (linksrekursiv) und bezüglich der Regelordnung,

- Behandlung der Unifikation durch den Parser, was besonders für LFG wichtig ist,
- Fehlersuch-Hilfen für den Regelschreiber.

Unter diesen Gesichtspunkten fiel die Entscheidung für deterministisches Parsing mit einem erweiterten LR-Algorithmus im Sinne der Vorschläge von Tomita zur Behandlung von mehrdeutigen kontextfreien Grammatiken. Es werden Repräsentationen der LFG erzeugt, wobei keine unnötigen (d.h. letztlich erfolglosen) Reduktionen durchgeführt werden (also: hoher Grad an Determinismus). Es gibt keine Beschränkungen bezüglich der Form und Ordnung der kontextfreien Regeln. F-Strukturen werden inkrementell erzeugt, so daß Ambiguitäten effizient behandelt werden können (Inkonsistenzen werden so früh wie möglich erkannt). Die Erweiterung eines Grammatikfragments wird besser unterstützt als bei einem top-down-backtracking Parser.

Am Donnerstag war Generierungstag. Er wurde eröffnet von Stephan Busemann (Berlin), der das im Berliner Teilprojekt der Begleitforschung implementierte GPSG-System vorstellte und insbesondere dessen Generierungskomponente. Das System wurde vor dem Hintergrund der möglichen Einbettung in ein komplexes natürlingsprachliches (Übersetzungs-) System entwickelt, wodurch sich erste Anforderungen an die semantische Repräsentationsebene ergeben, die die Schnittstelle zu GPSG bildet.

Ein wichtiges Design-Kriterium war die strikte Trennung von Kontrollwissen und deklarativ und prozedural repräsentiertem sprachlichen Wissen. Damit bleibt klar, was "GPSG-Anteile" des Systems sind und was diese "zum Laufen bringt" (Kontrollwissen). Dieser in der KI oft mit "Wissensbasiertheit" bezeichnete Ansatz ermöglicht den Austausch von Grammatiken, was im Hinblick auf maschinelle Übersetzung entscheidend ist. Zugleich eröffnet sich damit die Möglichkeit, Analyse und Generierung mit ein- und derselben Repräsentation des sprachlichen Wissens zu betreiben.

Die Generierung verläuft in der gegenwärtigen Implementation top-down (depth-first). Sie geht von einer vorläufigen semantischen Repräsentation aus, die lexikalisches Material auf zwei Ebenen strukturiert, der Formel- und der Termebene. Alles Material ist zu verbalisieren, d.h. zunächst in eine syntaktische Struktur abzubilden. Die Funktionsweise des Generators sowie der verwendeten Operationalisierung der GPSG-Theorie wurde an einem Beispiel erläutert.

Jürgen Wedekind und Jochen Dörre (Stuttgart) stellten den im Stuttgarter Teilprojekt entwickelten Ansatz zur Generierung mit LFG in einem mehr theoretischen ersten und in einem mehr technischen zweiten Teil ihres Beitrags dar. Der Stuttgarter Generator-Compiler benutzt den von Jürgen Wedekind ausgearbeiteten monostratalen Ableitungsbegriff für LFG, der ohne explizite Erzeugung einer C-Struktur auskommt. Ausgangspunkt der Gene-

rierung sind gerichtete azyklische Graphen (DAGs), die partiellen F-Strukturen entsprechen, die mit zusätzlichen Informationen über die abgeleiteten Symbole und deren lineare Abfolge angereichert sind. Der Generator-Compiler funktioniert praktisch so, daß die Wohlgeformtheit der Input-Strukturen überprüft und die gewünschte Endkette als Kontrollstruktur erzeugt wird. Durch die Bedingung, daß bestimmte Constraints an den Kanten erfüllt sein müssen, wird sichergestellt, daß alle vorhandene Information tatsächlich abgearbeitet wird. Hinzufügen von Information ist nur unter ganz bestimmten Bedingungen erlaubt (z.B. Ergänzung von Kasus-Information bei in dieser Hinsicht unterspezifizierten F-Strukturen, wie sie als Ergebnis des Transfers auf F-Struktur-Ebene zu erwarten sind). Wenn eine Input-Struktur vom Generator als nicht wohlgeformt erkannt wird, soll sie an die (noch zu implementierende) Transferkomponente zurückgegeben werden.

Danach stellte Russell Block (Hamburg) zwei Möglichkeiten zur Auswahl: "Generating Natural Language Without Transformations or Without Tears? - The Evidence From LFG". Drei Punkte standen im Mittelpunkt seines Vergleichs von LFG und Standardtheorie.

Der erste betraf Ansprüche hinsichtlich psychologischer Plausibilität in LFG, die nicht erfüllt zu sein scheinen. Es ist plausibler, Regeln unabhängig vom Wortschatz eines Sprechers anzunehmen, denn Sprecher können auch passivieren, wenn sie keinen Lexikonbeitrag für ein (unbekanntes) Verb haben. Das bedeutet auch, daß Grammatik und Sprachgenerierung bzw. Sprachanalyse als eigenständige Ebenen der sprachlichen Kompetenz zu unterscheiden sind.

Die Überlegenheit der LFG gegenüber der Standardtheorie muß nachgewiesen werden, um die Entwicklung neuer Begrifflichkeiten und Systeme zu rechtfertigen. Block zeigt an Permutationen der Wortstellung ("Bahnhofsätze"), daß für eine LFG (des Deutschen) eine große Menge Phrasenstrukturregeln (im Beispiel 18) notwendig ist, um alle akzeptablen Varianten zu generieren. Stattdessen sei es einfacher, mit einer PS-Regel und einigen Transformationen zu arbeiten. Weitere Beispiele (u.a. Passivierung) untermauern die These, daß LFG den Nachweis der Überlegenheit auch bezüglich Einfachheit und neuer Einsichten in das Funktionieren von Sprache nicht erbringt.

Der dritte Hauptpunkt betrifft die Eignung von LFG für Generierung im Gegensatz zur Analyse. Während LFG für Parsing gut geeignet scheint, ja geradezu darauf zugeschnitten ist, erweist sich die Theorie als ineffizient für die Generierung. Denn es muß auf F-Struktur- und auf C-Struktur-Ebene generiert und ständig müssen Konsistenz- und Vollständigkeitstests durchgeführt werden. Dabei kommt es zu beträchtlicher Übergenerierung von F-Strukturen, die wieder ausgefiltert werden müssen. Solche Probleme ergeben sich nicht bei einem Generator auf der Grundlage der Standardtheorie.

Dietmar Rösner (Stuttgart) ging das Thema Generierung wieder anders an. In seinem Vortrag "Textphänomene und rhetorische Strukturen" beschrieb er Anwendungen des SEMSYN-Generators für die Verbalisierung geometrischer Konstruktionen (GEOTEX) und für die Erzeugung von Arbeitsmarktberichten aufgrund statistischer Daten (SEMTEX). Der Generator geht von den zu verbalisierenden Inhalten aus und bestimmt, wie der Text aussehen soll. Das Verfahren enthält eine explizite Textplanung, die die Reihenfolge der zu äuffernden Fakten bestimmt (z.B. nach dem Schema: wichtigste Daten, evtl. Elaborierung, Kommentare und ggf. Bewertung). Auf dieser Strukturierung operieren kontextabhängige Mechanismen zur Verbesserung der 'Lesbarkeit' (unmittelbare Oberflächengenerierung ergäbe eine nicht nur stilistisch schlechte Folge von Einzelsätzen). Es wird anhand von Synonymlexikons vermieden, Wörter zu wiederholen; temporale Information wird ausgelassen, wenn sie unverändert bleibt; Tempusbestimmung erfolgt anhand von Informationen über Sprechzeitpunkt und Referenzzeitraum. Die zur Textstrukturierung verwendeten rhetorischen Schemata sind fest mit bestimmten Satzbaumustern korreliert. Sie werden aufgrund des zu verbalisierenden Materials instanziiert. Das System ordnet Konstituenten anhand der Information über *neu* vs. *bekannt*. Bestimmte Möglichkeiten zur Elliptifizierung und Pronominalisierung runden das Spektrum semantisch-pragmatischer Mittel ab, um eine bemerkenswerte Performanz des System zu erzielen.

Auf Anregung von Christopher Habel fand am Donnerstag Nachmittag eine Diskussion über Generierung statt, die die Kontroversen des Vormittags aufgriff und kanalisierte. Es wurde eine ungewöhnlich anregende Debatte über inhaltliche und paradigmatische Ansprüche. Ausgehend von Habels Unterteilung in drei komplementäre (?) Sichtweisen (theoretisch-mathematisch, kognitiv adäquat, anwendungsbezogen) wurden u.a. die unterschiedlichen Intentionen klar, mit denen die Exponenten der verschiedenen Sichtweisen ihre Systeme implementieren. Für Wedekind ist die Implementation eine geeignete Testmöglichkeit der Grammatik. Block steht die effiziente Modellierung kognitiver Prozesse an. Bei Rösner steht die Performanzleistung in einem Diskursbereich im Vordergrund. Es wurde unterschieden zwischen Sprachwissenschaft, der prozedurale Aspekte relativ gleichgültig seien und Sprecher-Hörer-Wissenschaft, die prozedurale Fragestellungen mit einschließt. Insgesamt war man sich bald einig, daß alle drei Sichtweisen sinnvoll und notwendig sind, jedoch bestanden unterschiedliche Auffassungen über die Art der gegenseitigen Beeinflussung: Kognitive Modellierung sollte nicht abwarten, bis die Sprachwissenschaftler ihre Arbeit erfolgreich abgeschlossen haben. Nun, nüchtern betrachtet führt jedenfalls kaum ein Weg daran vorbei, für ein gut funktionierendes MÜ-System, wie es mit EUROTRA

anvisiert wird, alle drei Aspekte unter einen Hut zu bringen...

Die beiden ersten Vorträge am Freitag beschäftigen sich in unterschiedlicher Weise mit dem Verhältnis zwischen Syntax und Semantik.

Zunächst stellte *Rudolf Hunze* (München) ein von der Siemens AG im Rahmen des Projektes WISBER entwickeltes System zur syntaktisch-semantischen Analyse von natürlichsprachlichem Input unter dem Titel "LFG, DRT und die Prinzipien A,B,C" vor. Der Vortrag beschäftigte sich mit der Erzeugung von Diskursrepräsentationsstrukturen (im Sinne von H. Kamp) aus einer LFG-Grammatik heraus. Besonderer Wert wurde auf eine flexible Anknüpfung des Semantiktells gelegt. So werden die DR-Strukturen nicht direkt aus den F-Strukturen erzeugt, sondern der Unifikationsmechanismus, den LFG bietet, wird ein zweites Mal benutzt, indem zu jeder syntaktischen Kategorie eine Gleichungsmenge assoziiert werden kann, deren Auswertung zum Aufbau der DRS führt. Dabei wird ein fester, auf der C-Struktur arbeitender Interpretationsalgorithmus benutzt. Durch Aufstellen entsprechender Gleichungen hat es der Grammatikschreiber weitgehend in der Hand, wie die DR-Strukturen aussehen, d.h. wie die Boxen zusammengebaut werden. Er muß lediglich die von Kamp verwendete Form (wonach jede Box aus Variablen und Bedingungen und evtl. eingebetteten Boxen besteht) beachten.

Zur Vorbereitung der Auflösung anaphorischer Referenzen kann der Grammatikschreiber zwei reservierte Prädikate benutzen, nämlich (DISJOINT *refo list*) und (FINDREF *refo*). (FINDREF *refo*) drückt aus, daß für das Referenzobjekt *refo* ein Antezedens zu bestimmen ist (dieser Schritt erfolgt erst, nachdem die DR-Strukturen aufgebaut sind) und (DISJOINT *refo list*) bedeutet, daß *refo* von den Referenzobjekten auf der *list* verschieden sein muß. Diese letzte Eigenschaft wird zur Beschreibung der Prinzipien A, B, C der Chomsky'schen Bindungstheorie eingesetzt.

Unter dem Titel "PASS" erläutert *Dieter Wunderlich* (Düsseldorf) einige neue Ideen zu einer mit Prädikat-Argument-Strukturen angereicherte Syntax/Morphologie.

PAS dient zur Integrierung der phonologischen, syntaktischen und semantischen Information eines Lexikoneintrags. Im weiteren Sinne lassen sich auch Funktoren als "Prädikate" auffassen.

Lexikalische Prozesse zur Ableitung neuer Lexikoneinträge können über PAS definiert werden. Typische lexikalische Ableitungsregeln sind z.B.:

- Tilgung von Argumenten
- Einführung von Argumenten
- Existenzquantifizierung über Argumente.

Lokale morphologische oder syntaktische Strukturen ergeben sich aus der Kombination einzelner PAS wie folgt:

1. A-A- oder A-P-Vernetzung der PAS
2. Projektion der einzelnen PAS auf die PAS des Produktes
(die semantische Repräsentation einer Mutter wird aus den semantischen Repräsentationen der Tochter berechnet).

Es wurden Beispiele der folgenden Typen diskutiert: Modifikation versus Funktionale Applikation; Komposita und Denominale Verben.

Vergleich zu LFG: F-Struktur ist in PAS enthalten. Vergleich zu GPSG: Merkmale können wahlweise als prädikatsspezifisierende Kategorien oder als PAS behandelt werden. Auch die Integration mit konzeptueller Information ist möglich.

Die durch Manfred Pinkal eher semantisch geöffnete Koordinationsklammer wurde durch *Tilman Hobbe* (Tübingen) auf eher syntaktische Weise geschlossen. Er erläuterte in seinem Vortrag "Koordination im Deutschen" an zahlreichen Beispielen einiges grundsätzliche zur symmetrischen Koordination:

1. richtig:

- (i) Es gibt phrasale Koordination von diversen Konstituententypen.
- (ii) Auf den durch (i) definierten Strukturen können nur
 - (a) Linkstügung (rechtsperiphere Elation in nicht-letzten Konjunkten, "Fight Node Kas sing") und
 - (b) Gapping (Elation eines unabhängigen Verbs und evtl. weiterer Konstituenten in nicht-ersten Konjunkten) operieren. (Nejt)

2. falsch:

- (i') Negation von (i). z.B. Es gibt phrasale Koordination nur von Sätzen (*Orisvow*) oder nur von N_{max} und V_{max} (5) (*George*).
- (ii') Negation von (ii)(a). Es gibt **Rechtstügung** (linksperiphere Elation in nicht-ersten Konjunkten). (*George, Godall, Orisvow*)
Negation von (ii)(b). Es gibt **Links-Gapping**. Es gibt "Gapping" in NP.
Negation von (ii). Es gibt (im Englischen) koordinative VP-Deletion. (*Orisvow*)

3. Folge von (1)

- (a) Jedes Konjunkt ist eine Konstituente
- (b) Ein erstes Konjunkt kann (nur) infolge von Linkstügung unvollständig sein.
- (c) Ein letztes Konjunkt kann (nur) infolge von Gapping unvollständig sein

Beispiel: Karl hat den Kindern ..., Heinz ... den Eltern ... und Fritz ... den Lehrern ein Buch geschenkt.

Besonders die letzten beiden Vorträge des Workshops boten soviel Stoff für ausgedehnte Diskussionen, daß diesem Bedürfnis im Rahmen der Veranstaltung (wir wollten rechtzeitig schließen, um allen Teilnehmern die Heimkehr vor dem Wochenende

zu ermöglichen) in keiner Weise Rechnung getragen werden konnte. Die jedem Vortragenden zur Verfügung gestellte volle Stunde (bei freier Wahl der Einteilung in Vortrag und Diskussion, was i.a. zu Lasten der Diskussionszeit ging) erwies sich überhaupt in einigen Fällen als zu kurz.

Wir hoffen jedoch, daß die Teilnehmer/innen in den Pausen und abends genügend Zeit gefunden haben, die dringendsten Diskussionswünsche zu befriedigen – man wird sich ja auch nicht das letzte Mal getroffen haben.

Noch ein paar Worte zum Atmosphärischen: Der Workshop fand aus Termingründen direkt im Anschluß an das Statusseminar statt, das vom Saarbrücker Projekt EUROTRA-D zusammen mit der Begleitforschung für Geldgeber, Gutachter und interessierte Fachöffentlichkeit am 16. und 17. Februar im Bürgerhaus Dudweiler organisiert wurde. Diese Nähe hatte Vor- und Nachteile: der wesentliche Vorteil bestand darin, daß die lokale Organisation des Workshops freundlicherweise von EUROTRA-D mit erledigt wurde (von den System-Vorführmöglichkeiten (beschränkt) bis zu den "Konferenzgetränken" (unbeschränkt) war für alles ge-

sorgt); der wesentliche Nachteil lag in dem nicht hundertprozentig geeigneten Räumlichkeiten: was für das Statusseminar statusmäßig angemessen war, war für unseren Workshop eine Nummer zu groß (obwohl die Faschingsdekoration und die dadurch ausgelösten psychedelischen Lichteffekte auf der Projektor-Leinwand durchaus stilvoll waren!). Ein weiterer (unwesentlicher) Nachteil ergab sich aus der Tatsache, daß die Teilnehmer, die zur Schnittmenge der beiden Veranstaltungen gehörten, schon etwas müde waren (besonders natürlich die Hauptakteure des Statusseminars). Die Teilnehmer-Mengen waren jedoch hinreichend disjunkt, um genügend frische Impulse in den Workshop zu bringen.

Wir hoffen jedenfalls, daß es den Teilnehmer/innen unseres Workshops trotz gewisser organisatorischer Probleme gefallen hat und natürlich, daß es nicht der letzte Workshop dieser Art zu MÜ-relevanten Themen war. Das Thema "Transfer" steht schließlich immer noch offen.

*Christa Hauenschild, Stephan Busemann, TU Berlin,
Projekt KIT/NASEV*

ANZEIGE

Intelligente Index/Registerherstellung

mit

ECOINDEX

Herstellung von Stichwortverzeichnissen, Indices,
Such- und Personennennsystemen

Im Unterschied zu anderen Textverarbeitungsprogrammen "liest" ECOINDEX das fertige Dokument selbstständig durch und erstellt einen "Vorschlag" für einen Index, der mit speziellen Editurfunktionen schrift- und einfach modifiziert werden kann.

ECOINDEX verarbeitet alle und beliebig viele ASCII-Textdateien, also auch eine ganze Datenbasis aus vielen Dokumenten.

ECOINDEX verarbeitet auch Dateien von Textverarbeitungssystemen wie **WordStar** (von Microsoft) und **MS-WORD** (von Microsoft).

ECOINDEX kann auch auf **Fotosatz** „gerechneten“ Text verarbeiten.

ECOINDEX kostet nur DM 575,70 → Demo DM 28,50

FORMALE BEGRIFFSANALYSE ERLÄUTERT AN EINEM WORTFELD

Uwe Kipke und Rudolf Wille

Technische Hochschule Darmstadt
 Fachbereich Mathematik, Forschungsgruppe Begriffsanalyse
 Schloßgartenstraße 7
 D-6100 Darmstadt

Die *Formale Begriffsanalyse* baut auf ein mengensprachliches Modell für hierarchische Begriffssysteme auf und erfordert von daher exakte Festlegungen. Sie kann in Form graphischer Darstellungen von Begriffsverbänden ein sehr übersichtliches und differenziert interpretierbares Ergebnis liefern. Bewährt haben sich insbesondere die *gestuften Liniendiagramme*, mit denen auch komplexe Begriffsverbände dargestellt werden können. Die begriffsanalytische Arbeit kann durch den Rechner effektiv unterstützt werden, wozu mehrere Programme entwickelt worden sind. Es hat sich gezeigt, daß die formale Begriffsanalyse in verschiedenartigsten Anwendungsbereichen ein nützliches Werkzeug sein kann. Im vorliegenden Beitrag sollen grundlegende Definitionen und Methoden der formalen Begriffsanalyse am Beispiel eines Wortfeldes erläutert und deren Anwendungsmöglichkeit im Bereich der Sprachwissenschaft aufgezeigt und weiter angeregt werden.

Die "Formale Begriffsanalyse", die auf ein mengensprachliches Modell für hierarchische Begriffssysteme gegründet ist (Wille 1982), stellt in wachsendem Umfang mathematische Methoden zur Untersuchung von Datenkontexten bereit. Wenn im folgenden anhand des Wortfeldes "Gewässer" Ansätze und Methoden der formalen Begriffsanalyse aufgezeigt werden, so geschieht das in der Überzeugung, daß die begriffsanalytischen Methoden auch im Bereich der Sprachwissenschaft mit Gewinn angewendet werden können. Einen Versuch, die Grundmuster möglicher Anwendungen der formalen Begriffsanalyse in ihrer ganzen Breite darzustellen, findet man in Wille 1987.

Ausgegangen wird in der formalen Begriffsanalyse von der Definition eines (*formalen*) *Kontextes* als eines Tripels (G, M, I) , wobei G und M Mengen sind und I eine binäre Relation zwischen G und M ist, d.h. $I \subseteq G \times M$; die Elemente von G bzw. M heißen *Gegenstände* bzw. *Merkmale*, und die Beziehung gIm wird gelesen: "der Gegenstand g hat das Merkmal m ". Von zentraler Bedeutung ist die Galois-Verbindung (Birkhoff 1967, S. 122 ff) zwischen den Teilmengen von G und M , die durch folgende Definition beschrieben wird: Für $A \subseteq G$ und $B \subseteq M$ sei

$$A' := \{m \in M \mid gIm \text{ für alle } g \in A\}$$

$$B' := \{g \in G \mid gIm \text{ für alle } m \in B\}$$

Ein Begriff, der nach der traditionellen Begriffslehre Wagner 1983) durch seinen Umfang (Extension) und seinen Inhalt (Intension) bestimmt ist (vgl. auch DIN 2330 (Din 1979) und DIN 2331 (Din 1980)), kann für einen Kontext (G, M, I) folgendermaßen definiert werden: Ein Paar (A, B) heißt ein (*formaler*) *Begriff* von (G, M, I) , wenn $A \subseteq G$, $B \subseteq M$, $A = B'$ und $B = A'$ gilt; A bzw. B wird der *Umfang* bzw. der *Inhalt* des Begriffes (A, B) genannt. Die Abstraktionsbeziehung zwischen Begriffen wird erfaßt durch die Definition

$$(A_1, B_1) \leq (A_2, B_2) \iff A_1 \subseteq A_2 \quad (\iff B_1 \supseteq B_2);$$

(A_2, B_2) heißt dann *Unterbegriff* von (A_1, B_1) und umgekehrt (A_1, B_1) der *Oberbegriff* von (A_2, B_2) . Die Menge (G, M, I) aller Begriffe des Kontextes (G, M, I) bildet bzgl. der Relation \leq einen vollständigen Verband, der der *Begriffsverband* von (G, M, I) genannt und mit $BV(G, M, I)$ bezeichnet wird.

Begriffsverbände sind das zentrale Untersuchungsobjekt der formalen Begriffsanalyse, deshalb sind Methoden für ihre Bestimmung grundlegendes Arbeitsmittel. Da für $X \subseteq G$ und $Y \subseteq M$ stets $X'' = X'$ und $Y'' = Y'$ gilt, kann man alle Begriffe des Kontextes (G, M, I) in der Form (X', X') bzw. (Y'', Y') gewinnen, diese Methode ist jedoch sehr aufwendig und wird darum nur bei der Bestimmung einzelner Begriffe angewandt. Für die Aufstellung des ganzen Begriffsverbandes ist es besser, zunächst alle Begriffsumfänge (Begriffsinhalte) als Durchschnitte der Mengen $\{m_i'\}$ mit $m_i \in M$ ($\{g_j'\}$ mit $g_j \in G$) abzuleiten und dann durch den zugehörigen Begriffsinhalt (Begriffsumfang) zu ergänzen, diese Methode wird gerechtfertigt durch die Formel

$$Y' = \bigcap_{m \in Y} \{m_i'\} \quad (X' = \bigcap_{g \in X} \{g_j'\}).$$

Die Bestimmung eines Begriffsverbandes soll zunächst an einem Kontext erläutert werden, den F. Hundsnurscher in seinem Buch "Neuere Methoden der Semantik" (Hundsnurscher, 1970) für das Wortfeld "Gewässer" angegeben hat (Fig. 1).

	Fluß	Bach	Wasserlauf	Bachlauf	Bachgraben	Bachrinne	Bachgrabenrinne	Bachlaufgraben	Bachlaufgrabenrinne
Fluß	*	*	*	*	*	*	*	*	*
Bach	*	*	*	*	*	*	*	*	*
Wasserlauf	*	*	*	*	*	*	*	*	*
Bachlauf	*	*	*	*	*	*	*	*	*
Bachgraben	*	*	*	*	*	*	*	*	*
Bachrinne	*	*	*	*	*	*	*	*	*
Bachgrabenrinne	*	*	*	*	*	*	*	*	*
Bachlaufgraben	*	*	*	*	*	*	*	*	*
Bachlaufgrabenrinne	*	*	*	*	*	*	*	*	*

Fig. 1

Als Begriffsumfänge der Form $\{m\}'$ mit $m \in M$ hat man offenbar

- $\{\text{fließend}\}' = \{\text{Fluß, Bach, Kanal, Graben}\},$
- $\{\text{stehend}\}' = \{\text{See, Tümpel, Teich, Becken}\},$
- $\{\text{natürlich}\}' = \{\text{Fluß, Bach, See, Tümpel}\},$
- $\{\text{künstlich}\}' = \{\text{Kanal, Graben, Teich, Becken}\},$
- $\{\text{groß}\}' = \{\text{Fluß, Kanal, See, Teich}\},$
- $\{\text{klein}\}' = \{\text{Bach, Graben, Tümpel, Becken}\}.$

Diese Mengen haben als Durchschnitte zwölf zweielementige und acht einelementige Begriffsumfänge sowie den leeren Begriffsumfang. Da auch die Menge aller Gegenstände ein Begriffsumfang ist, besteht der Begriffsverband des Kontextes in Fig.1 aus genau 28 Begriffen. Wie diese Begriffe bzgl. der Relation "Unterbegriff-Oberbegriff" geordnet sind, zeigt das Liniendiagramm in Fig.2 an: Die kleinen Kreise des Diagramms stellen die Begriffe dar und die Strecken zwischen ihnen die Ordnungsrelation, so bedeutet ein aufsteigender Streckenzug von einem Begriff β_1 zu einem Begriff β_2 , daß β_1 ein Unterbegriff von β_2 ist. Ist ein kleiner Kreis mit einem Gegenstand g bzw. Merkmal m bezeichnet, so stellt er den Begriff $\gamma g := \{g\}'$ bzw. $\mu g := \{m\}'$ dar. Diese Bezeichnung ermöglicht, für einen beliebigen Begriff (A, B) Umfang und Inhalt aus dem Diagramm abzulesen, denn es gilt: $A = \{g \in G \mid \gamma g \leq (A, B)\}$ und $B = \{m \in M \mid \mu m \geq (A, B)\}$; so steht z.B. der kleine Kreis in der mittleren Reihe ganz links für den Begriff mit dem Umfang {Fluß, Bach} und dem Inhalt {fließend, natürlich}. Ebenfalls kann man den Kontext im bezeichneten Liniendiagramm wiederfinden, da $g \mu m$ äquivalent zu $\gamma g \leq \mu m$ ist. Diese Tatsache wird vor allem für die Kontrolle eines zu einem Kontext aufgestellten Liniendiagramms genutzt; das Diagramm ist genau dann korrekt, wenn es einen vollständigen Verband darstellt und zwar so, daß jedes Verbandselement Supremum bzw. Infimum von Elementen in der Form γg bzw. μm ist und wenn



Fig. 2

$g \mu m$ äquivalent zu $\gamma g \leq \mu m$ ist. Die beschriebenen Zusammenhänge folgen aus einem grundlegenden Satz über Begriffsverbände, der deshalb hier noch einmal aufgeführt werden soll.

Satz (Wille 1982): Für einen Kontext (G, M, I) ist $BV(G, M, I)$ ein vollständiger Verband, in dem sich Infimum und Supremum folgendermaßen beschreiben lassen:

$$\bigwedge_{i \in T} (A_i, B_i) = (\bigcap_{i \in T} A_i, \bigcup_{i \in T} B_i), \quad \bigvee_{i \in T} (A_i, B_i) = (\bigcup_{i \in T} A_i, \bigcap_{i \in T} B_i)$$

Allgemein ist ein vollständiger Verband V genau dann isomorph zu $BV(G, M, I)$, wenn Abbildungen $\gamma: G \rightarrow V$ und $\mu: M \rightarrow V$ existieren, so daß γg supremum-dicht in V ist (d.h. $V = \bigvee \{X \mid X \subseteq \gamma G\}$), μm infimum-dicht in V ist (d.h. $V = \bigwedge \{X \mid X \subseteq \mu M\}$) und $g \mu m$ äquivalent zu $\gamma g \leq \mu m$ für alle $g \in G$ und $m \in M$ ist; insbesondere gilt $V \cong BV(V, V, \leq)$.

	fließend	natürlich	groß	klein	künstlich	stehend
fließend	x	x	x	x	x	x
Bach	x	x	x	x	x	x
Fluß	x	x	x	x	x	x
See	x	x	x	x	x	x
Kanal	x	x	x	x	x	x
Teich	x	x	x	x	x	x
Graben	x	x	x	x	x	x
Tümpel	x	x	x	x	x	x
Becken	x	x	x	x	x	x
See	x	x	x	x	x	x
Fluß	x	x	x	x	x	x
Teich	x	x	x	x	x	x
Graben	x	x	x	x	x	x
Becken	x	x	x	x	x	x
See	x	x	x	x	x	x
Fluß	x	x	x	x	x	x
Teich	x	x	x	x	x	x
Graben	x	x	x	x	x	x
Becken	x	x	x	x	x	x

Fig. 3

Zur Diskussion weitere Methoden der formalen Begriffsanalyse empfiehlt es sich, größere Kontexte zu betrachten. Fig.3 gibt einen Kontext zum Wortfeld "Gewässer" wieder aus der Schrift "Kontrastive Untersuchung von Wortfeldern im Deutschen und Englischen" von G.L. Karcher (Karcher 1979). Dieser Kontext hat schon so viele Begriffe, daß es nicht mehr so leicht ist, den zugehörigen Begriffsverband zu bestimmen. Für derartige Kontexte hat sich folgende gestufte Bestimmungsmethode (Wille 1984, S. 40) bewährt: Zunächst zerlegt man die Merkmalsmenge M in Teilmengen M_1, M_2, \dots, M_k und bestimmt die Begriffsverbände $(G, M_i, I \cap G \times M_i)$ für $i = 1, 2, \dots, k$; dann erhält man (bis auf die Isomorphie) den Begriffsverband (G, M, I) , indem man im direkten Produkt der $BV(G, M_i, I \cap G \times M_i)$ ist, der g in seinem Umfang hat.

Will man die gestufte Bestimmungsmethode auf den Kontext in Fig.3 anwenden, muß man zunächst überlegen, wie man die Menge aller Merkmale geeignet unterteilt. Obwohl es sich empfiehlt die Unterteilung so zu wählen, daß die einzelnen Begriffsverbände möglichst klein werden, sollten inhaltliche Kriterien für die Unterteilung im Vordergrund stehen. Da das gestufte Liniendiagramm das sich aus der gestuften Bestimmungsmethode ergibt, die einzelnen Begriffsverbände verschieden stark hervorhebt, ist ein Gesichtspunkt, die Merkmale nach ihrer Wichtigkeit zu gruppieren. G.L. Karcher hat in seiner Untersuchung die Merkmale (Noeme) nach der Häufigkeit ihres Vorkommens angeordnet, so daß eine Unterteilung nahegelegt wird, die die ersten k Merkmale zu einer Teilmenge zusammenfaßt und die restlichen Merkmale zu einer zweiten Teilmenge. Bezieht man noch den Kontext "waters" in Fig.4 (Karcher 1979) in die Begriffsbestimmung mit ein, so drängt sich folgende Zweiteilung auf:

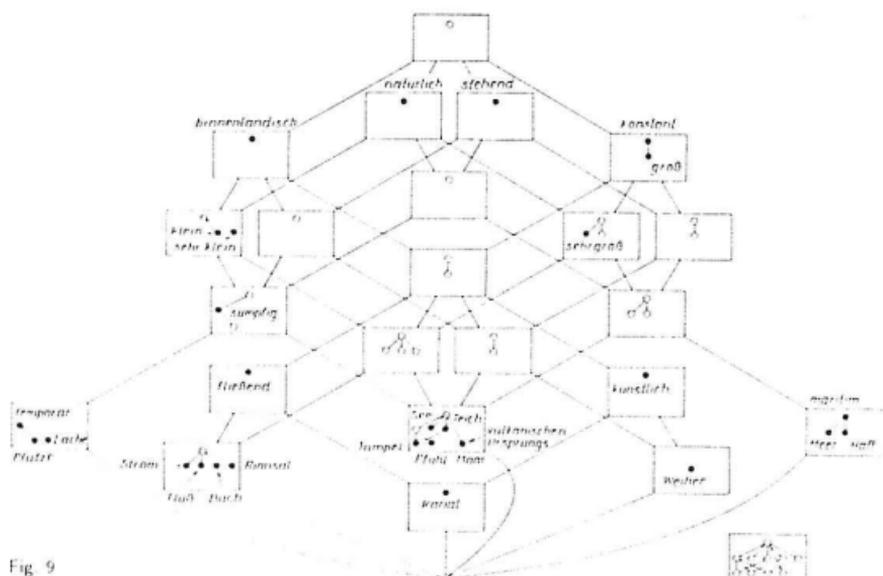


Fig. 9

Was sieht man nun an den Liniendiagrammen in Fig. 9 und 10. Sicherlich kann eine umfassende Interpretation nur von einem Sprachwissenschaftler geleistet werden, doch soll wenigstens beispielhaft aufgezeigt werden, was man an den Diagrammen ablesen kann. Stellen wir zunächst fest, daß der Kontext in Fig. 3 genau 53 Begriffe zuläßt, von denen 15 die vorgegebenen Gegenstandsnamen tragen. Allgemein kann jeder Begriff durch eine Kombination von Merkmalsnamen beschrieben werden; so stellt z.B. der kleine Kreis in der Mitte des Liniendiagramms in Fig. 10 den Begriff "großes, binnländisches Gewässer" dar, der als Umfang {Fluß, Tümpel, Teich, Weiher} hat. Will man den Begriff finden, der dem oft umfassenderen Gebrauch des Wortes "See" für Gewässer wie Teich, Maar und

Weiher entspricht, so hat man das Supremum der Begriffe dieser Gewässer zu bilden, was den Begriff "konstantes, binnländisches, stehendes Gewässer" mit dem Umfang {Pfuhl, Tümpel, Teich, Weiher, See, Maar} ergibt. Als feinere Unterteilung der Gewässer fällt folgende Klasseneinteilung ins Auge, deren Klassen Begriffsumfänge sind:

- temporäres Gewässer: Lache, Pflütze
- fließendes, natürliches Gewässer: Rinnsal, Bach, Fluß, Strom
- konstantes, binnländisches, stehendes, natürliches Gewässer: Pfuhl, Tümpel, Teich, See, Maar
- künstliches Gewässer: Kanal, Weiher
- maritimes Gewässer: Haff, Meer

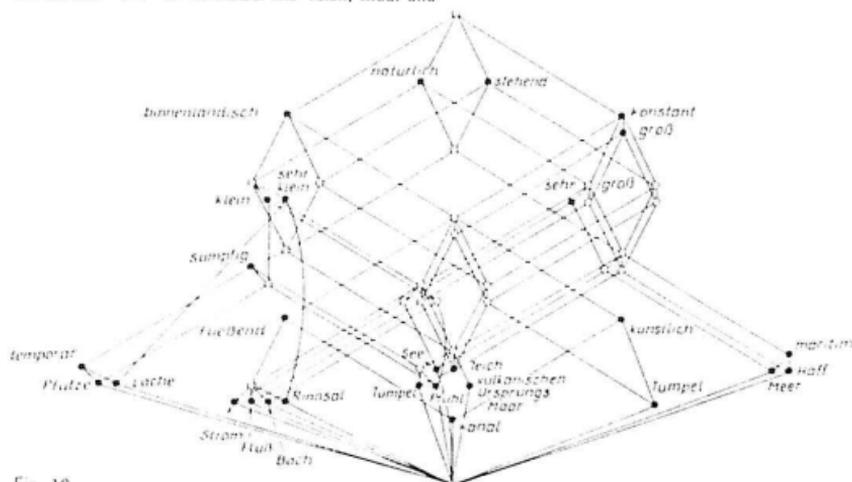


Fig. 10

An den Liniendiagrammen sind alle Implikationen zwischen den aufgeführten Merkmalen ablesbar (vgl. Ganter / Wille 1986). So erkennt man, daß "fließendes Gewässer" ein Unterbegriff von binnenländischem Gewässer" ist und somit das Merkmal "fließend" stets das Merkmal "binnenländisch" zur Folge hat. Da das Infimum der Begriffe "kleines Gewässer" und "stehendes Gewässer" ein Unterbegriff von "sumpfigem Gewässer" ist, implizieren die Merkmale "klein" und "stehend" zusammen das Merkmal "sumpfig". Die Gleichheit der Liniendiagramme in Fig. 5 und 6 besagt, daß zwischen den Merkmalen "natürlich", "künstlich", "stehend", "fließend", "binnenländisch", "maritim", "konstant" und "temporär" im Deutschen und Englischen dieselben Implikationen bestehen. Das Teildiagramm über dem Begriff "konstantes, binnenländisches, stehendes, natürliches Gewässer", das die Kantenstruktur des vierdimensionalen Würfels darstellt, zeigt an, daß die vier genannten Merkmale vollständig unabhängig voneinander sind, d.h. daß keine drei von ihnen das jeweils vierte implizieren. Es ist ein allgemeines Ziel der formalen Begriffsanalyse in einem Begriffsverband interpretierbare Teilstrukturen wie z.B. den n-dimensionalen Würfel (Booleschen Verband) herauszuarbeiten und im Liniendiagramm sichtbar zu machen.

G.L. Karcher hat in Karcher 1979 nicht nur das Vorhandensein eines Merkmals (mit +) angezeigt, sondern auch das Nicht-Vorhandensein (mit -) und die Irrelevanz eines Merkmals (mit o), so wie es in Fig. 11 wiedergegeben ist. Da in der formalen Begriffsanalyse die Bildung von Begriffen auf Merkmalen mit nur einer Ausprägung (x) beruht, hat man Merkmale mit mehreren Ausprägungen in Kollektionen von Merkmalen mit jeweils nur einer Ausprägung umzuinterpretieren. Für den Kontext in Fig. 11 liegt es nahe, ein "mehrwertiges" Merkmal jeweils in zwei

Fig. 11

"einwertige" Merkmale aufzulösen, von denen eines + als seine Ausprägung hat und das andere - als seine Ausprägung. Das Merkmal "natürlich" ist dann offenbar gleichwertig mit dem Merkmal "künstlich + " und das Merkmal "sumpfig + " gleichwertig mit dem Merkmal "klar - ". Zu neuen Begriffsumfängen, die der Kontext in Fig. 3 nicht besitzt führen die Merkmale "sehr groß - ", "sehr klein - ", "groß - " und "klein - ". Um den Begriffsverband zum "mehrwertigen - " Kontext in Fig. 11 zu bekommen, hat man die gestufte Bestimmungsmethode mit dem Begriffsverband in Fig. 12 an Stelle des Begriffsverbandes in Fig. 7 durchzuführen.



Fig. 12

An dem Liniendiagramm in Fig. 12 sieht man, daß der mehrwertige Kontext in Fig. 11 den wenig sinnvollen Begriff "nicht sehr kleines Gewässer" mit dem Umfang {Bach, Strom, Meer, Pfütze, Pfuhl, See} zuläßt. Eine solche Begriffsbildung kann man vermeiden, wenn man die Merkmale "sehr groß", "sehr klein", "groß", und "klein" durch die einwertigen Merkmale "kleiner als klein", "kleiner als groß", "kleiner als sehr groß" ersetzt, was den Begriffsverband in Fig. 13 ergibt. Aus Sicht der begriffsanalytischen Merktheorie (Ganter et al. 1986b) ordnet dann das mehrwertige Merkmal "Größe" jedem Gegenstand einen der Werte "sehr klein", "klein", "groß", "sehr groß" zu aus der ungerichteten Ordinalskala mit dem Kontext bzw. Begriffsverband.



Fig. 13

der in Fig. 14 bzw. 15 aufgeführt ist. Bei der vorangehenden Interpretationen des Kontextes von Fig. 11 sind die Werte des Merkmals "Größe" aus der Booleschen Skala in Fig. 16, deren Begriffsverband in Fig. 17 abgebildet ist. Im Kontext von Fig. 3 wird das Merkmal "Größe" aus der Booleschen Skala in Fig. 16, deren Begriffsverband in Fig. 17 abgebildet ist. Im Kontext von Fig. 3 wird das Merkmal "Größe" durch eine Nominalskala interpretiert. Für die Unterschiede zwischen den Liniendiagrammen in

nicht klein			x	x		
klein			x			
groß				x	x	
nicht groß					x	x

Fig. 14.

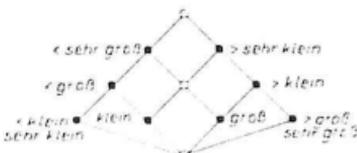


Fig. 15.

nicht klein			x	x		
klein			x			
groß				x	x	
nicht groß					x	x

Fig. 16.

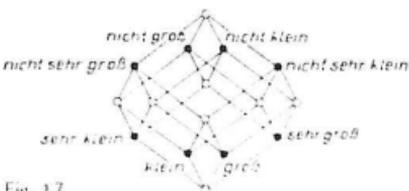


Fig. 17

Fig. 12 und 13 sind demnach die verschiedenen Interpretationen des Merkmals "Größe" verantwortlich. Bei mehrwertigen Merkmalen ist stets sorgfältig zu überlegen, durch welche begrifflichen Skalen sie interpretiert werden sollen, was dann seinen Niederschlag im Begriffsverband und dessen Liniendiagramm findet.

Abschließend sei noch einmal betont, daß die formale Begriffsanalyse exakte Festlegungen voraussetzt; dafür verhindert sie unklare Darstellungen wie z.B. die Kästchendiagramme in Karcher 1979, S. 197, deren Aussagefähigkeit und Ableitbarkeit aus den vorgegebenen Daten im Dunkeln bleibt. Ein Vorteil der mengensprachlichen Modellierung ist, daß begriffsanalytische Untersuchungen durch elektronische Datenverarbeitung effektiv unterstützt werden können, wozu verschiedene Computer-Programme zur Verfügung stehen (Ganter 1987; Ganter et al. 1986a, Luksch et al. 1986).

Literatur

- Birkhoff, G. 1967: Lattice theory. Third edition. Amer. Math. Soc., Providence, R.I., 1967
- Din 1979: (Deutsches Institut für Normung): DIN 2331: Begriffe und Benennungen, Allgemeine Grundsätze. Beuth-Verlag, Köln 1979.
- Din 1980: (Deutsches Institut für Normung): DIN 2331, Begriffssysteme und ihre Darstellung. Beuth Verlag, Köln 1980.
- Ganter, B. 1987: Algorithmen zur Formalen Begriffsanalyse. In: B. Ganter, R. Wille, K.E. Wolff (Hrsg.): Beiträge zur Begriffsanalyse. B.I.-Wissenschaftsverlag Mannheim 1987.
- Ganter, B. / Rindfrey, K. / Skorsky, M. 1986a: Software for formal concept analysis. In: W. Gaul, M. Schader (ed.): Classification as a tool of research. North-Holland Amsterdam 1986, S. 161-167.
- Ganter, B. / Stahl, J. / Wille, R. 1986b: Conceptus measurement and many-valued contexts. In: W. Gaul, M. Schader (ed.): Classification as a tool of research. North-Holland, Amsterdam 1986, S. 169-176.
- Ganter, B. / Wille, R. 1986: Implikationen und Abhängigkeiten zwischen Merkmalen. In: P.O. Degens, H.-J. Hermes, O. Opitz (Hrsg.): Die Klassifikation und ihr Umfeld. Index Verlag, Frankfurt 1986, S. 171-185.
- Hundsniischer, F. 1970: Neuere Methoden der Semantik. Max Niemeyer Verlag, Tübingen 1970.
- Karcher, G.L. 1979: Kontrastive Untersuchungen vor Wortfeldern im Deutschen und Englischen. Peter lang, Frankfurt 1979.
- Luksch, P. / Skorsky, M. / Wille, R. 1986: On drawing concepts lattices with a computer. In: W. Gaul, M. Schader (ed.): Classification as a tool of research. North-Holland Amsterdam, 1986, S. 269-274.
- Wagner, H. 1973: Begriff. In: Handbuch philosophische Grundbegriffe. Kösel, München 1973, S. 191-209.
- Wille, R. 1982: Restructuring lattice theory: an approach based on hierarchies of concepts. In: I. Rival (ed.): Ordered sets. Reidel, Dordrecht-Boston (1982), S. 445-470
- Wille, R. 1984: Liniendiagramme hierarchischer Begriffssysteme. In: H.H. Bock (Hrsg.): Anwendungen der Klassifikation: Datenanalyse und numerische Klassifikation. Index Verlag, Frankfurt 1984, S. 32-51.
- Wille, R. 1987: Bedeutungen von Begriffsverbänden. In: B. Ganter, R. Wille, K.E. Wolff (Hrsg.): Beiträge zur Begriffsanalyse, B.I.-Wissenschaftsverlag, Mannheim 1987.

DIE EXTREMWERTANALYSE IM ZEITBEREICH
ALS MÖGLICHKEIT DER AUTOMATISCHEN SPRACHERKENNUNG

Thomas und Patrick Schweisthal [1]

Arbeitskreis Spracherkennung, Sprachgenerierung und Phonetische Datenbanken
der Gesellschaft für Linguistische Datenverarbeitung (GLDV)
Institut für Phonetik und sprachliche Kommunikation
Universität München
Schellingstraße 3 V6
8000 München 40

Zukünftige Computersysteme sollten in der Lage sein, gesprochene Sprache ohne die bisher üblichen Einschränkungen (wie lange Trainingsphasen für Sprecher bei geringem Wortschatz) zu erkennen. Der Beitrag des Phonetikers hierzu kann eine Transkriptionsmaschine sein, die gesprochene Sprache in phonetische Zeichen umsetzt. Im nachfolgenden Artikel wird unter Berücksichtigung neuerer Erkenntnisse aus der Physiologie des Gehörs die Extremwertanalyse im Zeitbereich als Möglichkeit der akustischen Sprachschallanalyse dargestellt. Inwiefern sie eine Alternative zu gängigen Verfahren darstellt, wird sich bei genauerer Erprobung zeigen.

Seit langer Zeit bemüht sich die Forschung um ein zufriedenstellendes Verfahren zur Automatischen Spracherkennung. Für isolierte Worterkennung ist man diesem Ziel schon sehr nahe gekommen, während die Untersuchung fortlaufender Rede bisher unter unverhältnismäßigem Aufwand (Aktas/Kämmerer/Küpper/Lagger 1986 S. 9 "Trotz vieler neuer Ansätze auf dem Gebiet der Spracherkennung (Verwendung von Markov-Modellen, auf phonetischem Wissen basierende Systeme) ist der klassische Mustererkennungssatz mit dynamischer Programmierung immer noch von vorrangiger Bedeutung... Bei großen Wortschätzen (mehr als 1000 Wörter) ist es jedoch schwierig, mit diesen sehr rechenintensiven Verfahren Echtzeitbedingungen einzuhalten.") noch unbefriedigende Ergebnisse vorzuweisen hat.

Wir glauben, daß man mit den bisherigen Mitteln der Spektralanalyse, in die die größten Hoffnungen gesetzt worden sind und noch werden, noch nicht den richtigen Weg einer akustischen Sprachanalyse eingeschlagen hat; also werden auch andere Bereiche des Signalphonetischen Bandes (Tillmann 1980, S. 193) zur Spracherkennung herangezogen, obwohl nach unserer Auffassung das akustische Signal die notwendige Information für eine am Hören orientierte Sprachschallanalyse enthält (Pomino 1978, S. 22 f. "Meines Erachtens ist der Produktionsmechanismus nicht die Grundlage der Perception, der Perzeptionsmechanismus vielmehr ist auditorisch schon so ausgestattet, daß sich der Zusammenhang zwischen Produktion und Perception ontogenetisch formen kann.")

Die Spektralanalyse zerlegt das Zeitsignal in seine Bestandteile verschiedener Frequenzbereiche, deren Energiekonzentrationen festzustellen sind. Daß aber das menschliche Ohr nicht nur wie die Filterbank eines Sonographen funktioniert, ist längst nachgewiesen; lediglich für Frequenzanteile über 5000 Hz mag das in etwa zutreffen. Für den Frequenzbereich unter 500 Hz wird eine direkte zeitliche Wahrnehmung angenommen, für den verbleibenden Bereich von etwa 500 bis 5000 Hz eine gemischte Zeit-Ort-Wahrnehmung (Volley-Theory) (vgl. Schouten 1938, 1940; Ainsworth 1976, S. 31 ff.; Spreng 1975, S. 138). Dem akustischen Rezeptionsapparat Ohr also wird die Spektralanalyse allein nicht gerecht, zumal bekannte Physiologen der Sprachfor-

schung den Rat gegeben haben, sich verstärkt mit der Zeitanalyse zu befassen (Spreng 1975, S. 138 "Alle diese Befunde zeigen, daß neben der reinen Frequenzerkennung durch entsprechende Ortsabbildung eine Zeitanalyse vor allem bei der Erkennung der Sprache eine entscheidende Rolle spielt, so daß man mit Recht als die hervorstechendste Eigenschaft des Gehörs die Analyse zeitlicher Muster bezeichnen kann.")

Schon im Jahre 1951 ermittelte man Nulldurchgangsdichten des Originalsignals und des differenzierten Signals (also die Extremwertdichte des Ausgangssignals), (Chang / Pohl / Essigmann 1951), nachdem zuvor Licklider und Pollak (Licklider / Pollak 1948; Ainsworth 1976, S. 97 f.) in Untersuchungen mit manipulierten Sprachsignalen darauf hingewiesen hatten, daß ein Großteil sprachlicher Information in den Extremwertabständen liegen müsse.

Peterson stellte des weiteren die bemerkenswerte Proportionalität von Nulldurchgangsdichte zum ersten Formanten und Extremwertdichte zum zweiten Formanten der Spektralanalyse heraus (Peterson 1951). Es bestand aber weithin Einigkeit, daß diese Verfahren im Zeitbereich zu eher schlechteren Ergebnissen führten als die Spektralanalyse (Kirstein 1977, S. 11). Eine Untersuchung zum Vergleich zwischen Formantextraktion durch Filterbanke und andererseits durch Nullstellenanalyse lieferte Scarr 1968, wenn auch spätere Untersuchungen dieser Ansicht widersprechen (Kirstein 1977, S. 4 ferner S. 99). Kirstein, der selber nach Nullstellen analysiert, empfiehlt, Extremwertabstände in jedem Fall in weitere Untersuchungen einzubeziehen – weiteres dazu siehe Literaturverzeichnis.

Unter Berücksichtigung dieser Erkenntnisse bieten unserer Ansicht nach die Extremwerte den besten Ansatzpunkt für die Zeitanalyse. Im Gegensatz zur Nullstellenanalyse verändert sich die Extremwertlage nicht mit einem veränderten Berechnungsverfahren der Nulllinie, die ja den über einen bestimmten Zeitraum gemittelten Wert wiedergibt.

Früher ist allerdings, wie man dem akustischen Apparat Ohr bei der Analyse am ehesten gerecht werden kann. Erhalten bleiben sollen bei einem solchen Verfahren jedenfalls die ursprünglichen Extremwertabstände, was bei herkömmlichen (linearen) Hoch- und Tiefpassen nicht der Fall ist, wohl aber bei einem System, das die Kurve in Grob- und Fein-

strukturen aufteilt, wobei die Grobstruktur den Bereich der rein zeitlichen Wahrnehmung (bis etwa 500 Hz), die Feinstruktur den der Volley-Theory (etwa von 500 bis 5000 Hz) vertritt. Das läßt sich aber über ein herkömmliches Glättungsverfahren erreichen, welches die einzelnen Punkte der digitalisierten Kurve neu definiert als die Mittelwerte ihrer Umgebungspunkte der Originalkurve, ein nichtlineares Filterverfahren also. Es ergibt sich eine neue Kurve, für die in einem festgesetzten zeitlichen Bereich die Gesamtdifferenz der neuen Werte zu den Originalwerten gleich null ist, wobei die neuen Werte möglichst wenig von ihren Nachbarwerten abweichen. Die Differenz aus der so gewonnenen Grobstruktur und der Originalkurve ergibt die Feinstruktur. Wie weit der Mittelungszeitraum am sinnvollsten zu setzen ist, ist experimentell zu bestimmen. Bei Grob- und Feinstruktur ist nebeneinander der zeitliche Abstand benachbarter Extremwerte festzuhalten. Die Reziprokwerte der Abstände, die "Extremwertfrequenzen", in Abhängigkeit von der Signalzeit abgetragen, ergeben ein Sonagramm nicht unähnliches Bild, weshalb wir Häufungen bestimmter Extremwertfrequenzen als die wirklichen Formanten, die "Realformanten" bezeichnen möchten. Ein vergleichbares Verfahren wandte Janet M. Baker (Baker 1974) an.

Wir konnten, natürlich unter Vernachlässigung mancher Unregelmäßigkeiten - entsprechend der Formantanalyse im Spektralbereich - folgende qualitative Zuordnungen der Zungenlage und Zungenhöhe formulieren:

- je tiefer die Zunge, desto höher die Extremwertfrequenz der Grobstruktur (des ersten Realformanten)
- je weiter vorn die Zunge und je höher die Zunge, desto höher die Extremwertfrequenz der Feinstruktur (des zweiten und eventuell weiterer Realformanten)

Für die gesamte hintere Reihe (u bis hinteres tiefes a) bedeutet das, daß sie durch den ersten Realformanten hinreichend charakterisiert ist.

Für den Zentralbereich gilt, daß sich die Werte des ersten und zweiten Realformanten einander annähern. Auf dieser Basis wollen wir eine "Transkriptionsmaschine" bauen.

Unsere Bemühungen werden sich dahin richten, jedem möglichen Wertepaar der Realformanten 1 und 2 ein Schriftzeichen zuzuordnen und hierzu eindeutige, lückenlose Definitionsbereiche zu finden. Jedes Segment ist somit durch ein Wertepaar definiert (Objektive Transkription). Im zweiten Arbeitsgang wird unter Berücksichtigung von Koartikulationserscheinungen, artikulatorischer Nachbarschaft und psychoakustischen Phänomenen eine "Subjektive Transkription" versucht, die der üblichen Unterrichtstranskription entsprechen soll. (Beispiel für eine objektive Transkription eines /a/:

[i|e|ä|a|...|a|];

die ersten drei Zeichen /i|e|ä/ stellen eine charakteristische Transition von /t/ zu /a/ dar, lassen sich also subjektiv als t vor a transkribieren. Die objektive Transkription transkribiert im stimmhaften Bereich periodenweise, im stimmlosen Bereich segmentweise.

Ist erst die "Hürde" einer brauchbaren unkomplizierten Transkriptionsmaschine genommen, so wird unter Berücksichtigung weiterer sprachlich relevanter Gegebenheiten wie Tonhöhe und Intensitätsbetrachtung über längere Zeiträume eine weitere Unterteilung auf Silbenebene und in der Endstufe durch weitergehende Anwendung von Methoden der Künstlichen Intelligenz eine Verschriftung

schließlich möglich sein. Für die Spracherkennung ergeben sich folgende Einzelschritte:

1. Digitalisierung des Sprachsignals
2. Objektive Transkription
3. Subjektive Transkription (Verschriftung nach Lautschrift)
4. Versilbung (unter Berücksichtigung längerer Intensitätsverläufe und der Tonhöhe)
5. Syntaktische Analyse und schließlich
6. Verschriftung

Anmerkung

[1] Die Autoren sind auch unter folgender Privatadresse zu erreichen: Thomas und Patrick Schweithal, Edenthalweg 31, 8069 Rohrbach, Tel. 08442-8444.

Literatur

- Ainsworth, W.A. 1976: Mechanisms of Speech Recognition, 1976
- Aktas, A. / Kämmerer, B. / Küpper, W. / Lagler, H. 1986: Spracherkennung für große Wortschätze mit schnellerer Zeitanpassung. In NTG-Fachberichte 94, 1986
- Baker, J.M. 1974: A new time domain analysis of fricatives and stop consonants. In Proc. IEEE Symp. Speech Recognition, Pittsburgh 1974, S. 134-141
- Chang, S.H. / Phil, G.E. / Essigmann, M.W. 1951: Representation of Speech Sounds and some of their statistical properties. In Proc. Inst. Radio Engrs. 39, 1951, S. 147-152
- Ito, M.R. 1972: Relationship between zero-crossing measurements for speech analysis and recognition. J. Acoust. Soc. Am. 51, 1972, S. 2061-2062
- Keller, Th., von 1976: Die Kennzeichnung von Sprachlauten durch Spektrum, Autokorrelationsfunktion und Nulldurchgangsabstände. Nachrichtentechn. Z. 20, 1976, S. 195-205
- Keidel, W.D. 1970: Kurzgefälltes Lehrbuch der Physiologie. Thieme, Stuttgart, 1970 (3. Aufl. 1973)
- Kirstein, M. 1977: Untersuchung zur Verteilung von Nulldurchgangsabständen bei Sprachsignalen, Buske, Hamburg 1977
- Kotten, K. 1971: Nulldurchgangszählung und Nulldurchgangsdichte in DAWID-II. Beiträge zur automatischer Spracherkennung. Forschungsberichte IPK, Band 36, Hamburg 1971, S. 81-111
- Licklider, J.C.R. / Pollak, I. 1948: Effects of differentiation, integration, and infinite peak clipping upon the intelligibility of speech. J. Acoust. Soc. Am. 20, 1948, S. 42-51
- Meyer-Eppler, W. 1969: Grundlagen und Anwendungen der Informationstheorie. Zweite Auflage Berlin, 1969
- Peterson, E. 1951: Frequency detection and speech formants. J. Acoust. Soc. Am. 23, 1951, S. 668, 674
- Peterson, G.E. / Hanne, J.R. 1965: Examination of two different formant estimation techniques. J. Acoust. Soc. Am. 38, 1965, S. 224-228
- Pompino, B. 1978: Hemisphärenunterschiede und der "phonetic speech processor". In Forschungsberichte 9, 1978 Institut für Phonetik und Sprachl. Kommunikation, München
- Reddy, D.R. 1967: Computer recognition of connected speech. J. Acoust. Soc. Am. 42, 1967, S. 329-347
- Scarr, W.A. 1968: Zero crossings as a means of obtaining spectral information in speech analysis. IEEE Trans. Audio Electroacoust. Au-16, 1968, S. 247-255
- Schouten, J.F. 1940: The perception of pitch. Philips techn. Rev. 5, 1940, S. 286-294
- Spreng, M. 1975: Mechanisch-elektrische Wandlung: Dynamik der Sinneszellen des Gehörs, Adaption und selektives Verhalten. In: Keidel W.D. (Hrsg.): Physiologie des Gehörs. Stuttgart 1975, S. 102-152

ERKLÄRUNG EINIGER TECHNISCHER GRUNDBEGRIFFE ZUR AUTOMATISCHEN SPRACHERKENNUNG

1. Filter:

Die Abbildung H eines Signals $s(t)$ auf ein Signal $s'(t)$

$$s'(t) : s' = H(s)$$

2. Formanten:

Energiekonzentrationen in verschiedenen Frequenzbereichen eines Sprachsignals

3. Hochpaß:

Filter mit der Eigenschaft, oberhalb einer Grenzfrequenz f_0 liegende spektrale Anteile unverändert zu lassen und Anteile unterhalb dieser Frequenz zu unterdrücken (Gegenteil: Tiefpaß)

4. Koartikulation:

Begriff aus der Phonetik; bezeichnet die übergreifende Artikulation (Beispiel: "ko": die Lippenrundung des "o" setzt schon beim "k" an).

5. Lineares Filter:

Ein Filter H ist immer linear, wenn folgendes gilt:

$$\begin{aligned} S'_1 &= H(S_1); S'_2 = H(S_2) \\ \implies S'_1 + S'_2 &= H(S_1 + S_2) \end{aligned}$$

6. Signalphonetisches Band:

Umfaßt sämtliche Stationen des Kommunikationsprozesses von Sprecher zu Hörer (z.B. akustische Vorgänge, neurolog. Prozesse).

7. Sonogramm:

Bildliche Darstellung des Sprachsignals durch den Sonographen, der das komplexe Signal mit Hilfe von Filterbänken spektral zerlegt. (Der x -Wert des Diagramms entspricht der Zeit, der y Wert den Frequenzen, der Schwärzungsgrad der Stärke der Energiekonzentration im bezeichneten Frequenzbereich.)

8. Zeitanalyse:

Untersuchung der Sprachschalloszillation im Verlauf der Zeit $s(t)$.

GLDV-Jahrestagung

ANALYSE UND SYNTHESE GESPROCHENER SPRACHE

Jahrestagung der Gesellschaft für Linguistische Datenverarbeitung

in Zusammenarbeit mit der Fachgruppe

Sprachanalyse der Gesellschaft für Informatik

Die Jahrestagung 1987 der Gesellschaft für Linguistische Datenverarbeitung unter der Thema *Analyse und Synthese gesprochener Sprache* fand vom 4. bis 6. März im Vortragssaal des Landesmuseums Bonn statt. Die Tagung war von Gerd Willée und Burkhard Schaefer (Organisationskomitee) sowie COLING-erprobten und bewährten Helfern vom IKP, Bonn vorbereitet worden. Das Programmkomitee bildeten H.G. Tillmann (Vorsitz), Winfried Lenders, Bernd S. Müller und Klaus G. Schweisthal. Der Tagungsband ist als Band 9 der Reihe "Linguistische Datenverarbeitung" bei Olms erschienen (Tillmann, H.G., Willée, G. (Hrsg.) *Analyse und Synthese gesprochener Sprache*, Hildesheim: Olms, 1987) und war bereits zu Tagungsbeginn den Teilnehmern verfügbar.

Folgende Beiträge wurden vorgestellt:

B. Kaspar, Lochschmidt, H. Wolf: Spracherkennung und -synthese in der Telekommunikation.

W. Menzel, I. Starke: Wortformenanalyse und Schrifttext-gesteuerte Sprachsynthese.

S. Bauer et al.: Graphem-Phonem-Umsetzung: Lexikon versus Regelkatalog.

K. Huber: Sprachsynthese ab Text.

J.P. van Hemert, U. Adriaens-Porzig, L.M.H. Adriaens: Speech Synthesis in the SPICOS-Project

R. Sommer: Segmentierung von Sprachmaterial: Plädoyer für Koartikulation berücksichtigende Segmente.

G.F. Meier: Semantische Probleme beim Sprachverstehen.

E. Nöth, A. Batliner, R. Lang, W. Oppenrieder: Automatische Grundfrequenzanalysen und Satzmodusdifferenzierung.

R. Bannert, K.-J. Engelbert, B. Keck: Die Intonation der Spracherkennung. Das Black-Board-Model des Ikaros-Projekts.

D. Mergel, H. Ney, A. Noll, A. Paeseler: Wechselwirkungen zwischen akustischer Erkennung und linguistischen Einschränkungen im SPICOS-Projekt.

G. Niedermair: Linguistische Analyse im System SPICOS.

U. Ehrlich, H. Niemann: Semantische Verarbeitung.

A. Boner: Vier Jahre computerunterstützte Sprachübersetzung - Erfahrungsbericht.

U. Jekosch: Umsetzung von Phonemwörtern in ihre orthographische Form für unbeschränkten deutschen Wortschatz.

G. Thutman: Der Einsatz semantischer Verfahren in sprachverstehenden Systemen.

H. Geisler, K. Kotten, K.G. Schweisthal: Velare a-Realisationen im Bairischen.

I. Hertrich: Interindividuelle Variabilität in Langzeitstrukturen des Sprachsignals.

G. Dreckschmidt: The Linguistic Component in the Speech Understanding System SPICOS.

B. Möbius: Untersuchungen zu mikroprosodischen Grundfrequenzvariationen im Deutschen.

T. Goldbeck, W.F. Sendmeier: Phonetische Manifestation der Intonation bei gleichzeitiger Realisierung unterschiedlicher linguistischer Funktionen.

E. Marschall, D. Schmidbauer: Berücksichtigung von akustisch-phonetischen Ambihütäten bei einer bottom-up Worthypothesen-Generierung.

A. Batliner: Der Einsatz der Diskriminanzanalyse zur Prädiktion des Modus.

K. Kohler: Funktionen von $F_{0,1}$ -Gipfeln im Deutschen.

C. Reuter: Mikroprosodie und Synthese von Intonation.

R. Gaitenberg: Zur Stilisierung von Intonationskonturen.

K. Fellbaum, L. Fliegner, M. Rusch: Erzeugung einer einfachen Wort- und Satzprosodie durch Modifikation der Sprachgrundfrequenz und der Lautdauer.

U. Heid, B. Weck, D. Roesner: Generierung deutscher Sätze aus semantischen Repräsentationen im Sinne von E. Steiner.

Eine Panel-Diskussion "Automatische Segmentierung fließender Rede" wurde von H.G. Tillmann moderiert.

BEMERKUNGEN ZU DEN VORTRÄGEN DER JAHRESTAGUNG 1987 DER GLDV

Vom 4. bis 6. März fand in Bonn die Jahrestagung der Gesellschaft für linguistische Datenverarbeitung (GLDV) statt.

Die Gesellschaft hat für diese Tagung die Analyse und Synthese lautlicher Sprachereignisse gewählt, eine Thematik, die selbst innerhalb der LDV stiefmütterlich behandelt wird, ja lange Zeit nur im Rahmen der Akustik oder Nachrichtentechnik zuhause war. Verstehen von gesprochener Sprache ist nach wie vor ein noch wenig erforschter Bereich der Kommunikation, so daß von einer Simulierung natürlicher Prozesse nicht gesprochen werden kann. Bei künstlicher Analyse durch technische Anlagen spielen letztlich auch die physiologischen und psychologischen Faktoren der menschlichen Perzeption und Lautvorstellung kaum eine Rolle, da nur die physikalischen Parameter der gesprochenen Sprache, d.h. der Vielzahl der Varianten individueller Spechakte, verarbeitet werden müssen. Die Auffindung der für die Umkodierung in Grapheme nötigen Invarianten (bei einzelnen fälschlich als Phoneme) setzt die Segmentierung voraus, die der jeweiligen Orthographie entspricht. Das Gehirn löst dieses Problem durch zahlreiche Rückkopplungs- und Assoziationsprozesse, wie durch präprozepptive Referenzen, Summationen und Inhibitionen, so daß sprachinterne Artikulationsgewohnheiten zur selektiven, d.h. relevanten Hörgewohnheit führen, die ihrerseits zur graphematischen Zuordnung über trial-error-Prozesse die Basis ist. Da eine (Mutter-)Sprache großenteils im Satzverband mit bedeutungstragenden Einheiten (Formativen) erlernt wird, ist der Verstehensprozeß weitgehend mit semantischen Rückkopplungen verknüpft, die ihrerseits durch situative und Wissensinformationen verstärkt, korrigiert oder präzisiert werden. Auf der Tagung wurde von mehreren Forschungsteams eine Lösungsmöglichkeit vorgelegt, die Silbe, Wort oder "Semantik" mehr oder minder einbeziehen will. Allerdings bauen die meisten Versuche auf DUDEN-Silben auf bzw. legen Silbengrenzen ad hoc fest. Silbentheoretische Erörterungen wurden – sicher auch aus Zeitgründen – nicht diskutiert, obwohl gerade die Silbengrenze eine umstrittene Frage ist. Die Silbentrennung – dies zeigen Ausführungen der SPICOS-Bearbeiter – wird von der Koartikulation überlagert, die sehr unterschiedliche Zeitstrecken bestreicht.

So wichtig und interessant die in den 60er Jahren durchgeführten Diphonuntersuchungen sind, so bleiben sie doch isolierte Kunstprodukte, da die Koartikulation oft silbengrenzenüberschreitend wirksam wird, was in einzelnen Sprachen Vokalassimilationen, Brechungen, Palatalisierungen, Umlautungen (innerhalb des Invarianzenbereiches) erzeugt. Der Zugriff zu größeren Einheiten, mindestens Disyllabeme oder -syeme bzw. Wörter in allen morphologischen Vari-

anten, paßt sich mehr an den natürlichen Verdichtungsprozeß im Gehirn an und erlaubt größere Toleranzen. Dies zeichnet sich bei mehreren Vorträgen zur Analyse und zur Synthese deutlich ab. Man könnte besonders für Dialogsysteme sogar – dies kommt bei der Forschung der Bundespost zum Ausdruck – von der Tatsache Gebrauch machen, daß sprachliche Äußerungen weitgehend auf feste Syntagmen und Redestereotypen aufgebaut sind. Erfolgreicher als die Analyseforschung ist zweifelsohne die (auf der Tagung repräsentierte) Syntheseforschung, der auch mehr Beiträge gewidmet sind. Es scheint, daß hier auch Koartikulation und damit verknüpft der Zeitfaktor noch ein gutes Stück Arbeit verlangt, daß aber "verständliche Sprache" selbst ohne Prosodik erzeugt werden kann, da als Endabnehmer ja der Mensch mit seiner relativ großen Verstehenstoleranz steht.

Über die Transkriptionssysteme wurde auf der Tagung wenig gesagt, zumindest scheint im wesentlichen etwa die Genauigkeitsstufe von API befolgt zu werden, die jedoch im Koartikulationsbereich (Pranasalierung, Präaspiration, Diphonübergänge usw.) selbst bei europäischen Standardsprachen nicht hinreicht. Nicht ganz klar ist – so besonders bei den Zürcher Arbeiten – die Rolle bzw. Notwendigkeit morphologischer Umkodierungen, wie z.B. die Deflektion bei der Lemmatisierung. Morphologische Analyse und Synthese bieten eine Menge von Fehlerquellen, sind teilweise recht aufwendig und sind bei der massenhaften Zunahme an Internationalismen und Originalwörtern (mit Originalausdrücke und Originalschreibung) regelgestört. Es empfiehlt sich auch in der Synthese mit Vollwortformen aus dem Speicher zu arbeiten. Einige Arbeiten haben sich mit der Satzmelodik, d.h. mit Grundfrequenzanalysen und entsprechenden Synthesen befaßt, wobei richtig die Folge der Silbenkernfrequenzen berechnet wird. Sowohl die mikroprosodischen als auch konturenorientierten Berichte zeigen gute Fortschritte gegenüber früheren Ergebnissen. Als Problem erscheint hier die Auswirkung unterschiedlicher Sprechgeschwindigkeiten auf die Intonation, wie dies im SPICOS-Projekt auch für die Findung von Silbenkernen der Fall ist. Recht interessant ist in diesem Zusammenhang die Untersuchung Kohlers, die wichtige Hinweise auf die Wechselbeziehung von Betonungsveränderungen und Verschiebungen des F_0 -Gipfels gibt. Einige Beiträge betonten die Notwendigkeit, daß auch für die akustische Analyse höhere Ebenen herangezogen werden müssen. Explizit wurde in 6 Vorträgen auf die Einbeziehung der Semantik (in unterschiedlicher Definition) bei Analyse und Synthese teils theoretisch, teils auch mit konkreten Angaben verwiesen. Teils sind damit semantische Minima-Merkmale für syntaktische Operationen gemeint, teils wird auch an Lexikosemantik gedacht. Es ist aber keinem Vortragenden gelungen (möglicherweise durch zeitliche

Beschränkung bedingt) zu zeigen, wie in der Analyse mittels semantischer Entscheidungsoperationen Eindeutigkeit hergestellt wird.

Für die Satzsyntax ist natürlich eine inhaltliche (semantische) Ausgangsbasis erforderlich, wofür einige Vorschläge gemacht wurden, wenngleich nur an Modellen. Ob und wie weit das STEINER'sche Verfahren sich hierfür eignet, läßt sich erst sagen, wenn die entsprechenden Veröffentlichungen vorliegen. Unsere eigenen prädikatenlogischen Untersuchungen haben im Laufe von 20 Jahren mehrfacher Korrekturen diese Verfahrensweise als einzig mögliche bestätigt, wir bezweifeln aber, daß die Case-Grammar-Methode und die STEINER'schen Verbklassen dafür ausreichen, zumindest, wenn EUROTRA-D multilingual komplexe Informationen verarbeiten will. Es ist auf der Tagung mehrmals die Frage nach dem Zweck und der Anwendungsmöglichkeit der Forschungen gestellt worden. Unabhängig davon, wie man zu dem reklameträchtigen Terminus "Künstliche Intelligenz" stehen mag, ist sicher, daß Sachverhalte und Denkpro-

zesse in Computern nur verarbeitet werden können, wenn das semantische Problem ernsthaft, d.h. nicht nur in theoretischen Modellen und an einzelsprachliche Syntax gebunden, für Analyse, d.h. Inhalts-Verstehen (Faktenrecherchen) und für die Synthese (Dialog-Antworten) in Angriff genommen wird.

Es ist der Verdienst der Veranstalter und Gastgeber, daß diese Problemschau für alle Teilnehmer sehr instruktiv und sicher auch zu Verbesserungen anregend war. Selbst für sprachhistorische Betrachtungen sind einige Ergebnisse aufschlußreich, wie die Münchner dialektologische Untersuchung zum bayrischen *a* gezeigt hat. Die Zukunft der LDV hängt ohne Zweifel davon ab, ob die sechste Computergeneration das hält, was die fünfte nur versprochen hat – und das hängt in erster Linie von dem linguistischen Beitrag ab, denn ohne semantische Kleinarbeit lassen sich keine Sachverhalts-Informationen verarbeiten.

Prof. Dr. Georg F. Meier, Olching

Gesellschaft für
Linguistische Datenverarbeitung e. V.
z. Hd. Frau Prof. Dr. B. Endres-Niggemeyer
FH Hannover
Fachbereich BID
Hanomagstraße 8
3000 Hannover 91

Absender

Name

Vorname

Strasse

() Ort

Tele



Aufnahmeantrag für persönliche Mitgliedschaft

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

Name	Vorname	
Akadem. Grad oder Berufsbezeichnung	beschäftigt bei	Student ja/nein
Interessengebiete		
Arbeitsschwerpunkte		
Datum	Unterschrift	

Tagungsberichte

SPRACHWISSENSCHAFT IM COMPUTERZEITALTER

2. Sommerschule der Deutschen Gesellschaft für Sprachwissenschaft vom 1. bis 19. September 1986 in München

31. August 1986

Ankunft in München bei strahlendem Sonnenschein - die äußeren Umstände um die SiC (Sprachwissenschaft im Computerzeitalter) sind also vorerst einmal hervorragend!

Inbesitznahme des zugewiesenen Zimmers (den Organisatoren sei ein herzlicher Dank für ihre unermüdete Zimmersuche, die es den SiC-Teilnehmern schließlich doch ermöglichte, ihr 'müdes Haupt' irgendwo in München bequem hinzubetten).

Da war ich nun also. Mit einem kürzlich abgeschlossenen Hochschulstudium der Germanistik, Nebenfach Informatik hinter mir vielleicht ein idealer SiC-Teilnehmer. Dies würde sich nun in den kommenden drei Wochen zeigen. Was mich besonders zur Teilnahme an der SiC bewegen hatte war die Hoffnung, dort einen umfassenden Überblick über die verschiedenen Möglichkeiten der Zusammenarbeit zwischen Sprachwissenschaftlern und Informatikern, die bestehenden Ansätze der Wissensrepräsentation, den Stand der Forschung auf dem Gebiet der Sprachverarbeitung bzw. Sprachgenerierung zu bekommen.

1. September 1986

Beginn der Sommerschule. Zunächst mußten die Anmeldeformalitäten erledigt werden. Danach ein erster Stadtbummel. Da wir von den Organisatoren alle mit einem Stadtplan, einem Veranstaltungskalender und einem eigens erstellten SiC-Brevier, in dem nicht nur das Programm der SiC noch einmal vorgestellt wurde, sondern auch alles Wissenswerte über Essen, Trinken, Fortbewegung und Freizeit in München zu finden war, ausgestattet worden waren, konnte dabei nichts mehr schief gehen.

14.00 Uhr "Einführung in die Arbeit mit dem PC". Erste Kontakte mit den SiC-Teilnehmern (insgesamt waren es immerhin 260) und dann der offizielle Eröffnungsempfang mit den üblichen Reden und anschließend bairischen Buffet. Alles in allem eine gelöste und fröhliche Atmosphäre.

2. September 1986

Nun ging es richtig los! Das Programm der Sommerschule war ziemlich dicht gedrängt, aber es lag ja an einem selbst, welche und wieviele Veranstaltungen man besuchen wollte. Man hätte die Sommerschule zum "full-time"-Job werden lassen können, aber davor wurde von den Veranstaltern immer wieder gewarnt: "München ist doch auch eine sehenswerte Stadt".

Das Programm sah folgendermaßen aus:

(1) Programmierkurse

Programmierkurse gab es in PROLOG und LISP jeweils für Anfänger und Fortgeschrittene. Die Kurse umfaßten sieben Doppelstunden für die Vermittlung der Theorie und weitere sieben Doppelstunden Programmierpraktikum. Außerdem standen die beiden "Computerräume" auch sonst zu Übungszwecken zur Verfügung.

Aber bei den Programmierkursen fingen die Probleme zwischen Sprachwissenschaftlern und Informatikern schon an. Sind Prädikate nicht Verben, Bäume nicht Pflanzen? Ganz entnervt war schließlich ein SiC-Teilnehmer als man ihm sagte, er sollte doch mal mit der Mouse (Maus) durchs Menü (Menü) fahren. Am Ende der SiC konnte man allerdings keinen Programmierkursteilnehmer mehr mit derartigen Äußerungen aus der Fassung bringen.

(2) Thematische Kurse

Auch die thematischen Kurse umfaßten mit einigen Ausnahmen jeweils sieben Doppelstunden. Da natürlich bei der Fülle des Angebots viele Kurse parallel lagen, hatte man oft die Qual der Wahl. Folgende 20 thematischen Kurse wurden angeboten:
Altman, Hans/ Jacobs, Joachim: Syntax und Intonation

Baustle, Rainer: Prozesse und Repräsentation

Ein interessanter Kurs, in dessen Mittelpunkt die Diskursrepräsentationsanalyse stand. Problematisch daran nur die Anfertigung einer Mitschrift, da der altbewährten Tafel der Vortrag von Folien gegeben wurde und die Anordnung der wichtigen Informationen an der Tafel oft zu Wünschen übrig ließ.

Cooper, Robin: Situation Semantics und Unification

Ehlich, Konrad/ Rehbein, Jochen: Kommunikation in Institutionen

Ein recht kleiner Kurs, der leider oft nur für Insider dieser Forschungsrichtung verständlich war. Zwar brachten die theoretischen Abhandlungen zum Thema gute Einblicke in dieses Forschungsgebiet, bei den praktischen Beispielen konnten aber leider nur die Insider richtig mitarbeiten.

Engdahl, Elisabet: Current Syntactic Problems from the Point of View of GPSG and GB. A Comparative Survey

Habel, Christopher: Repräsentation und Verarbeitung von Wissen

Für mich von Aufbau und Vortrag her ein Kurs, bei dem so mancher Hochschullehrer etwas lernen könnte. Auch inhaltlich sehr in-

teressant. Es wurde auf die verschiedenen Repräsentationstypen eingegangen und Regelsysteme zur Verarbeitung von Standardannahmen vorgestellt.

Hausser, Roland: Linksassoziative Grammatik und der Parser NEWCAT

Hoepfner, Wolfgang: Sprachgenerierung in der Künstlichen Intelligenz. Woher wissen natürlich-sprachliche Systeme was sie sagen?

Kallmeyer, Werner: Konversationsanalyse

Kay, Martin: Unification Grammar

Kilbury, James: Parsing-Ansätze mit PROLOG

Krause, Jürgen: Anwendungsorientierte natürlichsprachliche Frage-Antwort-Systeme

Kuhlen, Rainer: Methoden und Systeme der Text-kondensierung

Der Kurs mit dem höchsten Aufwand an technischen Geräten und einem Manuskript, das mit dem Koffer auf der Reise von Chicago nach München verloren ging. Dennoch ein Kurs, der mich - obwohl er nur in der dritten SiC-Woche und zwischen 9.00 und 11.00 Uhr gehalten wurde, stets zum frühen Aufstehen bewegte.

Löbner, Sebastian: Quantifikation

Morik, Katharina: Natürlichsprachliche Systeme der Künstlichen Intelligenz

Ein sehr lockerer Kurs, dessen nette Atmosphäre nicht zuletzt auf die Kursleiterin zurückzuführen war. Es wurden mehrere Ansätze in Bezug auf natürlichsprachliche Systeme vorgestellt, in deren Mittelpunkt natürlich das HAM-ANS-Projekt stand.

Sells, Peter: Syntactic and Semantic Aspects of Anaphora

Tillmann, Hans Günther/ Schiefer, Lieselotte: Digitale Sprachsignalverarbeitung und experimentelle Sprachforschung

Uzskoreit, Hans: Das Grammatikmodell der Generierten Phrasenstrukturgrammatik (GPSG)

Zaefferer, Dietmar: Frage-Antwort-Linguistik

Zaenen, Annie: Introduction to Lexical-Functional Grammar (LFG)

(3) Rahmenprogramm

Für alle diejenigen, die nach achtstündigen SiC-Tag (9.00 bis 18.00 Uhr, mit einer Stunde Mittagspause) immer noch Energie hatten und weiterhin wißbegierig waren, bot sich das Rahmenprogramm an, das auch für die Öffentlichkeit zugänglich war. Es fanden jeden Abend meist noch zwei Veranstaltungen statt, in denen entweder ein Vortrag zum Thema der SiC gehalten, ein Überblick über bestimmte Forschungsaktivitäten gegeben oder ein spezielles Thema diskutiert wurde.

Vorträge

Altmann, Hans/ Jacobs, Joachim: Syntax und Intonation

Bäuerle, Rainer: Prozesse und Repräsentationen

Cooper, Robin: Situation Semantics and Unification

Ehlich, Konrad/ Rehbein, Jochen: Kommunikation in Institutionen

Habel, Christopher: Repräsentation und Verarbeitung von Wissen

Karttunen, Lauri: Lexical Unification Grammar

Kay, Martin: Some Promising Trends in Computational Linguistics (Eröffnungsvortrag)

Krause, Jürgen: Frage-Antwort-Systeme

Morik, Katharina: Natürlichsprachliche Systeme und Künstliche Intelligenz

Uzskoreit, Hans: Strategien in neueren Grammatikmodellen

Zeevat, Henk: Unification Categorical Grammar

Überblick über die deutsche und internationale Forschung

Christaller, Thomas: Verbundvorhaben im Bereich "Wissensbasierte Systeme", exemplarisch dargestellt anhand von WEREX

Hoepelmann, Jaap: ESPRIT: Beispiele aus der Projektarbeit

Lehmann, Egbert: Forschungen zur Verarbeitung natürlicher Sprache bei Siemens

Rohrer, Christian: Sprachtheorie und Maschinelle Übersetzung. Zum Stand der Forschung in Deutschland

Rollinger, Claus Rainer: Zur Entwicklung sprach-verstehender Systeme bei IBM

Sgall, Petr: Automatische Übersetzung und Frage-Antwort-Systeme in Prag

Tsujii, Jun-ichi: Activities of Computational Linguistics at Kyoto (Japan)

Panel-Veranstaltungen

Wissensrepräsentation - Sprache der Gehirne?

Computerlinguistik und Arbeitswelt

Arbeitsfeld Computerlinguistik: Perspektiven für Ausbildung und Beruf

Künstliche Intelligenz und Rüstungstechnologie: Die soziale Verantwortung des Computerlinguisten

Wissenschaftstheoretische und wissenschaftspolitische Auswirkungen des Computers auf die Linguistik

Soviel zum 'ernsthaften' Teil der Sommerschule. Ich fand das gesamte Programm sehr interessant. Dis SiC hat mir sehr viel Spaß gemacht, viele meiner Fragen sind beantwortet und diverse Denkanstöße für meine weitere Arbeit geliefert. Zu erwähnen

wäre noch die eigens eingerichtete SiC-Bibliothek, in der Literatur zum Thema der Sommerschule eingesehen werden konnte und in der sich auch Skripte zu einigen Kursen finden ließen. Eine schöne Einrichtung, die auch eifrig von den SiC-Teilnehmern genutzt wurde.

Der 'weniger ernsthafte' Teil der SiC sah von organisatorischer Seite her neben dem Eröffnungsempfang einen Biergartentreff, einen gemütlichen Abend bei Wein und Kerzenlicht (bei dem auch getanzt wurde), einen geselligen Abschlußabend und eine offizielle Schlußversammlung vor. Zur offiziellen Schlußversammlung am Freitag, den 19. September kam es allerdings nicht mehr, da man sie, weil viele Teilnehmer bereits im Laufe des Tages abreisen wollten, kurzer Hand mit dem geselligen Abschlußabend 'unifiziert' (Zitat) hatte. Der SiC-Bazillus namens 'Unifikation' hatte zu diesem Zeitpunkt schon weit um sich gegriffen.

19. September 1986

Ende der SiC. Übrig blieben zwei Ordner voller Mitschriften, Literaturlisten, kopierter Aufsätze zu dem einen oder anderen Thema, das mich besonders interessiert hatte und viele schöne Erinnerungen an Begegnungen und Gespräche mit SiC-Teilnehmern, an eine - zumindest für mich - Stadt mit Atmosphäre, die sich in diesen Spätsommertagen von ihrer schönsten Seite gezeigt hatte, so daß man doch viele Freistunden im Freien verbringen konnte. Für mich war die SiC jedenfalls eine gelungene Veranstaltung, die ich buchstäblich bis zur letzten Minute genossen habe, da unser unentwegter IISP-Dozent unbedingt noch seinen Stoff durchziehen wollte, was er bis punkt 18.00 Uhr auch mit doppelter Sprechgeschwindigkeit und einem neuen Rekord im Folienauflegen geschafft hatte. Dafür erholte er sich dann aber auch mit uns bei einem gemütlichen Abschlußtrunk in einer nahegelegenen Bierkneipe...

Marion Grabka, Universität Frankfurt

COMPUTEREINSATZ

IN DER PRAKTISCHEN LEXIKOGRAPHIE

Tutorial "MASCHINELLE LEXIKOGRAPHIE"
in Königswinter-Thomasberg, März 1987

Der viertägige, von der Gesellschaft für Linguistische Datenverarbeitung e.V. (Arbeitskreis Lexikographie), der Akademie des Deutschen Beamtensbundes e.V. sowie dem Bundesminister für Forschung und Technologie durchgeführte bzw. geförderte Kurs bot zum ersten Mal einem breiteren Interessentenkreis Gelegenheit, sich mit solchen Möglichkeiten des Einsatzes des Computers in lexikographischer Arbeit vertraut zu machen, die über den bereits weit verbreiteten Einsatz der EDV im Bereich der Drucklegung lexikographischer Werke hinausgehen. Abgesehen von Universitäten und anderen öffentlichen Einrichtungen, an welchen maschinelle sprachverarbeitende Forschung betrieben wird, rekrutierte sich der Teilnehmerkreis weiterhin aus Vertretern von Industrie und Verlagshäusern.

Die vier Veranstaltungstage waren unterschiedlichen Themenkreisen gewidmet. Am Anfang stand (A) ein Propädeutikum/Repetitorium zu allgemeinen linguistischen Grundlagen von Lexikologie und Lexikographie. (Alternativ hierzu wurde eine kurze Einführung in den Umgang mit Arbeitsplatzrechnern angeboten, welche im folgenden nicht weiter kommentiert werden soll.) Es schlossen sich an (B) die Auswertung maschinenlesbarer Textkorpora, (C) Wörterbuchstrukturen und ihre Modellierung in Datenbanken sowie schließlich (D) ein Ausblick auf die Gestaltung eines computerbasierten Arbeitsplatzes für Lexikographen bzw. für Sprachverarbeitung im weitesten Sinne.

Das in (A) eingehender besprochene linguistische Modell war das des taxonomischen Strukturalismus. Grundlegende Annahmen wie Zeichenmo-

dell, Verfahren zur Segmentierung und Klassifizierung, Prinzipien der (Teil-)Wortschatzstrukturierung (genauer gesagt deren Beschreibungskategorien), semantische Merkmale bzw. semantische Primitive im Allgemeinen, der Feldbegriff, die Dichotomie von Semasologie und Onomasologie wurden erörtert. Ansätze wie Dependenz- und Valenzgrammatik, Kasusgrammatik, Formalisierungspunzipien von Wortbedeutungen in der Künstlichen Intelligenz (semantische Netze, Frames, Scripts etc.) fanden eher am Rande Berücksichtigung, entweder als Ausblick über die strukturalistischen Verfahren hinaus oder als Unterpunkte der Wörterbuchtypologie, welche den zweiten Block des einführenden Teils bildete. Sehr illustrativ war hier der exemplarische Überblick über Artikel aus verschiedensten Wörterbüchern. Gelingen ist auch der Versuch, herauszuarbeiten, in welcher Weise linguistisch-lexikologische Prinzipien in der lexikographischen Praxis berücksichtigt bzw. ignoriert werden.

Schon das Konzept des einleitenden Tages deutet an, daß sich das Tutorial eher auf eine deskriptive maschinelle Lexikographie konzentrierte, ganz im Sinne der Definition von Lexikographie als 'wissenschaftliche Praxis mit dem Ziel der Erstellung von [gedruckten?] Wörterbüchern', welche sich schon bei der Gewinnung und Modellierung lexikographischer Materials, und nicht erst bei der Drucklegung, EDV-Verfahren bedienen will. Eine 'simulative Lexikographie', Prinzipien der maschinellen Modellierung von lexikalischem (und außerlexikalischem) Wissen in der Mensch-Maschine(Computer)-Kommunikation, waren kein explizites Thema, obwohl einige der im weiteren angesprochenen Verfahren u.U. auch für diesen Bereich eine gewisse Relevanz haben.

Der eher praxisorientierte Abschnitt (B) demonstrierte eindrucksvoll, welch mächtiges Werkzeug dem Lexikographen mit einem Dialog-Konkor-

danz-Programm wie BYU-Concordance selbst auf Kleinrechnern zur Verfügung steht. Ein wesentlicher Punkt ist hier natürlich das eigentliche Konkordanzcharakteristikum des direkten, selektiven Zugriffs auf ein großes Textkorpus, wobei jedoch im Vergleich zu gedruckten Konkordanzen die Zugriffsmöglichkeiten wesentlich gezielter sind, da die Selektionsbedingungen auch bezüglich des Kontextes mannigfach spezifizierbar sind. Das zweite Moment ist die im Programm integrierte Möglichkeit, gefundene Textpassagen mit Relevanz für die Wörterbucharbeit direkt für weitere Verarbeitungsschritte in besondere Dateien zu transferieren, von wo aus sie in Texte oder Datenbanken eingegliedert sind, was exemplarisch vorgeführt wurde. Exzerpierungen von Hand und Zettelkarteien überbrücken sich natürlich bei diesem Vorgehen, sofern, und das darf nicht übersehen werden, die Grundvoraussetzungen der Maschinenlesbarkeit der Texte erfüllt ist. Verfahren zur Gewinnung derartiger Korpora, Probleme der Lemmatisierung und Homographendambiguierung wurden im weiteren erörtert. Auch wurde eine Skizze des Einsatzes dieser Verfahren im Projekt EUROTRA (maschinelle Übersetzung in Institutionen der Europäischen Gemeinschaft) geboten, soweit er am Institut für Kommunikationsforschung und Phonetik der Universität Bonn erfolgt.

Am Anfang von (C) stand ein überaus komprimierter Einstieg in die Theorie des sog. relationalen Datenmodells. Aufgrund seiner 'Benutzerfreundlichkeit' hat sich dieses im Datenbankbereich weitgehend durchgesetzt und findet, obwohl zunächst für wirtschaftliche Anwendungen konzipiert, z.Zt. stets weitere Bedeutung für sprachverarbeitende Zwecke. Als Brückenschlag zur Lexikographie wurde die Umsetzung von in (A) diskutierten Wörterbuchstrukturen in Relationen theoretisch und exemplarisch auf dem Hintergrund der Arbeiten am COLEX-Projekt des Germanistischen Seminars der Universität Heidelberg skizziert. Abendvorträge zu diesen drei Themenkomplexen von ausgewiesenen Wissenschaftlern rundeten das eigentliche Tutorial ab und boten weitere Möglichkeiten zur Diskussion von Theorie und Praxis. Abschließend wurden in (D) zum einen die (z. T. zukünftigen) Möglichkeiten eines Computerarbeitsplatzes für Sprachverarbeitung im weitesten Sinne (Texterschließung, wissenschaftliche Textverarbeitung nicht nur für Lexikographie, Enzyklopädie etc.) aufgezeigt. Immer wieder ist in der Fachliteratur der 'handwerkliche' Aspekt der Lexikographie unterstrichen worden - was ihr nicht die wissenschaftlich-theoretischen Grundlagen entziehen sollte. Es sollte gerade für den Geisteswissenschaftler ein Desiderat sein, Handwerkliches die Maschine übernehmen zu lassen. Hier muß der Rechner verschiedenste Funktionen in sich integrieren, vom Textarchiv über Zettelkasten, bibliographische und Protokollkartei, Notiz- und Exzerptsammlung, (statistischer) Rechner zum Ergebnisarchiv und textestellender Komponente und zwar - dieses war zum

anderen Thema von (D) - auf eine menschenfreundliche Art und Weise.

Die Veranstaltung - das war einhelliges Fazit von Teilnehmerseite - war durch ihr überaus breit angelegtes Programm von der strukturalistischen Ausdruck-Inhalt-Dichotomie bis zur Logik relationaler Datenbanken und ihren Abfragesprachen und den Ambitionen fast aller Dozenten, jede relevante Unterproblematik ihrer Themen anzusprechen, überaus kondens. Dieses hatte den Vorteil, daß bei durchaus heterogenem Hintergrund der Teilnehmer ein jeder mit diversen Anregungen oder auch Bestätigungen nach Hause fahren konnte. Das Programm war durch ein stattliches Lesebuch vorbildlich dokumentiert, und zu Recht (jedoch mit zu geringer Frist) wurde den Teilnehmern abverlangt, sich mit dessen Inhalt vor Beginn vertraut zu machen. Das Problem terminologischer Diskrepanzen zwischen Informatik einerseits und Linguistik andererseits zeigte sich mitunter deutlich. Zu übersehen ist auch nicht, daß bei einer derartigen Komprimierung von Inhalten auch die Gefahr eines 'Overkills' gegeben ist. Um keinen Zweifel zu lassen - auch das war einhelliges Teilnehmerfazit - das Tutorial, von Gert Frackenpohl exzellent organisiert, ist gelungen, seine Institutionalisierung (bei Ausweitung auf einen studentischen Teilnehmerkreis) wünschenswert. Es bot eine Möglichkeit des Dialogs innerhalb und zwischen forschender und anwendender Seite. Mit der zunehmenden Bedeutung sprach- und wissensverarbeitender Prozesse in z. B. Industrie und Verwaltung wird auch ein 'Technologietransfer' oder, bescheidener ausgedrückt, ein (nicht unbedingt einseitiger) Gedankenaustausch zwischen theoretischer und angewandter sprachwissenschaftlicher Forschung zum einen und Anwendern zum anderen unabdingbar im beiderseitigen Interesse. Ein erneutes Tutorial zur maschinellen Lexikographie könnte neben deskriptiven Aspekten auch der Modellierung von sprachlichem und außersprachlichem Wissen in Lexika der Künstlichen-Intelligenz-Systeme sowie den Wortschatzkomponenten in der maschinellen Übersetzung weitere Beachtung schenken. Dazu müßten den Veranstalter jedoch in verschiedener Hinsicht die Möglichkeiten zur Differenzierung des Konzepts gegeben werden.

Gerd Hentschel, Seminar für Slavische Philologie der Universität Göttingen, Nikolausberger Weg 9c, 3400 Göttingen

ANWENDUNG VON REPRÄSENTATIONSTECHNIKEN

Workshop an der TH Darmstadt, Februar 1987

Gut 40 Teilnehmer versammelten sich am 12. und 13. Februar 1987 in Darmstadt, um über die Anwendung von Wissensrepräsentationstechniken zu diskutieren. Die Veranstalter, die Arbeitsgruppe "KI und Informationswissenschaft" in Zusammenarbeit mit der Fachgruppe Wissensrepräsentation im FA 1.2 der GI, waren sich darüber einig, daß eine Methodenlehre der Wissensrepräsentation ein Desiderat sei. Unterschiedlich waren die Prognosen, wie weit man mit einem ersten Workshop auf dem Wege dahin kommen könne. Recht behielten die gemäßigten Skeptiker: Es wurde ein Anfang gemacht, der - wie Christopher Habel in seiner Zusammenfassung konstatierte - auch beispielhafte Antworten brachte, aber insgesamt fruchtbarer in Fragen als in Antworten war.

Gefragt wurde nach Grundvorstellungen von Repräsentationsmethoden und -techniken. Anklang fand eine Vier-Ebenen-Sicht der Wissensrepräsentation, von Bernd Neumann entwickelt und von Bernhard Nebel ergänzt:

- Konzepte der Modellierung
- Konzepte der Formalisierung
- Repräsentationsformalismen
- Realisierung

Christopher Habel konnte mit einem Zitat aus der Diskussion des Workshops seinen Verdacht belegen, daß in der normalen Wissensrepräsentation die Implementierung weit stärker als die Analyse des Domains den Gang der Repräsentation bestimmt: "Ich habe etwas zu repräsentieren, das ist einigermaßen bekannt. Wie bekomme ich es auf ein Tool?" Eine Reduktion der Repräsentationsmethodik auf die

bloße Verwendung von Repräsentationswerkzeugen ist unbefriedigend. Sie erklärt sich durch den Mangel an allgemeineren Prinzipien und Methoden der Wissensrepräsentation.

Immerhin gab es Ansätze, den Wunsch nach einer stärker durch den dargestellten Sachverhalt geleiteten Repräsentationsarbeit zu erfüllen:

Gleich im ersten Beitrag hatte Brigitte Endres-Niggemeyer einen echten Anwendungsfall vorgestellt: Ein Vergleich mehrerer Repräsentationen eines Inhaltsverzeichnisentrags aus einer Fachzeitschrift führte sie zu der Forderung nach problemnäheren Repräsentationssprachen, im konkreten Fall für Informationsobjekte.

Bernd Neumann beschäftigte sich mit der Repräsentation des Gegenständlichen - nach der beherzigenswerten Maxime, daß der Umgang mit Abstrakta leichter fällt, wenn man zunächst mit Objekten der Anschauung zurechtkommt. Er führte das schon erwähnte Stufenmodell der Wissensrepräsentation ein, wies auf die Notwendigkeit hin, zwischen Repräsentationskonzepten, Repräsentationsmethoden und Repräsentationstechniken zu unterscheiden und demonstrierte höchst anschaulich verschiedene Techniken der Objektrepräsentation aus der Computer Vision.

Lothar Rostek berichtete ausführlich über SMALL-TALK und leider etwas knapp über seine Realisierung einer Thesaurusrepräsentation mit SMALLTALK.

Am nächsten Morgen erinnerte Dietmar Roegner an das Thema des Workshops, bevor er detailliert und anschaulich über die Generierung deutscher Texte aus einer frameartigen Repräsentation berichtete. Eine ebenfalls detailgenaue Darstellung der Repräsentationswerkzeuge des Konstanzer Frame-Repräsentationsmodells verdankten die Workshopteilnehmer Ulrich Reimer.



Abb.: Eine Entscheidungssituation im Vorstand der Supersoftware AG. Illustration eines Szenarios als Grundlage der Diskussion auf dem Workshop *Anwendung von Repräsentationstechniken*

Bernhard Nebel klärte in dankenswerter Weise seine Auffassung von Wissensrepräsentation im Dienste einer Theorie menschlicher Schlußfolgerung, bevor er zu seinem eigentlichen Thema kam, der Anwendung von KL-TWO in der Textgenerierung. Er sparte nicht mit Hinweisen auf andere Repräsentationsformalismen, wies auf die Wichtigkeit des Anwendungsbezuges für die Weiterentwicklung der Repräsentationsmethoden hin. Am Beispiel eines Konzeptes für "Commercial Exchange" (Kaufen) machte er deutlich, wie schwierig eine angemessene Repräsentation einfach erscheinender Konzepte sein kann.

Für Diskussionsstoff am Abend sorgte ein gemeinsames Repräsentationsproblem: Der Vorstand der Supersoftware AG muß entscheiden, ob man sich auf dem japanischen Markt engagieren will. Brigitte Endres-Niggemeyer skizzierte eine für die AUTOCAT-Repräsentation typische Auffassung des Szenarios als komplexes Informationsobjekt. Das Szenario wurde auf seine vielfältigen Ansatzpunkte für andere Repräsentationsstrategien und -methoden hin diskutiert.

In der Schlußdiskussion wurde anvisiert, eine eventuell verbesserte Fassung des SuSo-Szenarios zum integrierenden Mittelpunkt eines nächsten Workshops zu machen, der im Herbst 1987 stattfinden kann.

Brigitte Endres-Niggemeyer, Fachhochschule Hannover, Fachbereich BID.

ANWENDER-WORKSHOP DES AUTOCAT-Projekts

Projektmitarbeiter stellen ihren ersten Prototypen zur wissensbasierten Formalkatalogisierung vor.

"Wissensbasierte Systeme, die einen wesentlichen Beitrag zur Effektivitätssteigerung im Bereich des Input in Literatur-Datenbanken und Bibliothekssysteme leisten können, kommen nicht etwa in 10 bis 20 Jahren auf den Bibliotheks- und LuD-Bereich zu - sie stehen vor der Tür" - so könnte man die zentrale Aussage der AUTOCAT-Projektgruppe paraphrasieren, die am 11. Mai 1987 zu einer Veranstaltung an die TH Darmstadt geladen hatte, um das bisher erreichte Projektergebnis vorzustellen und Aspekte der Anwendung zu diskutieren. Das Projekt AUTOCAT wird seit 1. Oktober 1985 an der TH Darmstadt durchgeführt und vom BMFT gefördert. Unterstützt wird das Projekt durch die FH Hannover, das Fachinformationszentrum Karlsruhe (FIZ 4), den Lehrstuhl Informatik IV der Universität Saarbrücken und die FH Darmstadt. Ziel ist die Entwicklung eines wissensbasierten Systems zur automatischen bibliographischen Erschließung von Literatur am Beispiel von Aufsätzen in Physik-Fachzeitschriften nach dem (internationalen) Standard des INIS-Regelwerkes. AUTOCAT will damit an der komfortableren Gestaltung

der Schnittstelle zwischen den Originalpublikationen und den davon abgeleiteten Informationserzeugnissen (hier also zwischen Verlagswesen und Fachinformationszentrum) mitarbeiten. Damit wird ein Einstieg in die Reorganisation des Literaturdatenbank-Inputs mit den Techniken der wissensbasierten Informationsverarbeitung erreicht.

Über 50 Teilnehmer aus den verschiedensten Bereichen waren der Einladung zu dem AUTOCAT-Anwenderworkshop gefolgt. Anwesend waren insbesondere Vertreter der Fachinformationszentren, der Bibliotheken, der bibliographischen Berufsverbände, der Verlage und Druckereien, der Hochschulen und Fortbildungsinstitute sowie der "Technik" (Softwarehersteller, Anbieter von optischen Lesern).

In ihrem Einführungsreferat stellte und beantwortete Brigitte Endres-Niggemeyer eine Reihe grundlegender Fragen: *Was ist und will AUTOCAT?*, *Was sind wissensbasierte Systeme bzw. Expertensysteme?*, *Was ist Formalerfassung?*. Anschließend stellte sie im Überblick die Arbeiten des AUTOCAT-Projekts vor und erläuterte die AUTOCAT-Konzeption.

Sehr reges Interesse fand erwartungsgemäß auch der zweite Beitrag der Veranstaltung, in dem Christina Schnellbach die Inputbeschaffung für das AUTOCAT-System erläuterte. Zunächst ging sie auf die Situation im Verlags- und Druckereibereich hinsichtlich einer computergestützten Satzerstellung ein. Die Entwicklung ist hier noch sehr im Fluß und Standards bilden sich gerade erst heraus. Die Alternative, nämlich die Erfassung durch einen geeigneten optischen Leser, entwickelte sich aufgrund der gegebenen Situation zur Standard-Input-Schnittstelle.

Die Grundlage der Modellierung von Wissen über die Informationsstruktur von Physikzeitschriften war die detaillierte Auswertung derjenigen Zeitschriften, mit denen das System zunächst arbeiten soll (Physikzeitschriften, die vollständig für die Physik-Datenbank PB ausgewertet werden; die sogenannten "Kernzeitschriften" des Faches Physik). Amelie von Below stellte insbesondere das Kategorienschema vor, auf dessen Grundlage die mehr als 300 Kernzeitschriften hinsichtlich Struktur und Layout beschrieben, in einer Datenbank gespeichert und ausgewertet wurden.

Ergebnis der Zeitschriftenauswertung war insbesondere die detaillierte Beschreibung der "Physikzeitschrift im allgemeinen" und der einzelnen speziellen Physikzeitschriften in ihrer hierarchischen Struktur (im wesentlichen der *Zusammensetzungsstruktur*). In der Repräsentation dieser Beschreibung sind den einzelnen Zeitschriften-Komponenten (z.B. Beitragskopf, Titel, Autor) sogenannte *Experten* zugeordnet, die bei der Erkennung für die Beschaffung der Information aus den Rohdaten verantwortlich sind. An mehreren Beispielen zeigte Ulrike Rauth auf, wie spezielle Grammatiken die Grundlage solcher *Experten* bilden. Das Konzept der Typisierung erlaubt

darüber hinaus, das Vorwissen über spezielle Zeitschriften für eine erwartungsgesteuerte und effiziente Analyse zu nutzen, ohne das System bereits vom Ansatz her inflexibel gegenüber auftretenden Varianten und Änderungen zu machen.

Die Architektur des AUTOCAT-Systems war das Thema des folgenden Vortrags von Peter Müller. Er erläuterte an einem besonders anschaulich aufbereiteten Beispiel die Arbeitsweise und Organisation des Systems und stellte die einzelnen AUTOCAT-Komponenten (Monitor, Interpreter, Wissensverwaltung, Input- und Output-Verwaltung) vor.

Nach einem kleinen Ortswechsel wurde das Thema anhand einer Vorführung des in MPROLOG realisierten Prototyps weiter vertieft. Peter Müller, Harald Pitz, Petra Götz, Renato Vinga-Martins, Roland Kuhn und Iris Bauer, die die Systemkonzeption ausgearbeitet und implementiert hatten, führten die Arbeitsweise des Systems an mehreren Zeitschriften vor. Die Diskussion mit den Teilnehmern ergab verschiedene Anregungen, die das Projekt aufgreifen wird.

Parallel wurde von Gerhard Knorz anhand eines (gegenüber AUTOCAT) stark vereinfachten, kleinen Demonstrationssystems die Analyse (und Weiterverarbeitung) der formalen Struktur von Geschäftsbriefen vorgeführt, wodurch ein möglicher weiterer Anwendungsfall von AUTOCAT illustriert wurde.

Als Einstieg in die Abschlussdiskussion versuchte Gerhard Knorz, von der gegenwärtig vorliegenden Prototyp-Version auf ein einsetzbares Softwareprodukt zu extrapolieren. Er stellte die Skizze eines (dezentralen) Arbeitsplatzsystems vor, das die Auf-

- Aufbau der Wissensbasis
- Überwachung des Routinebetriebs einer automatischen formalen Erfassung
- Interaktion mit dem System bei "Problemfällen" und damit verbunden
- Anpassung und Erweiterung der Wissensbasis

mittels einer komfortablen graphisch-interaktiven Benutzerschnittstelle zu einem transparenten und flexiblen Werkzeug des Benutzers macht. Gerade vor dem Hintergrund von Vorbereitungen, ein AUTOCAT-II-Projekt zu definieren, dessen Ziel das einsetzbare AUTOCAT-Produkt ist, wurde nun nach dem Beitrag gefragt, den ein solches System zur Lösung des Input-Problems im IuD-Bereich aus der Sicht des Anwenders leisten kann, welche Probleme für und bei der Anwendung eines solchen Systems gesehen werden.

Die Diskussion brachte eine Vielzahl von Anregungen. Eine schriftliche Zusammenfassung der Diskussion wurde inzwischen zur Ergänzung und Stellungnahme an die Teilnehmer verschickt. Unmittelbare Vorteile einer AUTOCAT-Anwendung sahen die Teilnehmer insbesondere darin, daß

- das Niveau der Arbeit bei der formalen Erschließung von der Ebene des in einfachen Fällen sich ständig wiederholenden Einzelfalls auf die abstraktere Ebene der Formulierung von Regeln angehoben werden kann,
- der Routinefall bei der formalen Erfassung einen erheblich geringeren Arbeitseinsatz erforderlich machen wird,
- insbesondere bei der Nutzung von Satzbändern der Input um Größenordnungen beschleunigt werden kann (im Fachinformationsbereich ein Argument mit hoher Priorität),
- insbesondere bei der Integration einer computergestützten inhaltlichen Erschließung organisatorische Vereinfachungen bringt,
- viele aufwendige formale Prüfroutinen beim und vor dem Datenbankinput entfallen können,
- das AUTOCAT-System als ein flexibles und vielseitiges Tool für eine Vielzahl weiterer Aufgaben, z.B. für die Auszeichnung von Manuskripten in Verlagen eingesetzt werden kann.

Daß der Einsatz von Verfahren der wissensbasierten Inputbearbeitung über technisch/organisatorische Fragen weit hinausgeht, wurde ebenfalls angesprochen (Fragen des Berufsbildes, der betrieblichen Interessensvertretung, der Ausbildung)

Soweit ein Veranstalter geeignet ist, dies festzustellen, wurde die Veranstaltung einhellig als sehr informativ und für weitere Überlegungen als anregend aufgenommen. Es ist geplant, zum Projektende eine weitere Veranstaltung vorzubereiten, die sich dann mehr den wissenschaftlichen und (entwicklungs-) technischen Aspekten widmen soll.

Gerhard Knorz, FH Darmstadt, Fachbereich IuD

"KÜNSTLICHE INTELLIGENZ" - SPRACHVERARBEITUNG -

Deutsche und französische Wissenschaftler trugen in Stuttgart am 11. und 12. Dezember 1986 über ihre Forschungsaktivitäten -bedingungen und -pläne vor

Die Tagung "Künstliche Intelligenz - Sprachverarbeitung", im Dezember in Stuttgart durchgeführt, wurde bereits im Frühjahr 1986 von Walther v. Hahn in Paris geplant. Gespräche mit einer größeren Zahl von französischen Kollegen hatten gezeigt, daß man über die konkreten Forschungsbedingungen und -pläne der Gruppen des jeweils anderen Landes viel zu

wenig wußte. Da Konferenzprogramme in der Regel diese Art von Information nicht liefern, entstand die Idee eines eher "technischen" Treffens ohne umfangreiche Vorträge sondern nur mit Berichten über die Forschungsaktivitäten der größeren Gruppen in der Bundesrepublik und in Frankreich.

Mit *Gerard Sabah* als französischem Mitorganisator und mit tatkräftiger Unterstützung von *Clemens Heller* (Maison des sciences de l'homme) sowie *Hinnerk Bruhns* (C.N.R.S.) wurde die Tagung vorbereitet. Die C.N.R.S. und die Maison übernahmen im Rahmen des Programme Franco-Allemand die Reise- und Aufenthaltskosten für die französischen Teilnehmer, die DFG dieselben für die deutschen Teilnehmer.

Zur Tagung im Novotel Stuttgart-Nord trafen dann die folgenden eingeladenen Kolleginnen und Kollegen ein:

Noëlle Carbonell (CRIN Nancy, für den eingeladenen J.-M. Piel), *Daniel Coulon* (CRIN Nancy), *Laurence Danlos* (LADL, Paris VII), *Monique Flasaquier* (Französische Botschaft Bonn), *Günther Göz* (Erlangen), *Maurice Gross* (LADL, Paris VII), *Y.Ph. Guilbaud* (Genf, für den eingeladenen Chr. Boitet), *Christopher Habel* (Hamburg), *Walther v.Hahn* (Hamburg), *Wolfgang Hoepfner* (Koblenz), *Daniel Kayser* (LIPN - CSP Orsay), *Gerhard Knorz* (Darmstadt), *Irin Kowarski* (IMAG Grenoble, für den eingeladenen J. Courtin), *Heinrich Niemann* (Erlangen), *Patrice Pognan* (CERTAL INALCO Paris), *Yolla Polity* (CRISS Grenoble, für den eingeladenen J. Rouault), *Christian Rohrer* (Stuttgart), *Gerard Sabah* (LIMS1, Orsay), *Wolfgang Wahlster* (Saarbrücken), *Henti Zingle* (Chambéry), *Michael Zock* (LIMS1, Orsay)

Eine Reihe von eingeladenen Kollegen hatten leider aus zwingenden Gründen absagen müssen.

Am Donnerstagmorgen begannen nach einer kurzen Begrüßung und allgemeinen Einleitung die Vorträge in eher zufälliger Reihenfolge, je ein deutscher und französischer Teilnehmer abwechselnd. Die Konferenzsprache war englisch. In der Regel folgte man bei der Darstellung dem Vorschlag der Einladung:

- Mitglieder der Gruppe
- Förderungsstruktur
- Forschungsschwerpunkt
- Zukünftige Vorhaben
- grundlegende Publikationen
- Apparative Ausstattung
- Bestehende Zusammenarbeit
- Zusammenarbeitsbedarf
- Adresse, Telefon, Netzadresse

Die Beiträge wurden kurz diskutiert und bei Nachfragen ergänzt. Insgesamt wurde deutlich, daß die deutschen Gruppen z.T. erheblich besser ausgestattet waren als ihre französischen Kollegen und daß eine Reihe französischer Forschungsgruppen aus "deutscher Sicht" eher einer computerorientierten

Linguistik als der sprachorientierten KI zuzuordnen waren.

Es stellten sich bei den Vorträgen eine Reihe bisher nicht offen zutage getretener Gemeinsamkeiten und Anknüpfungspunkte heraus, die in persönlichen Gesprächen vertieft werden konnten. Es ergab sich durch den gewählten Rahmen nach einhelliger Ansicht ein besonders enger Kontakt zwischen den Teilnehmern, der intensiv zu Fachgesprächen und zweiseitigen Vereinbarungen genutzt wurde.

Am Abend des ersten Tages folgten die meisten Teilnehmer einer Einladung zu einem Empfang der SEL-Stiftung für technische und wirtschaftliche Kommunikationsforschung in das Forschungslabor der SEL AG.

Am Freitagmorgen waren die Teilnehmer zu einem Empfang des Wirtschaftsministers des Landes Baden-Württemberg, Herrn Dr. Herzog eingeladen, der insbesondere die Wichtigkeit der Forschung im Bereich "Künstliche Intelligenz" und Sprachverarbeitung herausstellte.

Nach den restlichen Vorträgen im Laufe des Freitags schlossen sich vier Diskussionspunkte an:

- (1) Darstellung der typischen Forschungsorganisation und der üblichen Semesterdaten in beiden Ländern (*Gross, v.Hahn*)
- (2) Bedingungen und Möglichkeiten der Projektbeantragung und -durchführung in beiden Ländern (*Gross, v.Hahn*)
- (3) Förderungsmöglichkeiten (*Flasaquier*)
- (4) Weiterführung der Aktivitäten, weitere Treffen? (Plenum)

Es wurde vereinbart, sich über Projekte zukünftig gegenseitig besser zu unterrichten und die Austauschangebote zu nutzen. Es wurde angeregt, auch auf der Ebene der Lehre Austausche oder Besuche vorzusehen und dabei die Mittel der Europäischen Gemeinschaft in Anspruch zu nehmen. So wird *Gerard Sabah* im April 1987 zu Vorträgen in der Bundesrepublik sein.

Die Teilnehmer beauftragten *W. v.Hahn*, nach etwa eineinhalb Jahren eine erneute Umfrage unter den Teilnehmern und weiteren Interessenten zu versenden und bei überwiegender Bedarf zu einem Folgetreffen in Frankreich einzuladen.

Die Möglichkeit, das LDV-Forum auf deutscher Seite als Plattform für die Darstellung von Forschungsaktivitäten auf französischer Seite zu nutzen, wurde begrüßt.

Die Veranstaltung wurde von allen Teilnehmern einhellig als sehr effizient und erfolgreich eingeschätzt. Es war Gelegenheit gegeben, die ausgelagerten Publikationen aus den Forschungsgruppen der Kollegen zu studieren und über Bestelllisten Kopien anzufordern. Die Unterlagen des Treffens (Folien, Texte, Literaturlisten, Infos) wurden im Mai 1987 an die Teilnehmer verschickt.

Walther v. Hahn, Universität Hamburg; Gerhard Knorz, FH Darmstadt

CHANCEN FÜR PHILOLOGEN?

Neuer Aufbaustudiengang "Linguistische Informations- und Textverarbeitung" an der Universität Würzburg

Der Präsident der Julius-Maximilians-Universität, Professor Theodor Berchem, überreichte Ende 1986 den ersten acht Absolventen des neuen Aufbaustudiengangs "Linguistische Informations- und Textverarbeitung" ihre Zeugnisse. Mit diesem Aufbaustudiengang, der im Wintersemester 1985/86 zunächst für eine zweijährige Erprobungsphase an der Philosophischen Fakultät II eingerichtet wurde, demonstriert die Universität Würzburg ihren Willen, Absolventen auch der geisteswissenschaftlichen Fächer zusätzliche Qualifikationsmöglichkeiten anzubieten, die sie vom Arbeitsmarktfeld Schule etwas unabhängiger machen können. Das Ausbildungsprogramm ist auf zwei Semester konzentriert. Es umfaßt vor allem die Arbeit mit verschiedenen Rechner- und Betriebssystemen, mehrere Programmiersprachen, den Umgang mit kommerziellen Text- und Dokumentationssystemen und die Fähigkeit zu fester Bewertung, fachliche Dokumentation, die Vermittlung EDV-spezifischer Englischkenntnisse, betriebswirtschaftlicher Grundbegriffe und nicht zuletzt ein Betriebspraktikum. Derzeit beginnt das Aufbaustudium jeweils im Wintersemester. Voraussetzung für die Einschreibung ist ein abgeschlossenes Hochschulstudium mit philologischem Schwerpunkt. Nach Abschluß der Erprobungsphase im SS 1987 soll der Studienbeginn in jedem Semester möglich sein.

Main Post vom 20./21. Dezember 1986

NICHTNUMERISCHE DATENVERARBEITUNG IN DEN GEISTESWISSENSCHAFTEN

Weiterbildungsveranstaltung in Göttingen, 1987

Die Gesellschaft für Wissenschaftliche Datenverarbeitung Göttingen (GMDG) und das Max-Planck-Institut für Geschichte (MPIG) veranstalten vom 27. Juli bis zum 8. August erstmals eine gemeinsame Weiterbildungsveranstaltung für Geisteswissenschaftler. Diese Veranstaltung soll eine praxisorientierte Ausbildung in vier Bereichen datentechnischer Arbeit gewährleisten, die für die Geisteswissenschaften allgemein, bzw. die historischen Disziplinen im Besonderen, derzeit von immer größerer Bedeutung werden.

Folgende vier Kursbereiche werden angeboten:

- Der Einsatz des Satzsystems \TeX für anspruchsvolle Druckvorhaben in den Geisteswissenschaften. (GWDG)

- Der Einsatz von Multi-font-Lesegeräten (Kurzweil Data Entry Machine = KDEM) für die Umsetzung gedruckter Texte in maschinenlesbare Datensätze (GWDG)
- Der Einsatz datenbankorientierter Software zur Unterstützung quellenorientierter Verfahren in den historischen Wissenschaften am Beispiel des Softwaresystems Cho/C (MPIG)
- Der Einsatz von Full Text-Retrieval Systemen am Beispiel der Softwaresysteme BYU-Concordance, Cho/C und Fatras (MPIG)

Teilnehmerkreis: Unabhängig von den einzelnen Kursbeschreibungen, wendet sich die Gesamtveranstaltung an Angehörige des Mittelbaus und Studenten, die für die Planung und Durchführung entsprechender Forschungsvorhaben verantwortlich sind oder in den kommenden Semestern die Verwendung der behandelten Techniken in eigenen Arbeiten planen.

Verlauf: Der Kursstoff wird grundsätzlich auf drei Ebenen vermittelt: in *Plenarveranstaltungen*, in denen das für den jeweiligen Bereich notwendige Hintergrundwissen vermittelt wird, in *Übungen*, bei denen an Hand vorbereiteter Daten und Anwendungsbeispiele die praktische Arbeit gelehrt wird und in *Projektberatungen*, bei denen die Teilnehmer in die Lage versetzt werden, weitere Übungen an Hand eigener Daten durchzuführen.

Teilnahmebedingungen: Die Teilnahme ist grundsätzlich kostenlos, für vor Kursbeginn versandte Materialien wird jedoch eine Gebühr von DM 30,- eingehoben, die nach Zusage der Teilnahme durch die Veranstalter fällig wird.

Da Übungen und Projektberatungen in Kleingruppen von in der Regel nicht mehr als fünf Personen erfolgen sollen, ist die Teilnehmerzahl strikt beschränkt. Erfolgen mehr Anmeldungen, als durch die Beschränkungen bei den einzelnen Kursen berücksichtigt werden können, wird über die Aufnahme auf Grund der bei der Anmeldung angegebenen Vorkenntnisse entschieden. Ein ausführlicher Kursprospekt wird auf Anforderung zugesandt.

Interessenten wenden sich bitte an eine der beiden folgenden Adressen:

Dr. Peter Scherber, Gesellschaft für Wissenschaftliche Datenverarbeitung, Am Fallberg, 3400 Göttingen

Dr. Manfred Thaller, Max-Planck-Institut für Geschichte, Hermann-Fogge-Weg 11, 3400 Göttingen

SCHWERPUNKT "MASCHINELLE SPRACHVERARBEITUNG" AN DEN UNIVERSITÄTEN STUTTART UND TÜBINGEN

An den Universitäten Stuttgart und Tübingen entstand ein neuer Forschungsschwerpunkt "Maschinelle Sprachverarbeitung", der vom baden-württembergischen Minister für Wissenschaft und Kunst am 8. Dezember 1986 in der Universität Stuttgart der Öffentlichkeit vorgestellt wurde.

Derzeit laufen bereits eine Reihe von Projekten an den beiden Universitäten, die von den Linguisten Prof. Dr. Christian Rohrer, Stuttgart, und Prof. Dr. Franz Günthner, Tübingen, mit insgesamt 30 Mitarbeitern betreut werden. Im Zuge des neuen Schwerpunktes sollen weitere Professoren und Mitarbeiter für die Bereiche Computerlinguistik, formale Logik und Sprachphilosophie und für experimentelle Phonetik (Akustik) hinzukommen. Zudem soll die bereits bestehende Kooperation mit z.B. Elektrotechnikern, Anglisten, Linguisten, Informatikern, mit der Industrie und anderen Forschungseinrichtungen, beispielsweise der Fraunhofer Gesellschaft, weiter vertieft und ausgebaut werden. Neben bereits existierenden Zentren für die Sprachverarbeitungsforschung in Saarbrücken und Hamburg und dem vom Bundesforschungsministerium geplanten Forschungszentrum für künstliche Intelligenz (Bericht in diesem Heft) soll die Stuttgarter und Tübinger Aktivität das vierte Zentrum dieser Art in der Bundesrepublik Deutschland werden.

Aktivitäten der Universität Stuttgart

Jährlich werden derzeit 45 Milliarden Mark weltweit für Übersetzungen ausgegeben. Die jährliche Zuwachsrate dieses Marktes liegt bei 15 Prozent. Der Sprachendienst der EG-Kommission umfaßt allein 1.100 Personen und rund 50 Prozent der Beamten des EG-Parlaments sind mit Übersetzungen beschäftigt. Durch die Erweiterung der Europäischen Gemeinschaft gibt es derzeit 72 Sprachpaare. (Quelle: Wochenzeitung "Die Zeit" vom 31.10.1986). Der Bedarf an genauen und schnellen Übersetzungen ist evident. Erste theoretische Voraussetzungen zur maschinellen Sprachverarbeitung wurden bereits Mitte der 50er Jahre mit Chomsky's transformationeller Grammatik geschaffen. Dies war eine rein geisteswissenschaftliche Fragestellung, die sich unabhängig von der Rechnerentwicklung vollzog. "Als Linguist hat mich seit den 60er Jahren die Frage sehr interessiert, inwiefern man natürliche Sprachen formal (und damit auch für den Rechner geeignet) beschreiben kann", erklärte Prof. Dr. Christian Rohrer. Er betont allerdings, daß die maschinelle Sprachverarbeitung heute immer noch in den Kinderschuhen stecke. "Realistisch kann man heute erst von maschinenunterstützten Übersetzungen, von maschinenlesbaren Wörterbüchern reden. Auf diesem Gebiet muß noch sehr viel Grundlagenforschung betrieben werden".

Rohrer sieht auf dem Gebiet der maschinellen Sprachverarbeitung auch gute Berufsaussichten für entsprechend qualifizierte Geisteswissenschaftler. Bereits jetzt suchen große Computerfirmen qualifizierte Leute für die sich im Aufbau befindlichen Forschungsruppen.

Derzeit arbeiten die Stuttgarter Informatiker und Linguisten unter anderem gemeinsam an einem japanisch-deutschen Übersetzungsprogramm Sem-syn, ein seit 1983 laufendes Projekt, das vom Bundesforschungsministerium gefördert wird und in Kooperation mit Japan läuft. Ziel ist dabei die automatische Übersetzung von japanischen Zeitschriften ins Deutsche aus dem speziellen Gebiet der Informationstechniken. Zudem arbeiten die Stuttgarter Linguisten in Verbindung mit dem EG-Programm Eurotra an Problemen der maschinellen Übersetzung vom Deutschen ins Französische.

Auf dem Markt werden bereits automatische Korrekturprogramme angeboten. Im Rechner ist ein Wörterbuch gespeichert. Wenn er ein Wort nicht kennt, gibt er entweder ein akustisches Zeichen oder korrigiert automatisch. Im Stadium der Grundlagenforschung befinden sich grammatikalische Analysen, mit deren Hilfe der Rechner auch Grammatik- und Syntaxfehler automatisch korrigieren kann.

Sinnvoll für Behinderte, für Auskunfts-systeme oder für computerunterstützten Unterricht ist auch die Entwicklung von Programmen, durch die der Rechner in die Lage versetzt wird, eingetippte Sätze auszusprechen. Dabei wird nicht etwa ein Tonband in Gang gesetzt, sondern jedes Wort, jeder Satz vom Rechner neu erzeugt.

Ein wichtiges Projekt ist auch die Entwicklung von Zugriffsmöglichkeiten auf Datenbanken in natürlicher Sprache. Hier arbeiten die Stuttgarter Wissenschaftler im Auftrag der Europäischen Gemeinschaft am Esprit-Programm daran, Beschreibungsverfahren zu entwickeln, die es auch ungeübten Benutzern erlauben, eine Datenbank in der jeweiligen Muttersprache abzufragen. Dies ist von großer Bedeutung zum Beispiel für internationale Institutionen, aber auch für international arbeitende Firmen.

Rohrer warnt vor allzu optimistischen Erfolgsvorsprechungen. "Es sind außerordentlich komplexe Forschungen notwendig, um syntaktische und semantische Strukturen der verschiedensten Sprachen zu beschreiben und das jeweils notwendige Fachwissen zu repräsentieren". Doch er betont, daß vor allem die USA und Japan große Anstrengungen unternehmen und die maschinelle Sprachverarbeitung massiv fördern. Dort sind Forschungsinstitute mit bis zu 70 Mitarbeitern keine Seltenheit. Auch in Frankreich und in Großbritannien wurden neue Forschungsinstitute auf diesem Gebiet eingerichtet.

Aktivitäten der Universität Tübingen

Der Schwerpunkt Natürlich-sprachliche Systeme ist in Tübingen an der Forschungsstelle für natürlich-

sprachliche Systeme (FNS), einer Einrichtung der Neophilologischen Fakultät, angesiedelt. Die FNS wird seit Mitte 1985 für die Dauer von fünf Jahren als Forschungsschwerpunkt auf dem Gebiet der Verarbeitung natürlicher Sprache und linguistischen Grundlagenforschung gefördert. Die Förderung umfaßt drei wissenschaftliche Mitarbeiter, eine halbe Stelle für Gastprofessoren, Hilfskräfte sowie Sachmittel. Zu den Arbeitsbereichen gehören neben der Grundlagenforschung in den Kernbereichen der Linguistik (Syntax, Semantik und Pragmatik) bestimmte Teilbereiche der Informatik, insbesondere der linguistischen Informationsverarbeitung.

Die FNS ist an verschiedenen durch Drittmittel geförderten Forschungsprojekten beteiligt:

- i) zur Entwicklung von Expertensystemen (Kooperationsprojekt mit dem Wissenschaftlichen Zentrum der IBM in Heidelberg);
- ii) zur Entwicklung von Methoden der Wissensrepräsentation und des Logikprogrammierens (COST-Projekt der EG in Zusammenarbeit mit dem Dept. of Computing, Imperial College, London);
- iii) zu Problemen der Textanalyse und -verarbeitung (im Verbund mit den Universitäten Hamburg, Osnabrück und Stuttgart sowie der IBM); und
- iv) zur inferentiellen Erschließung von Information aus Texten in einem von der National Science Foundation geförderten Projekt in Zusammenarbeit mit der Universität Georgia, USA)

Der Hauptarbeitsbereich und Schwerpunkt an der FNS ist dabei die natürlich-sprachliche Schnittstelle zu "intelligenten" Systemen; zur Informationsgewinnung aus Texten bzw. der Umsetzung von Informationen in Texte.

Daraus ergeben sich theoretische Probleme folgender Art:

- 1) Die formale Analyse der "Bedeutung" natürlich-sprachlicher Ausdrücke vor dem Hintergrund vorhandenen Wissens. In diesem Bereich muß das Instrumentarium der formalen Semantik dahingehend weiterentwickelt werden, daß neben einzelnen Sätzen auch textuelle Zusammenhänge erfaßt werden können.
- 2) Die Manipulation dieses Wissens im Sinne von "Schlußfolgerungen" bzw. der Informationserweiterung durch Texteingabe. Neben den relationalen Datenbanken, in denen Fakten explizit genannt werden, benötigen "intelligente" Systeme den Zugriff auf deduktive Datenbanken z.B. Methoden des Umgangs mit Regeln und Fakten), die neue Fakten durch Inferenzen erzeugen.
- 3) Die theoretischen Aspekte des Logikprogrammierens in der Anwendung auf unseren Gegenstandsbereich.

Was betrifft die Situation der Eigenschaften bestimmter in den anderen Themenbereichen verwen-

deter Formalismen der semantischen und logischen Repräsentation mit den Mitteln der in den meisten Ansätzen verwendeten Programmiersprache Prolog.

Praktische Anwendungen der erzielten theoretischen Fortschritte werden derzeit in mehreren Pilotstudien erprobt. Zu nennen sind hier Programme und Methoden zur Übersetzung von natürlichsprachlichen Sätzen und Texten in semantische Repräsentationen; juristische Expertensysteme zu Fragen des Straßenverkehrsrechts und des Arbeitsförderungsgesetzes; Programme zum computerstützten Unterricht sowie Inferenzmethoden im Rahmen bestimmter Logikkalküle.

Besonderen Wert legt die FNS auf die interdisziplinäre Zusammenarbeit mit Linguisten, Mathematikern, Informatikern, Psychologen und Philosophen an der Universität Tübingen. Mit dem Studiengang für "Allgemeine Sprachwissenschaft und Nebenfächer" wirkt sich diese Zusammenarbeit bereits auch in der Lehre aus.

Kontakte

Zu einer Reihe von Forschungseinrichtungen unterhalten die Stuttgarter und Tübinger Wissenschaftler seit Jahren sehr gute Kontakte, insbesondere mit der Stanford University/USA, der Universität Edinburgh, der Universität Kyoto, der Universität Marseille und der Universität Amsterdam.

Redaktion (aus Pressemitteilungen der Universitäten Stuttgart und Tübingen)

FORSCHUNGSZENTRUM FÜR KÜNSTLICHE INTELLIGENZ

Entscheidung für Standort Kaiserslautern / Saarbrücken gefallen

Die Weichen für das zukünftige neue Zentrum für Künstliche Intelligenz sind gestellt. In ausführlichen Beratungen mit Vertretern der Industrie im BME I am 15. und 16. Januar 1987 hat ein Kreis von Industrieunternehmen eine mehrheitlich industrielle Trägerschaft für die Errichtung des Zentrums für Künstliche Intelligenz beauftragt und Interesse bekundet, sich an dieser Trägerschaft zu beteiligen.

Bundesforschungsminister Dr. Heinz Riesenhuber erklärte am 27. März hierzu, er begrüßte sehr das zustandgekommene Industrieinteresse für das Zentrum für Künstliche Intelligenz. Damit sei ein wichtiger Meilenstein für eine stärkere Zusammenarbeit zwischen Industrie und Wissenschaft für die notwendige Konzentration der Forschung auf diesem zukunftssträchtigen Gebiet in Sicht. Er sei bereit, das Zentrum für Künstliche Intelligenz mit substanzieller Projektförderung zu unterstützen, damit dort international anerkannte Spitzenforschung betrieben werden könne, deren Ergebnisse die Wissenschaft anspornen und von der Industrie aufgegriffen werden.

Wegen der zukünftigen Marktchancen findet der Bereich der Künstlichen Intelligenz (KI) international zunehmend Beachtung. Expertensysteme sind als wissensbasierte Systeme eine der wichtigsten

und vielversprechendsten Anwendungen von Methoden der Künstlichen Intelligenz. Obwohl in letzter Zeit sehr viel über Expertensysteme veröffentlicht wird, gibt es derzeit weltweit nur ganz wenige Systeme im praktischen Einsatz. Dies ist eines der Indizien dafür, daß – nicht nur für Expertensysteme – noch in großen Umfang grundlagenorientierte Forschung für Methoden und Verfahren der Künstlichen Intelligenz und ihrer praxisorientierten Anwendung betrieben werden muß, um dieses Gebiet zu einer wirklich tragfähigen und breit anwendungsfähigen Technologiebasis auszubauen. Das neue Zentrum für Künstliche Intelligenz soll hierzu in den kommenden 10 Jahren einen zukunftsorientierten Beitrag leisten, der durch seinen vorwettbewerblichen Charakter ein breites industrielles Interesse findet.

In dem Gespräch am 15./16. Januar hatten auch die sich um den Standort des Zentrums für Künstliche Intelligenz bewerbenden Bundesländer Baden-Württemberg, Bayern, Hamburg sowie Rheinland-Pfalz gemeinsam mit Saarland Gelegenheit, ihre Vorschläge vorzutragen und zur Diskussion zu stellen, um in Überlegungen für den am besten geeigneten Standort eintreten zu können. Aufgrund einer gemeinsam durchgeführten Standortbewertung bestand bei den Industrievertretern nach sorgfältiger Abwägung der jeweiligen Gegebenheiten Einigkeit darüber, daß der Standort Kaiserslautern/Saarbrücken für das Zentrum für Künstliche Intelligenz der geeignete Ausgangspunkt für die weiteren Verhandlungen ist.

Als nächster Schritt ist geplant, mit einem erweiterten IndustrieKreis die in einigen Monaten vorgesehene Gründungsversammlung für das Zentrum für Künstliche Intelligenz vorzubereiten. Die Gesellschaft für Mathematik und Datenverarbeitung (GMD) und die Fraunhofer-Gesellschaft (FHG) werden voraussichtlich ebenfalls einbezogen. Gleichzeitig soll ein Förderverein gegründet werden, um darüberhinaus eine größere Zahl von Industrieunternehmen den Dialog mit dem KI-Zentrum zu ermöglichen.

Pressemitteilung des Bundesministers für Forschung und Technologie, Dr. Heinz Riesenhuber, vom 27. März 1987

OFFENER BRIEF ZUR EINRICHTUNG DES KI-ZENTRUMS

Betroffene Wissenschaftler an den Universitäten Kaiserslautern und Saarbrücken beteiligen sich offensiv an der Diskussion um Gestaltung und Zielsetzung des geplanten KIZ

Viele Leser des LDV-Forum werden gleichzeitig den KI-Rundbrief beziehen. Spätestens nach der Lektüre der Nummer 46 vom Oktober 1986 werden diese von der geplanten Einrichtung eines Forschungszentrums für Künstliche Intelligenz (KIZ) in Kaiserslautern und Saarbrücken informiert gewesen sein: ein entsprechender Bericht, basierend auf einer aktuellen Pressemitteilung des BMFT, ist in dieser Ausgabe abgedruckt. Was aus dieser Darstellung, im Gegensatz zu den Stellungnahmen im KI-Rundbrief, nicht hervorgeht, ist die Tatsache, daß sowohl die geplante Einrichtung als insbesondere der Verlauf der Planung

selbst äußerst kontrovers diskutiert werden. Es ist nicht übertrieben, wenn von der ersten wirklich ersten Belastungsprobe des "KI-Lagers" gesprochen wird.

Die betroffenen wissenschaftlichen Mitarbeiter an den Universitäten sind nun in die Offensive gegangen, um für attraktive Arbeitsbedingungen und eine sachdienliche inhaltliche Forschungsorientierung zu sorgen. In einem offenen Brief vom 9. April wurden entsprechende (weitgehend formale) Bedingungen an die Arbeitssituation und -ausstattung formuliert, die mit einer Unterschriftenliste dem BMFT und den Trägern des KIZ zugeleitet werden sollen. Wissenschaftler, die diese Aktion mit eigenen Unterschriftenlisten unterstützen oder sich zunächst informieren wollen, sollten sich an den Bezieher des KI-mod (e-mail-Version des KI-Rundbriefs) wenden (message 034), an die Redaktion des LDV-Forum oder an Hans-Jürgen Bürkert, Universität Kaiserslautern (e-mail: buerkert@ukl.klr.b.uuep). Gleichzeitig wird nachdrücklich darauf hingewiesen, daß Ideen und Diskussionsbeiträge zu einem Leitmotiv für das KIZ aus Sicht der Wissenschaftler in den Universitäten (bzw. der GMD) dringend erwünscht sind.

Spekulationen, weshalb die eigentliche Primärinformation nicht im LDV-Forum angeboten wird, erübrigen sich – die Nachricht kann einfach zu spät und war technisch nicht mehr aufzunehmen. Es wäre sehr zu wünschen, wenn auch das LDV-Forum speziell die nächste Ausgabe, von seinen Lesern als Diskussionsplattform für eine solche wichtige Meinungsbildung genutzt werden würde.

Gerhard Knorz, FH Darmstadt

PARALLELVERARBEITUNGSMODELLE FÜR PROBLEMLÖSUNGEN IN DER KI

Neues Projekt am Institut für Informatik der Technischen Universität München

Seit August 1986 läuft bei Prof. Dr. W. Brauer am Institut für Informatik der Technischen Universität München ein von der DFG gefördertes Projekt mit dem Titel: *Parallelverarbeitungsmodele für Problemlösungen in der Künstlichen Intelligenz*. Ziel des Projekts ist die Entwicklung und Erprobung von parallelen Problemlösungsansätzen im Bereich der Verarbeitung natürlicher Sprache. Diese Untersuchungen werden im Rahmen der Implementierung eines Frage-Antwort Systems durchgeführt, welches Wegauskünfte gibt. Weitere Informationen können bei Erwin Klöck, Institut für Informatik der Technischen Universität München, Arcisstraße 21, D-8000 München 2 angefordert werden.

Rezensionen

Lenders, Winfried / Willée, Gerd:

LINGUISTISCHE DATENVERARBEITUNG – EIN LEHRBUCH.

Opladen: Westdeutscher Verlag, 1986. 201 Seiten.

Angesichts einer sehr differenzierten Angebots- und Bedarfssituation im Bereich der LDV an den bundesdeutschen Hochschulen (ausgebaute Studiengänge mit Tradition einerseits, Aufbaufänge andererseits und schließlich noch Interesse-Bekundungen) sind in letzter Zeit mehrere Publikationen mit einführenden Intentionen erschienen, die jedoch alle jeweils andere Schwerpunkte setzen und sehr unterschiedlichen Ansprüchen und Erwartungen gerecht zu werden versuchen. Das Anwenderinteresse am Gebiet der LDV ist enorm gewachsen; es fehlte bislang jedoch an Einführungen, die diesen Titel verdienen. Dazu werden – eigene mühevollte Erfahrung! – praxisorientierte Basisinformationen benötigt. Das Buch von *Lenders/Willée* wendet sich an Lehrende und Lernende, die zur Erforschung von Sprache und im Umgang mit Sprache die Hilfsmittel der Datenverarbeitung einsetzen wollen. Ihnen sollen die grundlegenden Verfahren der Verarbeitung von Texten thematisch und praktisch vermittelt werden, ... (S. 10) Als "Verarbeitung von Texten" ist aber offenbar nicht auch "Textverarbeitung" zu verstehen, da diese kurz vorher von der LDV abgegrenzt wird.

Im ersten Kapitel (Kommunikationstechnologie und Sprache) werden zunächst allgemeine theoretische Überlegungen zur Beziehung zwischen Sprache und moderner Kommunikationstechnologie angestellt. Von hier wird zur Maschinellen Übersetzung als einem Hilfsmittel der internationalen Kommunikation übergeleitet, deren drei Grundschritte (Analyse, Transfer, Synthese) in aller Kürze abgehandelt werden. In Kapitel 2 (Linguistische Grundlagen) werden Möglichkeiten des Computereinsatzes bei sprachlichen Analyseproblemen erörtert (z.B. Lemmatisierung, morphologische Operationen, Laut- und Silbensegmentierung). Mit der grundlegenden Methodik von maschineller Sprachverarbeitung (Algorithmen) beschäftigt sich Kapitel 3 (Verarbeitungstechniken); an Einzelbeispielen wird das Verfahren der Blockdiagrammdarstellung bei der Problemanalyse erläutert.

Die Vorteile des Einsatzes eines Computers in den in Kapitel 2 und 3 angesprochenen Problemfeldern werden in Kapitel 4 an relevanten Aufgaben deutlich gemacht, alles Arbeiten, die früher ohne Maschine nur mit großem, kostspieligem Arbeitsaufwand und der Gefahr der Fehlerhaftigkeit erledigt werden konnten.

Den entscheidenden Schritt zur Praxis linguistischer Arbeit mit dem Rechner will dann Kapitel 5 tun (Praktische Umsetzung von Algorithmen in Programme). Hier werden Grundstrukturierungen von Computerprogrammen und deren Verwendungsmöglichkeiten an interessanten Beispielen erläutert. Daß die Autoren sich dabei auf die Verwendung einer der möglichen Programmiersprachen festlegen (hier: PL/1), ist selbstverständlich, und die Auswahl ist auch begründlich, da die Autoren Zugang zu einer IBM-Anlage haben. Diese Entscheidung ist dennoch bedauerlich, da PL/1 m.E. durchaus nicht so stark verbreitet und so leicht erlernbar ist, wie behauptet, vor allem aber weil diese Sprache auf anderen als IBM-Anlagen oft nicht optimal implementiert und daher mit größeren Schwierigkeiten bei ihrer Verwendung zu rechnen ist. Schade ist auch, daß sich die Autoren mit dem Lehrbuch zwar an Interessenten wenden, die "sich zum ersten Mal mit den Möglichkeiten der Datenverarbeitung" (S. 12) befassen, später jedoch eingestehen müssen, daß "Grundkenntnisse und eine gewisse Vertrautheit mit PL/1" vorausgesetzt werden. Einem Anfänger kann man das Buch daher als erste Informationsquelle nicht vorbehaltlos empfehlen; gleichwohl haben die Verfasser große Verdienste damit erworben, sich mit großer Sachkenntnis und interessanter Darstellung auf eine Expedition in begehrter werdendes Neuland begeben zu haben. Für eine Neuauflage, die dieses Buch sicherlich bald erhalten wird, wäre eine konsequentere Orientierung auf die angesprochene Zielgruppe wünschenswert.

Martina Schwanke, Universität Kiel

Gerhard Lustig (Hrsg.):

AUTOMATISCHE INDEXIERUNG ZWISCHEN FORSCHUNG UND ANWENDUNG*Linguistische Datenverarbeitung, Band 5. Hildesheim: Olms 1986. 182 Seiten.*

In der GLDV-Reihe *Linguistische Datenverarbeitung* im Olms-Verlag ist der Band "Automatische Indexierung zwischen Forschung und Anwendung", herausgegeben von G. Lustig, erschienen. In diesem Band werden Vorgehen und Ergebnisse der Darmstädter Projekte WAI, AIR und AIR/PHYS, die seit 1977 zum Gebiet der automatischen Indexierung in Darmstadt durchgeführt wurden, zusammenfassend dargestellt.

Mit den Darmstädter Projekten ist es gelungen, einen Forschungsansatz von den ersten experimentellen Stadien bis zu einem System weiterzuentwickeln, das unter realen Arbeitsbedingungen eingesetzt wurde. Dieser Einsatz wurde in einem großangelegten Retrievaltest vorbereitet und konnte weiter forschend begleitet werden.

In dem Band sind in Beiträgen der jeweiligen Bearbeiter die einzelnen Komponenten beschrieben.

Im ersten Beitrag zur Konzeption des Darmstädter Verfahrens von G. Lustig werden die wesentlichen Faktoren genannt, die das Konzept in die Landschaft des Information Retrieval einordnen. Indexierung wird konsequent als Problematik der Behandlung von Massendaten, wie sie in den gewählten Anwendungsbereichen (realer Informationssysteme) FSTA und Physik anfallen, verstanden. Indexierung bedeutet damit den Übergang von dem Text englischsprachiger Referate zu einem vorgegebenen Deskriptorensystem. Diese Aufgabe wurde und wird auch im größeren Rahmen bis heute überwiegend intellektuell gelöst. Die erste wesentliche Komponente bei der Automatisierung der Indexierung ist die Erstellung eines Indexierungswörterbuches, das linguistische Aspekte (Anweisungen zur Identifikation von Grund- und Stammformen sowie Mehrwortgruppen), dokumentarische Aspekte (z. B. Zuordnung von Facetten zu Deskriptoren) und statistische Aspekte (Deskriptorenfrequenzen zur Signifikanzabschätzung) in sich vereinigt. Wesentlich ist die Gewinnung von sog. z -Relationen zwischen Texttermen und Deskriptoren, die auf der Basis intellektuell indexierter Dokumente gewonnen und im Wörterbuch notiert werden. Die automatische Indexierung setzt eine Relevanzbeschreibung (z -Werte) der zu indexierenden Dokumente voraus, nach konkurrierenden Verfahren wird dann die Menge der zuzuordnenden Deskriptoren ausgewählt.

Gegenstand des zweiten Kapitels ist die Entwicklung von Indexierungswörterbüchern (Gewinnung von z -Werten, aufschlußreiche Zahlen zum partiellen Parsing und zur Formelbehandlung in großen Textmengen). Anschließend werden im dritten Kapitel die beiden (im folgenden Retrievaltest konkurrierenden)

Indexierungsverfahren nach dem Booleschen und nach dem Polynomansatzverfahren dargestellt.

Die Beschreibung des Retrievaltests in Kapitel vier macht in der Bewertung der Ergebnisse dieses doch recht umfangreichen Retrievaltests die begrenzte Aussagefähigkeit der gängigen Bewertungsmaße deutlich. An einigen Punkten, wie z. B. der geringen Überlappung der intellektuell und der automatisch erschlossenen relevanten Antwortdokumentmengen, zeigt sich, daß der Bereich realer Retrievalpraxis und ihrer Bewertung noch keineswegs abschließend geklärt ist. Das Zusammenwirken von automatischer und intellektueller Indexierung, von Deskriptor- und Freitextsuche, die Berücksichtigung weiterer Faktoren wie Indexierungstiefe, verschiedener Benutzer- oder Fachgebietsaspekte weisen auf ein Arbeitsgebiet hin, das bisher – sicher aufgrund der Schwierigkeiten, die ein großangelegter Retrievaltest mit sich bringt – keineswegs voll erschlossen ist, jedoch ein breites Feld für weitere Untersuchungen darstellt.

N. Fuhr geht dem Aspekt des Ranking mit gewichteter Indexierung im anschließenden Abschnitt nach, die dargestellten Ergebnisse (Tanimoto und Cosinus-Maß vs. Fuzzy Retrieval und Extended Boolean Retrieval) wecken die Neugier auf seine vor kurzem abgeschlossene Dissertation, die auf den Daten des Retrievaltests beruht. Abschließend diskutiert G. Knorz die Möglichkeiten und Bedingungen des Einsatzes der automatischen Indexierung, die er (analog zur Entwicklung in anderen Bereichen der maschinellen Sprachverarbeitung) in einer intelligenten Kooperation mit dem (menschlichen) Indexierer sieht, der so von Routineaufgaben entlastet, Freiraum für (immer nötige) Konsistenzprüfungen und Adaptionen an neue Entwicklungen erhält.

Insgesamt bietet der Band eine zusammenhängende Dokumentation zu dem gesamten Darmstädter Indexierungsprojekt und zeigt Ergebnisse, Probleme und Dimensionen auf, mit denen unter realen Einsatzbedingungen zu rechnen ist.

Die diffusen Ergebnisse des Retrievaltests rühren möglicherweise von der Konstellation der verwendeten Indexierungsverfahren (Adaption des automatischen Verfahrens an der intellektuellen Indexierung) her. Bereits mit den im Band selbst angesprochenen Fragestellungen (s. o.) zeichnet sich jedoch die Möglichkeit ab, mittels weiterer analytischer Untersuchungen anhand der Daten des Retrievaltests Aufschlüsse über Indexierungs- und Retrievalprozesse zu gewinnen. Wie sich auch hier wieder gezeigt hat, ist eine Weiterentwicklung der Forschung auf diesem Gebiet auf der Grundlage globaler Bewertung kaum möglich; wohl nur durch die Analyse der unter realen / realitätsnahen Bedingungen ablaufenden Vorgänge ist ein Aufbrechen der Komplexität des Information Retrieval und damit eine Weiterentwicklung zu leisten.

Christine Schneider, *Linguistische Informationswissenschaft, Regensburg*

Publikationen

Für die Bibliographie "NEUERE VERÖFFENTLICHUNGEN IM BEREICH DER LINGUISTISCHEN DATENVERÄRBEITUNG" werden dem LDV-FORUM vom

GID-Informationszentrum (GID-IZ)
Lyoner Strasse 44-48
D-6000 Frankfurt 71

regelmäßig die Standardprofile Linguistische Datenverarbeitung und Künstliche Intelligenz der Datenbasis INFODATA zur Verfügung gestellt. Arbeiten, die vor 1985 erschienen sind, werden nicht berücksichtigt. Arbeiten, deren Bezug zu dem Bereich "natürliche Sprache" nicht erkennbar ist, werden ebenfalls nicht übernommen. Dies gilt insbesondere für solche Veröffentlichungen, die sich in nur sehr allgemeiner Weise mit Wissensrepräsentation beschäftigen (die Grenze ist fließend und verständlicherweise subjektiv gezogen).

Die Zitate beschränken sich 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, D-6000 Frankfurt 71
Tel.: (069) 66 87-374 oder 66 87-373

Verarbeitete Ausschnitte der INFODATA-Datenbank: Standardprofile vom zweiten Halbjahr 1986 (ab Erf-Jahr 8606, Erf-Nr 1601 bis Erf-Jahr 8611, Erf-Nr 4000).

Aufbau einer Dokumentationseinheit:

Autor, A.: Der Titel in Fettdruck, mit einem Punkt abgeschlossen. Die Quelle, mit Seitenangabe, (bei aufeinanderfolgenden Arbeiten mit gleicher Quelle wird diese nur beim ersten Zitat angegeben, die folgenden Zitate werden eingerückt), Deskriptor₁; Deskriptor₂; ...; Deskriptor_n, Bestellnummer

Bidd, D.; Chevigny, L. de: **PRECIS for subject access in a national audiovisual information system.** Canadian library journal, 43(1986)3, S. 177-184. *Automatic indexing; Index preparation; Information service; Film; Audio-visual document; Canada 86-03996*

Yong-yuan, Y.: **GID: Aufbau eines deutsch-chinesischen Terminologie-Datenbanksystems. Zwischenbericht.** Stand: Aug. 1985. Frankfurt (1985) 13 S. *Database; Terminology; Database organization; German; Chinese; Microcomputer; Reprography; Dictionary compilation; Testing 86-01603*

Singh, M. G.; Cook, R.: **A new class of intelligent knowledge-based systems with an optimisation-based inference engine.** Decision support systems, 1(1985)4, S. 299-312. *Expert system; Decision support system; Inference; Draft 86-01635*

Pau, L. F.: **Inference of the structure of economic reasoning from natural language analysis.** Decision support systems, 1(1985)4, S. 313-321. *Natural language processing; Artificial intelligence; Text analysis; Computational linguistics; Inference; Model 86-01636*

Wingert, F.: **An indexing system for SNOMED.** Methods of information in medicine, Stuttgart, 25(1986)1, S. 22-30. *Automatic indexing; System of concepts; Medicine; Algorithms 86-01683*

Mosier, J. N.; Smith, S. L.: **Application of guidelines for designing user interface software.** Behaviour and information technology 5(1986)1, S. 39-46. *Guideline; User support; Man-machine communication; Software; User friendly; Inquiry 86-01725*

Smith, S. L.: **Standards versus guidelines for designing user interface software.** Behaviour and information technology, 5(1986)1, S. 47-61. *Guideline; Standard; Comparison; Man-machine communication; User support; Software; User friendly; Human factor; Workplace design 86-01726*

Navathe, S. B.; Kerschberg, L.: **Role of data dictionaries in information resource management.** Information and management, 10(1986)1, S. 21-46. *Database system; Draft; Information management; Systems research; Model; Semantics; Requirement 86-01766*

Nagao, M.; Tsujii, J. I.; Nakamura, J. I.: **The Japanese government project for machine translation.** Computational linguistics, 11(1985)2/3, S. 91-110. *Automatic translation; Governmental agency; Computational linguistics; Text analysis; Japanese; English; Abstract; Dictionary compilation; Software; Evaluation; Project; Japan 86-01777*

Bennett, W. S.; Slocum, J.: **The LRC machine translation system.** Computational linguistics, 11(1985)2/3, S. 111-121. *Automatic translation; Information system; Computational linguistics; Project; Texas; USA METAL 86-01778*

Vasconcellos, M.; Leon, M.: **SPANAM and ENGSPAN - Machine translation at the Pan American Health Organization.** Computational linguistics, 11(1985)2/3, S. 122-139. *Automatic translation; Information system; Computational linguistics; Health service; Medicine; Corporation; Spanish; English 86-01779*

Brewer, A.; Feneyrol, C.; Ritzke, J.; Stegentritt, E.: **ASCOF - a modular multilevel system for French-German translation.** Computational linguistics, 11(1985)2/3, S. 137-154. *Automatic translation; Information system; Computational linguistics; Text analysis; French; German; Federal Republic of Germany 86-01780*

- Johnson, R.; King, M.; Tombe, L. des: **EUROTRA: a multilingual system under development.** Computational linguistics, 11(1985)2/3, S. 155-169. *Automatic translation; Information system; Computational linguistics; Software; Prototype; EC; Project; EC-countries* 86-01781
- A machine(-aided) translation bibliography.** Generally restricted to currently accessible documents written in English, French or German during the years 1973-1984. Computational linguistics, 11(1985)2/3, S. 171-183. *Automatic translation; Publication; English; French; German* 86-01782
- Barton, G. E.: **On the complexity of ID/LP parsing.** Computational linguistics, 11(1985)4, S. 205-218. *Computational linguistics; Linguistics; Grammar; Text analysis; Algorithm* 86-01783
- Jacobs, P. S.: **PHRED: a generator for natural language interfaces.** Computational linguistics, 11(1985)4, S. 219-242. *Natural language processing; Man-machine communication; Computational linguistics; Text analysis; English; Spanish* 86-01784
- Bischoff, R.: **Software-Ergonomie - Einführung, Überblick, Akzentuierungen.** Handbuch der modernen Datenverarbeitung, 22(1985)126, S. 3-20. *Man-machine communication; Software; Information-seeking behavior; Human factor* 86-01791
- Gunzenhäuser, R.; Knopik, T.: **Wissensbasierte Mensch-Computer-Schnittstellen in der Software-Ergonomie.** Handbuch der modernen Datenverarbeitung, 22(1985)126, S. 119-128. *Man-machine communication; Software; Artificial intelligence; Computer aided learning* 86-01797
- Lehmann, E.: **Natürlichsprachliche Zugangssysteme.** Handbuch der modernen Datenverarbeitung, 22(1985)126, S. 137-142. *Man-machine communication; Speech input; Fact documentation; Geography* 86-01799
- Hoepelman, J.; Rigoll, G.: **Sprachverarbeitung. Die Integration von Signalverarbeitung und künstlicher Intelligenz.** FhG-Berichte, München (1986)1, S. 26-29. *Speech processing; Computational linguistics; Speech recognition; Speech output; Project* 86-01809
- Wagner, E.: **Post-editing SYSTRAN. A challenge for commission translators.** Terminologie et traduction, (1985)3, S. 1-7. *Automatic translation; Correction; Method; Translation; Translator; Information-seeking behavior* 86-01811
- Pigott, I.: **Machine translation as an integral part of the electronic office environment.** S. 9-15. *Automatic translation; Office organization; Microcomputer; Compatibility* 86-01812
- Stamoulis, C.: **The terminological component of the French-English M.T. system SYSTRAN (franz.).** S. 17-29. *Automatic translation; Terminology; Evaluation; Language statistics; Text analysis; Dictionary; French; English; Testing* 86-01813
- Lau, P.: **Elements of EUROTRA today.** S. 31-39. *Automatic translation; Software; Generative Grammar; Computational linguistics* 86-01814
- Pilgrim, J.: **On the interaction of man and EDP use as work activity.** International journal of man-machine studies, 22(1985)5, S. 493-505. *Man-machine communication; User friendly; DP; Trend* 86-01816
- Scagliola, C.: **Language models and search algorithms for real-time speech recognition.** International journal of man-machine studies, 22(1985)5, S. 523-547. *Speech processing; Speech recognition; Algorithms; Language; Model* 86-01818
- Forsyth, R.; Rada, R.: **Machine learning. Applications in expert systems and information retrieval.** Chichester: Horwood u. New York: Halsted Press (1986) 277 S., *Expert system; Learning; Method; Information retrieval; Automatic indexing; Algorithm; Case study; Medicine* 86-01829
- Kallmeyer, W. (Hrsg.): **Kommunikationstypologie - Handlungsmuster, Textsorten, Situationstypen. Jahrbuch 1985 des Instituts für deutsche Sprache.** Institut für deutsche Sprache, Düsseldorf: Schwann (1986) 429 S., *Communication science; Linguistics; Language; Typology; Human factor* 86-01884 (nicht ausleihbar!)
- Bátori, I. (Hrsg.); Weber, H. J. (Hrsg.): **Neue Ansätze in maschineller Sprachübersetzung. Wissensrepräsentation und Textbezug.** Tübingen: Niemeyer (1986) 277 S. *Automatic translation; Text analysis; Computational linguistics; Knowledge representation; Artificial intelligence* 86-01889
- Batori, I.: **Paradigmen der maschinellen Sprachübersetzung.** S. 3-27. *Automatic translation; DP; Computational linguistics; Syntax; Grammar; Artificial intelligence; Research* 86-01890
- Pause, P. E.: **Zur Modellierung des Übersetzungsprozesses.** S. 45-74. *Automatic translation; Translator; Text analysis; Semantics; Natural language processing* 86-01891
- Rohrer, C.: **Maschinelle Übersetzung mit Unifikationsgrammatiken.** S. 75-99. *Automatic translation; Grammar; Dictionary; Semantics; Phrase-structure grammar* 86-01892
- Boitet, C.; Gerber, R.: **Expert systems and other new techniques in M(a)T.** S. 103-119. *Automatic translation; Expert system; Software engineering; Grammar; Knowledge representation* 86-01893
- Rösner, D.: **SEMSYN - Wissensquellen und Strategien bei der Generierung von deutsch aus einer semantischen Repräsentation.** S. 121-137. *Automatic translation; Title; Semantic network; Knowledge representation; Frame; Dictionary; Japanese; German* 86-01894
- Saluveer, M.; Oim, H.: **Rules and reasoning in text comprehension.** S. 139-163. *Text analysis; Communication theory; Inference; Logic; Automatic translation* 86-01895
- Hauenschild, C.: **KIT/NASEV oder die Problematik des Transfer bei der maschinellen Übersetzung.** S. 167-195. *Automatic translation; Knowledge representation; Semantics; Phrase-structure grammar; Logic; Text analysis KIT/NASEV* 86-01896
- Rothkegel, A.: **Textverstehen und Transfer in der maschinellen Übersetzung.** S. 197-227. *Automatic translation; Text analysis; Syntax; Case grammar; Frame; Data structure* 86-01897

- Weber, H. J.: **Faktoren einer textbezogenen maschinellen Übersetzung. Satzstrukturen, Kohärenz- und Koreferenz-Relationen, Textorganisation.** S. 229-261. *Automatic translation; Text analysis; Syntax; Dictionary; Semantics; Artificial intelligence; German; English COAT; OVERCOAT* 86-01898
- Tou, J. T.: **Knowledge engineering revisited.** *International journal of computer and information sciences*, 14(1985)3, S. 123-133. *Expert system; Typology; Knowledge representation; Inference* 86-01906
- Biermann, A. W.; Fineman, L.; Gilbert, K. C.: **An imperative sentence processor for voice interactive office applications.** *ACM transactions on office information systems*, 3(1985)4, S. 321-346. *Speech processing; Natural language processing; Text analysis; Command language; Word processing; Office organization; Prototype* 86-01908
- Peels, A. J. H. M.; Janssen, N. J. M.; Nawijn, W.: **Document architecture and formatting.** *ACM transactions on office information systems*, 3(1985)4, S. 347-369. *Office organization; Word processing; Text; Text analysis; Unification* 86-01909
- ADBS; ANRT (Veranst.): **IDT 85 - information, documentation, knowledge transfer. Texts of lectures Versailles june 1985, 6th national congress for information and documentation.** Paris (1985) 279 S. *Information policy; Computational linguistics; Expert system; Technology transfer; Microcomputer; Information retrieval system; Documentalist; Information system; Library automation; Database; Training; France* 86-01916
- Laurian, A. M.: **Post-editor, a new profession linked with knowledge transfer through machine translation (in franz.).** S. 13-17. *Automatic translation; Translation; Evaluation; Job outline; Training; Translator; SYSTRAN* 86-01918
- Zajac, R.: **Interactive edition of technical documents with regard to machine translation (in franz.).** S. 18-24. *Automatic translation; Computational linguistics; Grammar; Man-machine communication; Error; Correction* 86-01919
- Martin, E.: **A database about the French language (in franz.).** S. 25-32. *Database; Lexicology; Fulltext database; French* 86-01920
- Buchanan, B. G.: **Expert systems: working systems and the research literature.** *Expert systems*, Abingdon 3(1986)1, S. 32-51. *Expert system; Market review; Artificial intelligence; Trend; Literature, Research and development* 86-01981
- Kehoe, C. A.: **Interfaces and expert systems for online retrieval.** *Online review*, 9(1985)6, S. 489-505. *Man-machine communication; User support; Expert system; Information retrieval; History; Trend* 86-02010
- Hawkins, D. T.; Levy, L. R.: **Introduction of online searching to end users at AT and T Bell Laboratories.** In: Williams, M. E.; Hogan, T. H., Online Review (Veranst.): National online meeting 1986. Proceedings of the seventh National Online Meeting, New York, May 1986. Medford: Learned Information (1986) S. 167-174. *User training; Dialog system; Man-machine communication; Software; Microcomputer* 86-02130
- Schneider, H. J. (Proj.); Emde, W.; Fuernsinn, M.; Günther, S.; Khenkhar, M.; Kilbury, J.; Luck, K. von; TU Berlin (FB Informatik, Inst. für angewandte Inf.): **Textverstehen. Interenssensgesteuerte Variation der Analysetiefe. Schlussbericht.** Berlin, Dez.(1985) 119 S. *Artificial intelligence; Text analysis; Computational linguistics; Knowledge representation; Method; Algorithm; Learning* 86-02191
- Thiel, U.; Hammwöhner, R.; Kuhlen, R. (Proj.); Univ. Konstanz, Bereich Informationswissenschaft: **TOPOGRAPHIC. Die graphisch interaktive Schnittstelle des Textkondensierungssystems TOPIC - Bericht zur ersten Projektphase.** Konstanz, Dez.(1985) 82 S. *Artificial intelligence; Natural language processing; Man-machine communication; User support; Text analysis; Software TOPOGRAPHIC; TOPIC* 86-02192
- Greenspan, S. J.; Borgida, A.; Mylopoulos, J.: **A requirements modeling language and its logic.** *Information systems*, 11(1986)1, S. 9-23. *Systems research; Model; Semantics; Time factor; Artificial language; Software engineering; Knowledge representation* 86-02197
- Mahapatra, M.; Biswas, S. C.: **Interdependence of PRECIS role operators. A quantitative analysis of their associations.** *Journal of the American Society for Information Science*, 37(1986)1, S. 20-25. *Classification research; Automatic indexing; Role indicator; Evaluation; Empirical study* 86-02234
- Fidel, R.: **Towards expert systems for the selection of search keys.** *Journal of the American Society for Information Science*, 37(1986)1, S. 37-44. *Information retrieval; Search strategy; Software; Decision support; Model; Expert system* 86-02236
- Nejt, A.: **Esperanto as the focal point of machine translation.** *Multilingua*, 5(1986)1, S. 9-13. *Automatic translation; Esperanto; Evaluation* 86-02243
- Quicheron, J. B.: **The interpreter and the multilingual obstacles.** *Multilingua*, 5(1986)1, S. 15-19. *Translation; Multilingual; Meeting* 86-02244
- Ammon, R. v.: **Evaluierungsergebnisse zur automatischen deutsch-japanischen Titelübersetzung SUSY-DJT.** *Multilingua*, 5(1986)1, S. 21-29. *Automatic translation; Title; German; Japanese; Evaluation; Inquiry SUSY-DJT* 86-02245
- Gile, D.: **Terminological work at conference translation.** *Multilingua*, 5(1986)1, S. 31-36. *Translation; Terminology; Dictionary; Meeting* 86-02246
- Berman, S.: **A semantic data model as the basis for an Automated Database Design tool.** *Information systems*, 11(1986)2, S. 149-165. *Database system; Database; Draft; Automatic; Model; Semantics; Man-machine communication; Systems research* 86-02272
- Bassiouni, M. A.; Ok, B.: **Double encoding - a technique for reducing storage requirement of text.** *Information systems*, 11(1986)2, S. 177-184. *Compression; Text; Code; Storage; Method* 86-02274
- Milne, R.: **Resolving lexical ambiguity in a deterministic parser.** *Computational linguistics*, 12(1986)1, S. 1-12. *Text analysis; Syntax; Context* 86-02281

- Fink, P. K.; Biermann, A. W.: The correction of ill-formed input using history-based expectation with applications to speech understanding. *Computational linguistics*, 12(1986)1, S. 13-36. *Speech processing; Speech recognition; Error; Correction; Information-seeking behavior; Testing; Method* 86-02282
- Hahn, U.; Univ. Konstanz, Bereich Informationswissenschaft: Industrialisierungstendenzen in der Künstlichen Intelligenz. Wissensindustrielle und förderpolitische Entwicklungen auf dem Markt für intelligente Informationssysteme. Konstanz, Feb.(1985) 72 S. *Artificial intelligence; Industry; Information industry; Information economy; Trend; Expert system; Information retrieval system; Comparison* 86-02314
- Hahn, U.; Univ. Konstanz, Bereich Informationswissenschaft: Informationslinguistik 1. Einführung in das linguistische Information Retrieval. Konstanz, Jan.(1985) 186 S. *Computational linguistics; Information retrieval; Information system; Morphology; Syntax; Semantics; Method* 86-02326
- Hahn, U.; Univ. Konstanz, Bereich Informationswissenschaft: Informationslinguistik 2. Linguistische und statistische Verfahren im experimentellen Information Retrieval. Konstanz, Jan.(1985) 183 S. *Computational linguistics; Information retrieval; Statistics; Text analysis; Information system; Natural language processing; Method* 86-02327
- Hahn, U.; Univ. Konstanz, Bereich Informationswissenschaft: Intelligente Informationssysteme. Verfahren der Künstlichen Intelligenz im experimentellen Information Retrieval. Konstanz, Feb.(1985) 388 S. *Artificial intelligence; Information retrieval; Expert system; Information system; Knowledge representation; Frame; Logic; Semantic network; Decision support system; Problem solving; Natural language processing; Industry; Trend* 86-02328
- Hahn, U.; Reimer, U.; Univ. Konstanz, Bereich Informationswissenschaft: The TOPIC project. Text-oriented procedures for information management and condensation of expository texts. Final report. Konstanz, 122 S. *Text analysis; Fulltext; Compression; Automatic abstracting; Information system; Computational linguistics; Artificial intelligence; Information management; Model; Method* TOPIC 86-02334
- Hellwig P. (Hrsg.); Lehmann, H. (Hrsg.): GLDV: Trends in der Linguistischen Datenverarbeitung. Beiträge der Jahrestagung 1983. Hildesheim: Olms (1986) 187 S. *Text analysis; Dictionary compilation; Natural language processing; Lexicology; Automatic translation* 86-02402 (nicht ausleihbar!)
- Brustkern, J.; Heinze, G.; Schulze, W.: Die Wortdatenbank des Deutschen. S. 26-33. *Dictionary; Lexicology; German; Data structure; Database system* 86-02403
- Hofmann, J.: Wissensdarstellung für die Behandlung von Thematisierung in Breitbanddialogen. S. 56-70. *Natural language processing; Knowledge representation; Cognition* 86-02404
- Hitzenberger, L.: Möglichkeiten bei der Anwendung phonologischer Verfahren bei der Namensuche in Datenbanken. S. 71-85. *Phonetics; Automatic; Generative Grammar; Name; Searching; Man-machine communication* 86-02405
- Weber, H. J.: Ein dreistufiges Lexikonsystem zur maschinellen Analyse von semantischen Textstrukturen. S. 92-109. *Text analysis; Semantics; Dictionary compilation; Automatic translation* 86-02406
- Kobsa, A.: Generating a user model from wh-questions in the VIE-LANG system. S. 125-140. *Natural language processing; Man-machine communication; Inference* 86-02407
- Guenther, F.; Lehmann, H.: Linguistische und logische Aspekte beim Aufbau eines juristischen Expertensystems. S. 141-150. *Expert system; Knowledge representation; Programming language; Logic; Law* 86-02408
- Maas, H. D.: Das Projekt SUSY-DJT. S. 171-177. *Automatic translation; Title; German; Japanese SUSY-DJT* 86-02409
- Rösner, D.; Lesniewski, A.: SEMSYN - ein Projekt zur japanisch/deutschen (Titel-) Übersetzung. S. 178-187. *Automatic translation; Title; Japanese; German; Knowledge representation; Frame* 86-02410
- Katzer, J.; Oddy, R. N.; Das-Gupta, P.: Syracuse Univ., School of Information Studies, Syracuse: A study of free-index phrases. Final report. Syracuse (1985) 85 S. *Keyword; Indexing techniques; Automatic indexing; Title; Abstract; Search result; Evaluation; Comparison* INSPEC 86-02414
- Winston, P. H. (Hrsg.); Prendergast, K. A. (Hrsg.): The AI business. The commercial uses of artificial intelligence. Cambridge: MIT Press (1985) 324 S. *Expert system; Artificial intelligence; Software engineering; Production engineering; Natural language processing; Man-machine communication* 86-02416
- Davis, R.: Amplifying expertise with expert systems. S. 17-40. *Expert system; Systems research; Trend; Inference; Case study* 86-02417
- Harris, L. R.: Natural language front ends. S. 149-161. *Natural language processing; Man-machine communication; INTELLECT* 86-02421
- Blatt, A.; Freigang, K. H.; Schmitz, K. D.; Thome, G.: Computer und Übersetzen. Eine Einführung. Hildesheim: Olms (1985) 345 S. *Automatic translation; Translation; Computer aided; Translator; Training* 86-02434
- MacGregor, D. R.; Malone, J. R.: An architectural approach to advances in information retrieval. In: *Aslib: Advances in intelligent retrieval. Proceedings of a conference*, S. 35-46. *Database system; Semantic network; Microelectronics; Hardware; Information retrieval* FACT 86-02443
- Enser, P. G. B.: Experimenting with the automatic classification of books. S. 66-83 *Automatic classification; Classification; Manually operated; Stock; Book; Comparison; Effectiveness* 86-02445
- Zarri, G. P.: Interactive information retrieval. An artificial intelligence approach to deal with biographical data. S. 101-119. *Expert system; Knowledge representation; Case grammar; Inference; Information retrieval; Science of history* RESEDA 86-02447
- Hancox, P.; Smith, F.: A case system processor for the PRECIS indexing language. S. 120-147. *Automatic translation; Syntactic indexing; Natural language processing; Case grammar; Semantics; Evaluation* PRECIS 86-02448

- Rouault, J.: Linguistic methods in information retrieval systems. S.148-162. *Information retrieval system; Document; Structure; Database system; Computational linguistics; Natural language processing* 86-02449
- Aragon-Ramirez, V.: Design of a system for the online elucidation of natural language search statements. S. 163-190. *Man-machine communication; Natural language processing; Search strategy; Thesaurus CIRCE* 86-02450
- Aktas, A.; Kämmerer, B.; Küpper, W.; Lagger, H.; Siemens, BMFT, (Auftrg.): Automatische Spracherkennung für große Wortschätze. Schlussbericht. München (1985) 128 S. *Speech recognition; Pattern recognition; Word; Phonetics; Method; Software* 86-02472
- Simons, G. L.: Die fünfte Computer-Generation. Konzepte und Wege - eine Einführung. München: Hanser (1986) 266 S. *Artificial intelligence; Expert system; Hardware; Software; Man-machine communication; Knowledge representation; Systems research; Trend* 86-02475
- Guha, B.; UNESCO (GIP); UNESCO (UNISIST): Study on the language barrier in the production, dissemination and use of scientific and technical information with special reference to the problems of the developing countries. Paris, FR (1985) 66 S. *Information flow; Language barrier; Translation; Literature; Information needs; Developing countries; Trend; India* 86-02487
- Bibel, W.: Wissensbasierte Software-Entwicklung. In: Brauer, W. (Hrsg.); Radig, B. (Hrsg.); GI (FA 6 Künstliche Intelligenz) (Veranst.): Wissensbasierte Systeme. GI-Kongress, München, Oktober 1985. Berlin: Springer (1985) S. 17-41. *Artificial intelligence; Software engineering; Software; System environment PROLOG* 86-02532
- Sowizral, H. A.: Expert systems. Annual review of information science and technology, 20(1985) S. 179-199. *Expert system; Knowledge representation; Cognition; Theorem proving; Programming; Trend* 86-02731
- Hofmann, J.; GLDV: Breitbandiger natürlichsprachiger Mensch-Rechner-Dialog. Skizze eines Verarbeitungsmodells am Beispiel der Thematisierung. Hildesheim: Olms (1985) 306 S. *Computational linguistics; Natural language processing; Man-machine communication; Semantics; Method; Inference; Software* 86-02752
- Jones, K. P.: Getting started in computerized indexing. The indexer, Hitchin 15(1986)1, S. 9-13. *Automatic index preparation; Automatic indexing; Expert system; Trend* 86-02756
- Wellisch, H. H.: Indexing and abstracting. A current-awareness bibliography. The indexer, Hitchin 15(1986)1, S. 29-36. *Indexing; Indexing techniques; Automatic indexing; Abstracting; Abstracting service; Database* 86-02757
- Beghtol, C.: Semantic validity. Concepts of warrant in bibliographic classification systems. Library resources and technical services, 30(1986)2, S. 109-125. *Classification research; Classification characteristic; Semantics; Classification system; History* 86-02801
- Fedorowicz, J.; Williams, G. B.: Representing modeling knowledge in an intelligent decision support system. Decision support systems, 2(1986)1, S. 3-14. *Systems research; Decision support system; Knowledge representation; Frame; Semantic network; Draft* 86-02814
- Endres-Niggemeyer, B. (Hrsg.); GLDV; GI (FA 1,2 Künstliche Intelligenz und Mustererkennung, Fachgruppe 3 Natürlichsprachliche Systeme): Sprachverarbeitung in Information und Dokumentation. Proceedings, Jahrestagung, Hannover, März 1985. Berlin: Springer (1985) 251 S. *Natural language processing; Expert system; Knowledge representation; Frame; Text analysis; Software; Automatic indexing; Automatic abstracting; Information retrieval* 86-02881
- Kuhlen, R.: Verarbeitung von Daten, Repräsentation von Wissen, Erarbeitung von Information. Primat der Pragmatik bei informationeller Sprachverarbeitung. S. 1-22. *Information science; Natural language processing; Semiotics; Pragmatics; Information retrieval; Facts; Knowledge representation; Trend* 86-02882
- Thiel, M.: Eine konzeptionelle Basis für natürlichsprachliche Systeme. S. 23-33. *Natural language processing; Knowledge representation; Grammar; Software* 86-02883
- Rösner, D.: Generierung von Deutsch aus einer semantischen Repräsentation Zu Stand und Perspektiven des Projekts SEMSYN. S. 46-55. *Automatic translation; Generation; Semantics; German; Japanese; Software* 86-02884
- Jarke, M.; Krause, J.: Konzeptualisierung und Training für natürlichsprachliche FAS Zur Reinterpretation des empirischen Materials des 'Advanced Language Project'. S. 56-66. *Natural language processing; Man-machine communication; Project; Evaluation* 86-02885
- Zöpprit, M.: Ein Sprachmodell für die Software-Ergonomie. S. 67-79. *Man-machine communication; Information retrieval; Command language; Category scheme; Software; User friendly; Draft* 86-02886
- Lehmann, H.: Ein juristisches Expertensystem auf der Grundlage von Linguistik und Logik. S. 92-100. *Natural language processing; Expert system; Law; Transportation; Knowledge representation* 86-02887
- Schirra, J. R.; Brach, U.; Wahlster, W.; Woll, W.: WILIE: ein wissensbasiertes Literaturerfassungssystem. S. 101-112. *Artificial intelligence; Data input; Formal description; Literature documentation; Frame; Inference; Knowledge representation; Microcomputer WILIE* 86-02888
- Arz, J.: TRICON: ein System für geometrische Konstruktionen mit natürlichsprachlicher Eingabe. S. 113-123. *Natural language processing; Data input; Figure; Knowledge representation* 86-02889
- Mueller, B. S.; Primio, F. d.: Interrogative Wissensrepräsentation als Grundlage für regelbasierte Expertensysteme. S. 124-140. *Knowledge representation; Model; Expert system; Natural language processing; Problem solving* 86-02890
- Rothkegel, A.: Textlinguistische Komponenten in der Informationserschließung. S. 151-164. *Natural language processing; Text analysis; Content analysis; Software* 86-02891

- Hahn, U.; Reimer, U.: Entwurfsprinzipien und Architektur des Textkondensierungssystems TO-PIC. S. 165-179. *Automatic abstracting; Text analysis; Compression; Method; Software; Knowledge representation; Fulltext searching* 86-02892
- Lustig, G.: Zur Pilotanwendung AIR/PHYS der automatischen Indexierung. S. 191-202. *Automatic indexing; Pilot application; Dialog service; Physics; Classification method; Ranking AIR/PHYS* 86-02893
- Knorz, G.: Unsicheres und unvollständiges Wissen in wissensbasierten Systemen. Kontrolle von Heuristik durch systematische Beobachtung. S. 203-215. *Weighting; Mathematical method; Factual data bank; Information retrieval* 86-02894
- Schwarz, C.: Textrecherche mit Mehrwortbegriffen. S. 216-225. *Information retrieval; Search strategy; Word; Context; Syntax; Method* 86-02895
- Rostek, L.; Fischer, D.: Konzepte für ein thesaurusbasiertes Information Retrieval am Arbeitsplatz. S. 226-243. *Information retrieval; Man-machine communication; Knowledge representation; Syntax; Method; Microcomputer* 86-02896
- Willis, W.: Methodische Probleme der maschinellen Übersetzung. *Lebende Sprachen*, 31(1986)2, S. 49-56. *Automatic translation; Method; Criticism; History; Human factor; Problem solving* 86-02906
- Krause, J.: Silbentrennung und Rechtschreibhilfe. Zum Vordringen sprachanalytischer Werkzeuge in der Textverarbeitung. *Lebende Sprachen*, 31(1986)2, S. 58-60. *Word processing; Orthography; Error; Control (Check); Software; Trend* 86-02907
- Milevski, S. N.: Artificial intelligence for the layman. *Bulletin of the American Society for Information Science*, 12(1986)4, S. 12-14. *Expert system; Trend; Military matters and technology; USA* 86-02910
- Ungvary, R.: Intensional splitting. An empirical examination of conceptual duality. *International classification*, 13(1986)1, S. 9-17. *Linguistics; Classification research; Semantics; Word; Synonym* 86-02913
- Gaines, B. R.; Shaw, M. L. G.: From timesharing to the sixth generation. Part 1 - The development of human-computer interaction. *International journal of man-machine studies*, 24(1986)1, S. 1-27. *Man-machine communication; Systems research; DP; Artificial intelligence; History; Trend* 86-02922
- Newton, S. J.: Voice in office systems. Manchester: NCC Publ. (1985) 77 S. *Speech recognition; Speech output; Method; Office organization; Communication system; Telephony; Electronic mail; Inhouse information* 86-02931
- Choueka, Y.; Lusignan, S.: Disambiguation by short contexts. *Computers and the humanities*, 19(1985)3, S. 147-157. *Computational linguistics; Text analysis; Context; Morphology; Method* 86-02933
- Stübbe, H. L. P.: International bibliographic standards in cartographic archives. *The National Map Collection, Public Archives of Canada, experience Inspel*, 20(1986)1, S. 5-33. *Archival registration; Automatic indexing; Map; Guideline; Canada PRECIS* 86-03023
- Keitz, W. von: Automatic indexing and the dissemination of information. *Inspel*, 20(1986)1, S. 47-67. *Automatic indexing; Fulltext; Descriptor; Method; Comparison; Automatic translation; Trend; AIR; PASSAT; CTX* 86-03025
- Selfridge, M.: A computer model of child language learning. *Artificial intelligence*, 29(1986)2, S. 171-216. *Natural language processing; Learning; Intelligence; Social groups; Model; Computer program CHILD* 86-03046
- Klenk, U. (Hrsg.): DGfS (Veranst.): Strukturen und Verfahren in der maschinellen Sprachverarbeitung. Deutsche Gesellschaft für Sprachwissenschaft (DGfS), Passau: Jahrestagung 1985. *Dudweiler: AQ-Verl.* (1985) 176 S. *Computational linguistics; Automatic translation; German; French; Text analysis; Syntax; Morphology; Semantics; Knowledge representation; Data structure; Phrase-structure grammar ASCOF; COAT; OVERCOAT* 86-03124
- Klenk, U. (Hrsg.): Kontextfreie Syntaxen und verwandte Systeme. Vorträge eines Kolloquiums in Ventron (Vogesen) im Oktober 1984. Tübingen: Niemeyer (1985) 154 S. *Text analysis; Language; Syntax; Context; Grammar; Computational linguistics; Automatic translation SUSY* 86-03125
- Panyr, J.: Automatische Klassifikation und Information Retrieval. Anwendung und Entwicklung komplexer Verfahren in Information-Retrieval-Systemen und ihre Evaluierung. Tübingen: Niemeyer (1986) 428 S. *Automatic classification; Clustering; Classification method; Information retrieval system; Search strategy; Search result*; 86-03141
- Kuhlen, R.; Univ. Konstanz, Sozialwissenschaftliche Fakultät, Konstanz: Informationelle Absicherung von Verwaltungshandeln durch wissensbasierte Informationsverarbeitung bei der Administration von Existenzgründung. Ein Positionspapier. Konstanz, März(1986) 31 S. *Decision support system; Model; Information needs; Expert system; Frame; Public administration; Business administration; Federal Republic of Germany; Baden-Wuerttemberg* 86-03153
- Beghtol, C.: Bibliographic classification theory and text linguistics Aboutness analysis, intertextuality and the cognitive act of classifying documents. *Journal of documentation*, 42(1986)2, S. 84-113. *Classification method; Class; Hierarchy; Computational linguistics; Text analysis; Comparison* 86-03163
- Schweda-Nicholson, N.: A United Nations interpreter survey. (Part 1). *The specialist/generalist controversy*. *Multilingua*, 5(1986)2, S. 67-80. *Translator; Subject-oriented; Trend; Inquiry; Corporation; International* 86-03169
- Löffler-Laurian, A. M.: Fast post editing and conventional post editing. (Part 1). *Two modes of a special activity (franz.)*. *Multilingua*, 5(1986)2, S. 81-88. *Automatic translation; Error; Translation; Comparison; English; French; SYSTRAN* 86-03170
- Tucker, M.: Expert systems blaze trails to AI success. *Mini-micro systems*, *Newton* 19(1986)4, S. 69-78. *Expert system; Microcomputer; Mainframe; Software; Market review* 86-03175
- Pigott, I. M.: Current Systran developments at the EC Commission. *Terminologie et traduction*, (1986)1: numero special, S. 10-16, en, ISSN 0256-7873 *Automatic translation; Effectiveness; Planning; EC SYSTRAN* 86-03230

- Evans, A. D. C.: Systran - the translator's viewpoint. S. 17-23. *Automatic translation; EC; Translator; English; French; Word processing SYSTRAN 86-03231*
- Habermann, F. W. A.: Provision and use of raw machine translation. S. 29-39. *Automatic translation; Effectiveness; Statistical survey; Literature; Nuclear technology SYSTRAN 86-03232*
- Mendez, J.: Machine translation in bureau service. S. 48-53. *Automatic translation; Translation service; Evaluation; Effectiveness; Translator Mendez Translations; SYSTRAN 86-03233*
- Siebenaler, L.: SYSTRAN for ESPRIT and ECAT bureau service. S. 54-60. *Translation service; Automatic translation; Online; Project; EC ECAT; SYSTRAN; ESPRIT 86-03234*
- Akazawa, E.: Systran Japanese systems. S. 78-89. *Automatic translation; Japanese; English; Text analysis SYSTRAN 86-03235*
- Walker, P. A.: The importance of word processing in the user environment. S. 102-111. *Automatic translation; Requirement; User; Word processing; Translation service; EC 86-03236*
- Pahl, T.: Ensuring compatibility in the Systran environment. S. 112-117. *Automatic translation; Compatibility; Unification; EC-countries SYSTRAN 86-03237*
- Homer, J.: Improving machine translation quality or eliminating the 'invisible idiot'. S. 118-145. *Automatic translation; Effectiveness; Software; Structure; Syntax; Semantics; Dictionary compilation SYSTRAN 86-03238*
- Guazzo-Jansen, J.: Dictionary coordination. S. 152-172. *Automatic translation; Consistency; Dictionary compilation; Method SYSTRAN 86-03239*

Buhlmann, R.: Merkmale geschriebener und gesprochener Texte im Bereich naturwissenschaftlich-technischer Fachsprachen. Eine Betrachtung unter didaktischen Gesichtspunkten. *Fachsprache, Wien, AT 7(1985)3/4, S. 98-125. Terminological research; Text; Language; Morphology; Syntax; Vocabulary; Comparison; Natural sciences; Engineering 86-03350*

Kucera, A.; Vollnhals, O.: Compact mit dem Computer. Maschinelle Umkehr eines zweisprachigen Fachwörterbuches. *Lexicographica, Tübingen (1986)2, S. 120-126. Lexicology; Dictionary; Computer aided; Preparation; Automatic translation; German; English; Natural sciences; Engineering; Software Compact 86-03363*

Schmuck, A.; Voigt, W.: Zweisprachige elektronische Wörterbücher Am Beispiel des Langenscheidt-Wörterbuchs 'alpha B'. *Lexicographica, Tübingen (1986)2, S. 284-290. Dictionary; Microcomputer; German; English; Use; Evaluation; Electronic publishing 86-03364*

Learned Information. 9th International online information meeting London, December 1985. 497 S. *Dialog service; Information retrieval; Search result; Comparison; Software; Microcomputer; Downloading; Information industry; Man-machine communication; Marketing; Trend 86-03423*

Williams, P. W.: The design of an expert system for access to information. S. 23-29. *Expert system; Information retrieval; Microcomputer; Software Userlink 86-03426*

Morita, A.; Sato, M.; Nishida, R.: Preparation of an online English language database for Japanese scientific and technical information. S. 61-67. *Dialog service; Technical information; Automatic translation; Japanese; English; Information network; Planning; Japan JICST; JOIS 86-03429*

Dierich, E.: Zur Anwendung von Methoden und technischen Mitteln der Mustererkennung für die Erschließung von Fachtexten. *Informatik, 33(1986)2, S. 66-67. Text analysis; Pattern recognition; Text; Structure 86-03467*

Schramm, R.; Drews, H.; Biela, K. D.: Grundlagen für Patentdatenbanken. Teil 2. Besonderheiten der Patentdatenbasis und ihrer Bearbeitung. *Informatik, 33(1986)2, S. 73-80. Patent documentation; Database; Patent; Relation; Patent classification; Automatic indexing; Evaluation; German Democratic Republic 86-03469*

Schramm, R.; Drews, H.; Biela, K. D.: Grundlagen für Patentdatenbanken. Teil 3. Aufbau ausgewählter Patentdatenbanken. *Informatik, DD 33(1986)3, S. 115-120. Patent documentation; Database; Automatic indexing; Text analysis; German; Testing; German Democratic Republic EPD; TEX 86-03475*

Rösner, H.: Generierung von Erklärungen aus Wissensrepräsentationen. *LDV-Forum, 4(1986)1, S. 3-19. Expert system; Natural language processing; Text analysis; Grammar; Model; Knowledge representation; Prototype; Federal Republic of Germany Nadorf; TWAICE 86-03492*

Görz, G.; Beckstein, C.: Unification-based speech parsing with a chart. *LDV-Forum, 4(1986)1, S. 20-26. Speech processing; Grammar; Syntax; Graphic representation EVAR; GuLP 86-03493*

Univ. Hamburg, Forschungsstelle für Informationswissenschaft und Künstliche Intelligenz, Hamburg; Univ. Hamburg, Fachbereich Informatik, Hamburg: Sprachorientierte KI-Projekte in Hamburg. *LDV-Forum, 4(1986)1, S. 27-32. Artificial intelligence; Natural language processing; Text analysis; Project; Federal Republic of Germany; Hamburg LOKI; MARVIN; WISBER 86-03494*

Hahn, U.; Hammwöhner, R.; Reimer, U.; Thiel, U.: TOPIC II/TOPOGRAPHIC II. Zur Entwicklung eines wissensbasierten Volltext-Informationssystems am Lehrstuhl für Informationswissenschaft der Universität Konstanz. *LDV-Forum, 4(1986)1, S. 33-37. Expert system; Fulltext; Information technology; University; Text analysis; Compression; Text; Automatic abstracting; German; Computer graphics; User interface; Knowledge representation; Prototype; Federal Republic of Germany; Baden-Wuerttemberg 86-03495*

Univ. des Saarlandes, Fachbereich Angewandte Mathematik und Informatik, Saarbrücken: Projekte 'XTRA', 'VITRA' und 'SC' an der Universität Saarbrücken. *LDV-Forum, 4(1986)1, S. 38-45. Natural language processing; Text analysis; Expert system; Finance; Image analysis; Transportation; Sports; Operating system; Project; Federal Republic of Germany; Saarland 86-03496*

- Seelbach, D.: Zum propositionalen und textuellen Wissen für das Verstehen französischer Wetterberichte. LDV-Forum, 4(1986)1, S. 46-62. *Terminology; Meteorology; Text analysis; Grammar; Automatic translation; French; German; Draft* 86-03497
- Brustkern, J.; Frackenpohl, G. Textbearbeitung mit dem Personal Computer. Ein Erfahrungsbericht über das BYU Concordance system. LDV-Forum, 4(1986)1, S. 63-67. *Information retrieval system; User support; User friendly; Microcomputer; Information retrieval; Search strategy; Concordance; Text analysis; Computational linguistics; Linguistics; Lexicology; USA BYU Concordance System* 86-03498
- Krzak, M.; Boras, D.: Lexical data base of the Croatian literary language. Informatologia Yugoslavica, 17(1985)3/4, S. 223-242. *Database; Terminology; Draft; Yugoslaviana* 86-03502
- Dierich, E.: Investigations of phrase-combining connections in German technical texts. International forum on information and documentation, 11(1986)2, S. 8-12. *Text analysis; Algorithm; Automatic indexing; Electrical engineering* 86-03510
- Dean, T. L.: Handling shared resources in a temporal data base management system. Decision support systems, 2(1986)2, S. 135-143. *Systems research; Database system; Use; Access; Hierarchy; Inference* 86-03548
- Stecher, P.; Hellemaa, P.: An "intelligent" extraction and aggregation tool for company data bases. Decision support systems, 2(1986)2, S. 145-158. *Systems research; Database system; Relation; Expert system; Implementation; Man-machine communication; Software SMARTY* 86-03549
- Somers, H. L. (Hrsg.): EUROTRA. Multilingua, 5(1986)3: special issue, S. 125-177. *Automatic translation; Computational linguistics; Software; EC; Project; History; Trend; EC-countries EUROTRA* 86-03553
- Moore, G. W.; Riede, U. N.; Polacek, R. A.; Miller, R. E.; Hutchins, G. M.: Group theory approach to computer translation of medical German. Methods of information in medicine, 25(1986)3, S. 176-182. *Automatic translation; Computer program; Literature; Publication; Medicine; German; English; USA; Maryland; Federal Republic of Germany; Baden-Wuerttemberg* 86-03577
- Ruske, G.; Weigel, W.: Automatische Erkennung fließender Sprache auf der Basis von Halbsilben. Informationstechnik, Fellbach, 28(1986)4, S. 227-235. *Speech recognition; Speech processing; Pattern recognition; Phonetics; Method* 86-03626
- Guethner, F.; Lehmann, H.: Verarbeitung natürlicher Sprache. Ein Überblick. Informatik-Spektrum, 9(1986)3, S. 162-173. *Natural language processing; Computational linguistics; Syntax; Semantics; Pragmatics; Text analysis; Information retrieval* 86-03640
- Gaul, W. (Hrsg.); Schader, M. (Hrsg.): Gesellschaft für Klassifikation: Classification as a tool of research. Proceedings of the 9th annual meeting, Karlsruhe, June, 1985. Amsterdam: North-Holland (1986) 515 S. *Classification research; Thesaurus research; Data analysis; Clustering; Documentary language; Artificial intelligence; Software* 86-03644
- Händler, H.: Probleme der inhaltlichen Erschließung von Dokumenten in internationalen Verbundsystemen. Mitteilungen der Gesellschaft für Bibliothekswesen und Dokumentation des Landbaues, Stuttgart (1986)38, S. 59-83. *Content analysis; Indexing; Foreign language; Semantics; Search result; Information exchange; International* 86-03671
- Doszkocs, T. E.: Natural language processing in information retrieval. Journal of the American Society for Information Science, 37(1986)4, S. 191-196. *Information retrieval; Man-machine communication; Language; Speech processing* 86-03687
- Williams, M. E.: Transparent information systems through gateways, front ends, intermediaries, and interfaces. Journal of the American Society for Information Science, 37(1986)4, S. 204-214. *Systems research; Dialog system; Man-machine communication; Information retrieval; Software; User friendly; Optical disc; Trend* 86-03689
- Zadeh, L. A.: Test-score semantics as a basis for a computational approach to the representation of meaning. Literary and linguistic computing, 1(1986)1, S. 24-35. *Computational linguistics; Knowledge representation; Semantics; Testing; Algorithm* 86-03717
- Piotrowski, R.: Text processing in the Leningrad Research Group 'Speech Statistics' Theory, results, outlook. Literary and linguistic computing, 1(1986)1, S. 36-40. *Automatic translation; Semiotics; Model; Artificial intelligence* 86-03718
- Gunzenhäuser, R.; Knopik, T. Neuere Entwicklungen des rechnerunterstützten Lernens. In: Maurer, H. (Hrsg.): Überblicke Informationsverarbeitung 1985. Mannheim: Bibliograph. Inst. (1985) S. 99-118. *Computer aided learning; Programmed instruction; Artificial intelligence; Software; Trend* 86-03724
- Leedham, C. G.; Downton, A. C.: On-line recognition of Pitman's handwritten shorthand. An evaluation of potential. International journal of man-machine studies, 24(1986)4, S. 375-393. *Manuscript; Text analysis; Transcription; Data input; DP; Conversion; Online; Software; Evaluation; Testing* 86-03794
- Goldman, K. J.; Goldman, S. A.; Kanellakis, P. C.; Zdonik, S. B.: ISIS: interface for a semantic information system. SIGMOD record, 14(1985)4, S. 328-342. *Systems research; Man-machine communication; User support; Graphic representation; Semantic network ISIS* 86-03818
- Robertson, S. E.; Thompson, C. L.; Macaskill, M. J.; Bovey, J. D.: Weighting, ranking, and relevance feedback in a front-end system. Journal of information science, 12(1986)1/2: Special issue, S. 71-75. *Information retrieval; Man-machine communication; Search result; Weighting; Ranking; Relevance; Feedback; Evaluation; Software* 86-03829
- Lamersdorf, W., Gesellschaft für Informatik (GI): Semantische Repräsentation komplexer Objektstrukturen. Modelle für nichtkonventionelle Datenbankanwendungen. Berlin: Springer (1985) 196 S. *Systems research; Database; Model; Semantic network; Data structure; Database organization; Office organization; Prototype* 86-03897
- Motro, A.: Constructing queries from tokens. SIGMOD record, 15(1986)2, S. 120-131. *Systems research; Database organization; Man-machine communication; Database system; Frame; Draft* 86-03978

Mattox, A. (Hrsg.): AFIPS: National computer conference 1986: Dynamic dimensions of computing, implications for management, professionals and users. June 1986, Las Vegas. Reston: AFIPS Press (1986) 580 S. Artificial intelligence; Software engineering; Computer network; Network architecture; Compatibility; User interface; Natural language processing; Software; Hardware 86-03922

Thompson, B. H.; Thompson, F. B.: Shifting to a higher gear in a natural language system. S. 349-354. Expert system; Natural language processing; Man-machine communication; Testing REL: POL 86-03934

Ballard, B. W.; Lusth, J. C.: An English-language processing system that 'learns' about new domains. S. 405-412. Natural language processing; English; Semantics; Office organization 86-03936

Busemann, S.; Höppner, W.; Marburger, H.; Morik, K.: Representing and processing copula and full-verb sentences in HAM-ANS. In: Stoyan, H. (Hrsg.): GI: GWAI-85. 9th German Workshop on Artificial Intelligence, Dassel/Solling, September 1985. Berlin: Springer (1986) S. 187-196. Natural language processing; Knowledge representation; German HAM-ANS 86-03943

Wirth, R.: Halbautomatische Erweiterung eines Thesaurus. S. 204-211. Natural language processing; Dictionary compilation; Computer aided; Text analysis 86-03944

Bergmann, H.; Paeseler, A.: Wissensakquisition für das natürlichsprachliche Zugangssystem HAM-ANS. S. 295-299. Knowledge representation; Frame; Natural language processing HAM-ANS; IVAS = Interaktives Wissensakquisitionssystem 86-03949



Frankfurter Linguistische Forschungen

Frankfurter Linguistische Forschungen (FLF) ist eine halbjährlich erscheinende Zeitschrift, die der Linguistik an der Universität Frankfurt ein Forum bietet. Inhaltlich vertreten sind Syntax, Phonologie, Morphologie, Logische Form, Psycho- und Neurolinguistik, Methodologie der Linguistik und Computerlinguistik. Jedes Heft enthält zwei längere Beiträge zu den o.a. Bereichen, sowie Kurzartikel zu Forschungsprojekten (z.B. Dissertationen), Berichte über Gastvorträge, Neuigkeiten aus dem Bereich den klinischen Linguistik und Rezensionen. Darüberhinaus erscheinen in unregelmäßiger Abfolge FLF-Sondernummern; die erste – eine Einführung in die Government-Binding-Theorie – erscheint im Herbst 1987.

Ein Blick in das Inhaltsverzeichnis von FLF2 (April 1987):

Theorie

Monika Klein: Überlegungen zu einer phonologischen Markiertheitstheorie

Arbeitsgruppe Psycholinguistik und Aphasieforschung: Referentielle Strategien und die Struktur des mentalen Lexikons – Evidenz aus der Aphasie

Forschungsprojekte

Angela Heuser: Neurolinguistische Aspekte transkortikaler Aphasien

Dagmar Wiegand: Kontaminationen und die Struktur des Sprachprozessors

Alicja Sakaguchi: Internationale geplante Sprachen Arbeitsgruppe Psycholinguistik und Aphasieforschung: Kontrolltheorie – Evidenz aus der Aphasie

Klinische Linguistik

Ulrike Kling-Lünser, Claudia Neubert, Margot Rudolf, Michaela Zeh: Prinzipien einer linguistisch orientierten Aphasie-therapie am Beispiel einer Geriatrischen Rehabilitationsklinik

Claudia Neubert, Margot Rudolf, Michaela Zeh: Berufsbild "Klinischer Linguist"

Gastvorträge

Franz-Josef Stachowiak: Kortikale Funktionen und Spracherwerb

Neuerscheinungen

Mary-Louise Kean (Hrsg.): Agrammatism

Noam Chomsky: Barriers

News

Herausgeber: Prof. Helen Leuninger, Institut für deutsche Sprache und Literatur II, Johann Wolfgang Goethe Universität, Grafstr. 76, 6000 Frankfurt II

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 datenstruktur-orientierter 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

Veranstaltungen

1987

JULI

22.6. - 3.7. DUBROVNIK

Sommerschule für Psycholinguistik, Inter-University Center of Postgraduate Studies, Frana Bulica 4, YU-50000 Dubrovnik

28.6. - 4.7. ELSENEUR

14ieme Colloque International de Linguistique Fonctionnelle, Societe International de Linguistique Fonctionnelle, C.P.H.E IVE Section, 45, rue des Ecoles, F-75005 Paris

29.6. - 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, Wueellnerstr. 3, D-5100 Aachen

29.6. - 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

1. - 3. LINZ

GI/ÖGI/SI-Fachtagung Entwurf von Informationssystemen, Informationsbedarfsermittlung und -analyse, Prof. Dr. Traunmüller, Universität Linz, Institut für Informatik, A-4040 Linz

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 Catani, Centro Internazionale di Semiotica e de Linguistica, Piazza del Rinascimento 7, I-61029 Urbino

11. - 13. MONTGOMERY

International Conference on Data Bases in the Humanities and the Social Sciences, Dr. Lawrence J. McCrank, AUM Library and Resource Center, Auburn University at Montgomery, Montgomery, Alabama, AL 36193-0401

13. - 17. SEATTLE, WA.

AAAI-87, 6th National Conference on Artificial Intelligence, Kenneth D. Forbus, Co-Programm Chairman, Science Program Qualitative Reasoning Group, University of Illinois, 1304 W. Springfield Avenue, Urbana, Ill. 61801

13. - 17. KARLSRUHE

14th International Colloquium on Automata, Languages, and Programming (14th ICALP) Gesellschaft für Informatik e.V. (GI), 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 Kin-

dersprache, Bodie Mattensson, Institut für Linguistik, Universität Lund, Helgenbacken 12, S-22362 Lund

20. - 24. BRISTOL

3rd International Conference on Social Psychology and Language, Prof. Anne Mallitte, School of Education, University of Bristol, 35 Berkeley Square, Bristol BS8 1JA, Great Britain

27. - 31. MÜNCHEN

IFAC '87, 10th World Congress on Automatic Control, Verein Deutscher Ingenieure IFAC '87, Abt. Tagungsorganisation, Postfach 1139, Graf-Recke-Straße 84, 4000 Düsseldorf 1

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

28.7. - 7.8. OSLO

2nd Advanced Course in Artificial Intelligence, ACAI '87, P.O. Box 5030, Majorstua, N-0301 Oslo 3

AUGUST

10. - 14. QUEBEC, KANADA

16. Internationaler Kongress für Onomastik, Prof. Dr. Rosemarie Gläser, Universität Leipzig, Sektion Fremdsprachen, Karl-Marx-Platz 9, DDR-7010 Leipzig

10. - 15. BERLIN (DDR)

14. Internationaler Linguistenkongress, Akademie der Wissenschaften der DDR, Otto-Nuschke-Str. 22-23, Postfach Linguistenkongress, DDR-1086 Berlin

10. - 15. HONOLULU

Second International Conference on Human-Computer Interaction, Gavriel Salvendy, Hawaii 1987, School of Industrial Engineering, Purdue University, West Lafayette, IN 47907

16. - 21. SYDNEY

AILA Congress (Association Internationale de Linguistique Appliquee), GAL-Sekretariat, Universität Trier, Postfach 3825, D-55500 Trier

17. - 19. TORONTO, KANADA

Workshop on Knowledge Representation, Management and Utilization in the Office, Dr. W. Lamersdorf, IBM Europ. Zentrum für Netzwerkforschung, Tiergartenstr. 15, D-6900 Heidelberg

17. - 19. VANCOUVER, B.C.

Second International Workshop on Natural Language Understanding and Logic Programming, Patrick Sain-Dizier, Dept. of Computing Science, Simon Fraser University, Burnaby, B.C., V5A 1S6 Canada

17. - 22. ANTWERPEN

International Pragmatics Conference, Jef Verschueren, Linguistics, University of Antwerpen, Universiteitsplein 1, B-2610 Wilrijk

VERANSTALTUNGEN

23. - 29. MAILAND

IJCAI 1987, 10th International Joint Conference on Artificial Intelligence, Organizing Secretariat, MGR, Piazza S. Ambrogio, 16, 20123 Milano, Italy

24. - 27. TRIER

I.C.H.O.L.S.IV., Geschichte der Sprachwissenschaft - Prof. Dr. Hans-Werner Niederer, I.C.H.O.L.S.IV, FB II, Universität Trier, Postfach 3825, D-5500 Trier

SEPTEMBER

1. - 4. STUTTGART

INTERACT '87 International Federation for Information Processing (IFIP), INTERACT '87 Secretariat, FB-1AG, Holzgartenstr. 17, D-7000 Stuttgart 1, Federal Republic of Germany

4. - 6. DUBLIN

First European Conference on Logo in Education, EUROLOG '87, Herbert Lötke, Jägerstr. 18, D-7053 Rummelshausen

7. - 9. EDINBURGH

Summer Conference on Category Theory and Computer Science, Prof. Dr. H.-D. Ehrlich, Institut für Programmiersprachen und Informationssysteme, Abteilung Datenbanken, TU Braunschweig, Postfach 3329, D-3300 Braunschweig

8. - 11. BUDAPEST

IFIP Conference on Governmental and Municipal Information Systems, IFIP-TC8 Conference Secretary, John von Neumann Society for Computing Sciences, Postfach 240, H-1368 Budapest

8. - 11. STRASBURG

ESEC '87, 1st European Software Engineering Conference, AFCET, Secretariat of ESEC 87, 156, Boulevard Perrier, 75017 Paris

9. - 12. AACHEN

6. Aachener Symposium für Signaltheorie - mehrdimensionale Signale und Bildverarbeitung, ASST '87, ASST '87, RWTH Aachen, Lehrstuhl für Melntechnik, D-5100 Aachen

14. - 17. PORTSMOUTH

13th Symposium on Microprocessing and Microprogramming, EUROMICRO 87, Art. Mrs. Ch. Snippe-Marlisa, 118 Twente, Gebouw TW/RC, P.O. Box 217, NL-7500 AE Enschede

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 Republic of Germany

21. - 26. GIESSEN

14. Deutscher Kongress für Philosophie, Dr. Winfried Franzen, Institut für Philosophie, Universität Gießen, Postfach, D-6300 Gießen

22. - 25. WIEN

Österreichische Artificial Intelligence Tagung 1987, Österreichische Gesellschaft für Artificial Intelligence, "ÖGAI-Tagung 1987", Postfach 177, 1014 Wien

28.9. - 2.10. GESEKE

GWAI 87, 11. Fachtagung über Künstliche Intelligenz, Dimitris Karagiannis, TU Berlin, Sekr. FR 5-8, Franklinstr. 29/29, 1000 Berlin 10

29.9. - 1.10. ERLANGEN

4. GI/NTG-Fachtagung, Sonja Wiffler, Institut für Mathematische Maschinen und Datenverarbeitung (Inform. IV), Martenstr. 3, 8520 Erlangen

29.9. - 1.10. TRIER

International Congress on Terminology and Knowledge Engineering, Prof. Dr. Dr. Hans Czup, Universität Trier, Postfach 3825, 5500 Trier

29.9. - 1.10. BRAUNSCHWEIG

9. DAGM-Symposium Mustererkennung Informatik, Prof. Dr.-Ing. E. Paulus, Institut für Nachrichtentechnik, TU Braunschweig, Schleinitzstr. 23, 3300 Braunschweig

OKTOBER

1. - 3. HEIDELBERG

Jahrestagung der Gesellschaft für Angewandte Linguistik, GAL-Sekretariat, Universität Trier, Postfach 3829, D-5500 Trier

1. - 3. HEIDELBERG

Angewandte Linguistik und Computer, 18. Jahrestagung der Gesellschaft für Angewandte Linguistik, GAL-Geschäftsstelle, Universität Duisburg, Postfach 10162, D-4100 Duisburg 1

5. - 7. AMSTERDAM

Working Conference on Distributed Processing, IFIP Secretariat, 16 Place Longemalle, CH-1204 Genf
ESSEN

5. Kongress der Deutschen Gesellschaft für Semiotik (DGS), Prof. Dr. Werner Enninger, FB 3, Universität GHS Essen, Universitätsstraße 2, D-4300 Essen

15. - 18. PENSACOLA

12th Annual Meeting of the Semiotic Society of America, Semiotic Society Secretariat, P.O. Box 10, Bloomington 47402, U.S.A.

16. - 17. DARMSTADT

Begriffsanalyse und Linguistik, Prof. Dr. Rudolf Wülfel, FB Mathematik, Technische Hochschule, Schloßgartenstr. 7, D-6100 Darmstadt

19. - 27. GENÈVE

TELECOM '87, 5th World Telecommunication Forum, VDE-Zentralstelle Tagungen, Stresemannallee 29, D-6000 Frankfurt/M. 70

20. - 23. MÜNCHEN

GI '87, 17. Jahrestagung der Gesellschaft für Informatik, Computerintegrierter Arbeitsplatz im Büro, Prof. Dr. Wilfried Brauer, TU München, Institut für Informatik, Postfach 202420, D-8000 München 2

20. - 27. MÜNCHEN

GI-Kongress auf der SYSTEMS 1987 "Wissenschaftliche Systeme", Prof. Dr. Wilfried Brauer, TU München, Institut für Informatik, Postfach 202420, D-8000 München 2

NOVEMBER

6. - 7. FAIRFAX, VA

Symposium on International Cultural Perspectives in Literature and Language. Prof. Jeffrey T. Chamberlain, Dept. of Foreign Languages and Literatures, George Mason University, Fairfax, VA 22030, U.S.A.

15. - 17. DORTMUND

Datenbanken für Software-Engineering, Software-Engineering für Nicht-Standard-Datenbanken. Volker Dirwehn, Universität Dortmund, FB Informatik, LS Software-Technologie, Postfach 500500, D-4600 Dortmund 50.

19. - 20. SAN FRANCISCO

Eleventh Western Educational Computing Conference. Judah Rosenwald, Extended Education, San Francisco State University, 1600 Holloway, San Francisco CA 94132, U.S.A.

DEZEMBER

WALZBURG

8. Symposium der Österreichischen Gesellschaft für Semiotik (OGS). Mag. Gloria Withalm, Viktoriagasse 24 B/45, A-1150 Wien

5. - 8. GENT

Conference on Communication and Cognition. Prof. Jan Van Diermael, CC 20, Blandijnberg 2, B-9000 Gent

7. - 9. PITTSBURGH

11th Annual International Conference on Information Systems for Management Sciences, ACM SIGBDP (Business Data Processing). William R. King, Graduate School of Business, University of Pittsburgh, Pittsburgh, PA 15260, U.S.A., Tel.: 412/648-1587

1988

JANUAR

1. - 8. CANBERRA

1st FIPLV-Weltkongress, FIPLV Head Office, Seestraße 247, CH-8038 Zürich

FEBRUAR

15. - 17. GÖTTINGEN

Neue Ansätze in der computerlinguistischen Morphologie und Phonologie. Dr. U. Klenk, Seminar für Romanische Philologie der Universität Göttingen, Nikolausberger Weg 23, D-3400 Göttingen

MÄRZ

6. - 10. ZÜRICH

1988 International Zurich Seminar on Digital Communications. P. Günzburger, Secretariat IZS 88, Haslerstr. 1, TDS, Belpstrasse 23, CH-3000 Bern 14

10. - 12. SAARBRÜCKEN

Computerlinguistik und ihre theoretischen Grundlagen. Jahrestagung 1988 der GLDV. J. Haller, IAI/LEBROTRA-D, Martin-Luther Str. 14, D-6600 Saarbrücken

8. - 13. CHICAGO

22nd Annual TESOL Conference. TESOL Central Office, 201 D.C. Transit Building, Georgetown University, Washington D.C. 20057, U.S.A.

MAI

5. - 8. PARIS

Symposium über Semiotik und Epistemologie Berlin. Prof. Dr. Roland Posner, Institut für Linguistik, TU Berlin, Sekr. TEL 6, Ernst-Reuter-Platz 7, D-1000 Berlin 10

JULI

8. - 22. URBINO

Semiotisches Sommerinstitut. Christina Calani, Centro Internazionale di Semiotica e de Linguistica, Piazza del Rinalcemento 7, I-61029 Urbino

24. - 29. LAUSANNE

European Conference on Computers in Education ECCE. Prof. Bernard Lévrat, ECCE 1988, Centre Universitaire d'Informatique, 12, rue du Lac, CH-1207 Genéve

24. - 31. ZAGREB

Symposium über Konismus der Syntax, anlässlich des 12. Internationalen Kongresses für anthropologische und ethnologische Wissenschaften. Prof. Marze E. Landsberg, Štikovna Straße 1, Bat. Galim, Hata 55014, Israel

AUGUST

17. - 21. BRIGHTON

18. Weltkongress für Philosophie. Conference Service, World Congress of Philosophy, 15 Bute Street, GB London, SW7 3EY

22. - 27. MÜNCHEN

12. Kongress der Internationalen Gesellschaft für Vergleichende Literaturwissenschaft, Raum und Grenzen. Prof. Dr. Roger Bauer, Abteilung für Vergleichende Literaturwissenschaft, Universität München, Schellingstr. 3, D-8000 München 40

22. - 27. BUDAPEST

Coling '88, 12th International Conference on Computational Linguistics, Coling '88 Secretariat, c/o MTE57 Congress Bureau, Kossuth ter 6-8, 1055 Budapest

SEPTEMBER

20. - 23. GUNZBURG

Interdisziplinäres Symposium über Text und Bild, Bild und Text. Prof. Dr. Georg Jäger, Institut für Deutsche Philologie, Universität München, Schellingstr. 3, D-8000 München 40

OKTOBER

5. - 7. ST. CHARLES, IL

Workshop on Spatial Reasoning and Multi-Sensor Fusion. Suixiang Chen, Dept. of Computer Science, University of North Carolina, Charlotte, NC 28223

17. - 20. BEIJING

IAPR - 9th International Conference on Pattern Recognition, 9ICPR Secretariat, Chinese Association of Automation, P.O. Box 2728, Beijing, China

Arbeitskreise

RECHTLICHE PROBLEME BEI DER NUTZUNG VON LEXIKOGRAPHISCHEN UND TERMINOLOGISCHEN DATENBANKEN FÜR WISSENSCHAFTLICHE ZWECKE [1]

1 Einleitung

Die Nutzung von Datenbanken wirft eine Fülle rechtlicher Fragestellungen auf. Einige dieser Probleme sind bereits weitgehend aus der wissenschaftlichen Diskussion herausgewachsen und haben ihren Niederschlag in speziellen gesetzlichen Regelungen gefunden. Ich nenne hier nur den Datenschutz, der in zahlreichen Ländern gesetzlich verankert ist. Andere Probleme der Datenbanknutzung können dagegen nur unter Rückgriff auf generelle Regelungen gelöst werden. Beispielhaft verweise ich auf das zivilrechtliche Vertrags- und das Urheberrecht. Die jeweiligen Gesetze, vor allem das Bürgerliche Gesetzbuch und das Urheberrechtsgesetz (UrhG) beruhen zudem in weiten Teilen auf einem Verständnis, das von der heutigen Rechtswirklichkeit weit entfernt aus der Vergangenheit herrührt. Ziel dieser Regelungen ist aber in jedem Fall ein gerechter Interessenausgleich zwischen den Beteiligten, also zwischen den Vertragsparteien bzw. zwischen Urheber und Werknutzer, und dadurch behalten sie ihre Gültigkeit. Die Umsetzung der normativen Bestimmungen auf den konkreten Fall der Nutzung lexikographischer und terminologischer Datenbanken möchte ich Ihnen für den Bereich des Urheberrechts demonstrieren. Ich werde mich dabei auf eine überblicksartige Darstellung beschränken. Grundlage des Beitrags ist das Urheberrecht der Bundesrepublik Deutschland; in einem eigenen Kapitel werde ich aber auch auf einige Grundsätze der internationalen Urheberrechtskonventionen eingehen.

2 Problemstellung

Die urheberrechtliche Problematik lexikographischer und terminologischer Datenbanken läßt sich am besten an Hand konkreter Fragestellungen erläutern. Dabei erscheinen mir folgende Aspekte besonders wichtig zu sein:

- (1) Sind lexikographische und terminologische Informationen urheberrechtlich geschützt?
- (2) Inwieweit ist die Übernahme solcher Informationen aus anderen Wörterbüchern in eine Datenbank bzw. das manuelle oder maschinelle Extrahieren dieser Informationen aus Texten urheberrechtlich zulässig?
- (3) Wem stehen die urheberrechtlichen Verwertungsrechte zu, wenn die Verfasser lexikographischer Informationen als Arbeitnehmer für den Arbeitgeber, z.B. den Datenbankenproduzenten, tätig werden?
- (4) Dürfen geschützte Werke in eine andere Datenstruktur umcodiert oder auf andere Datenträger, z.B. CD-ROM, überspielt werden?
- (5) Genießen lexikographische und terminologische Datenbanken Urheberrecht?
- (6) Ist die Übermittlung der gespeicherten Informationen auf Abruf urheberrechtlich relevant?

- (7) In welchem Umfang dürfen diese Informationen vom Nutzer frei verwendet werden? Welche Bedeutung haben Nutzungsverträge für die Weitergabe der Informationen an Dritte?
- (8) Welche Konsequenzen ergeben sich aus den internationalen Urheberrechtskonventionen für die Nutzung von Datenbanken?

Zum besseren Verständnis seien vorab die Begriffe "lexikographische Information" und "terminologische Information", so wie sie im folgenden verwendet werden, näher erläutert.

- (1) Unter einer lexikographischen Information verstehe ich eine Erläuterung zu einem Stichwort, die nicht nur den Begriff selbst erläutert, sondern u.U. weitere Informationen zu dem Stichwort gibt, z.B. die geschichtlichen Entwicklungen des "Absolutismus" oder die rechtliche Relevanz der "Wilenserklärung" beschreibt.
- (2) Eine terminologische Information beschränkt sich dagegen auf eine Begriffserläuterung.

In beiden Fällen sind die Informationen in knapper fachsprachlicher Form abgefaßt

3 Urheberrechtliche Probleme lexikographischer und terminologischer Datenbanken

3.1 Urheberrechtlicher Schutz lexikographischer und terminologischer Informationen

Als Schriftwerke sind die genannten Informationen grundsätzlich dem Urheberschutz zugänglich. Urheberschutz genießen sie aber nur dann, wenn sie eigenschöpferische geistige Leistungen im Sinne des § 2 Abs. 2 UrhG darstellen. Dies ist der Fall, wenn das Werk als eine Entäußerung der Persönlichkeit des Urhebers erscheint, d.h. diese muß sich in dem Werk widerspiegeln, etwa in einem eigentümlichen gestuften Sprachstil oder bei einer Sammlung in der schöpferischen Anordnung, Auswahl und Zusammenstellung der Einzelbeiträge. Dann gründen die Individualität des Werkes und damit die engen geistigen Beziehungen des Urhebers zu seinem Werk. Dementsprechend schützt das Urheberrecht den Urheber in seinen geistigen und persönlichen Beziehungen zum Werk und in der Nutzung des Werkes, § 11 UrhG. Dabei beschränkt sich der Urheberschutz auf Form und Inhalt eines Werkes [2]. Geschützt sind dagegen nicht z.B. wissenschaftliches und technisches Gedankengut [3], auch wenn dieses erst in mühevoller Arbeit gewonnen wurde [4]. Begründet ist diese Wertung dann, daß wissenschaftliche Ideen und Erkenntnisse frei und jedermann zugänglich sind [5]. Aber auch Form und Gestaltung eines wissenschaftlichen Werkes sind nur in engen Grenzen schutzfähig. Einschränkungen ergeben sich z.B. hinsichtlich des verwendeten Sprachstils [6]. Denn eine eigenschöpferische Gedankenformung und -führung des dargestellten Inhalts liegt dann nicht mehr vor, wenn die Gedankenführung sich aus der Natur der Sache ergibt und die im jeweiligen wissenschaftlichen Fachbereich übliche Ausdrucksweise verwendet wird [7]. Diese Grundsätze machen deutlich, daß eine urheberrechtliche Beurteilung lexikographischer und

terminologischer Informationen nicht pauschal sondern nur im konkreten Einzelfall vorgenommen werden kann. Die Kürze der Begriffserläuterungen wird allerdings nur wenig Raum für eine fachsprachlich individuelle Gestaltung des Textes lassen. So etwa dann, wenn das Verfassen lexikographischer Informationen eine tiefe Durchdringung der zu beschreibenden Sachverhalte und deren Analyse voraussetzt, um sie knapp und leicht verständlich wiederzugeben [8]. In diesen Fällen liegt es nahe, von einer eigenständigen schöpferischen Leistung auszugehen. Ob dies auch für terminologische Informationen gilt, erscheint im Hinblick auf die gebotene Kürze sowie die Verwendung der allgemein gebräuchlichen Fachsprache fraglich. Zu beachten ist freilich auch hier, daß Urheberrecht u.U. schon dann gewährt wird, wenn ein anderer Verfasser die Begriffserläuterung möglicherweise anders abgefaßt hätte [9].

3.2 Übernahme lexikographischer bzw. terminologischer Informationen

Soweit lexikographische bzw. terminologische Informationen urheberrechtlich geschützt sind, stellt deren Speicherung in maschinenlesbarer Form eine Vervielfältigung im Sinne des 16 Abs. 1 UrhG dar. Sie ist damit grundsätzlich dem Urheber vorbehalten und nur mit dessen ausdrücklich oder stillschweigend erteilter Zustimmung zulässig.

Ausnahmen vom Zustimmungserfordernis bestehen allerdings im Rahmen der gesetzlichen Schranken des Urheberrechts. Als solche kommen hier die Zitierfreiheit nach 51 UrhG sowie die Vervielfältigungsfreiheit nach 53 Abs. 2 UrhG in Betracht.

Nach 51 Nr. 2 UrhG dürfen aus einem selbständigen Werk einzelne Textstellen in das eigene Werk aufgenommen werden. Voraussetzung ist allerdings, daß die Vervielfältigung einem Zitatzweck dient, z.B. der Erläuterung dieser Stellen oder der Auseinandersetzung mit einer darin geäußerten Auffassung [10]. Es ist nicht zulässig, das zitierte Werk um seiner selbst willen dem Leser zur Kenntnis zu bringen. So liegen die Dinge aber, wenn lexikographische oder terminologische Informationen nur übernommen werden, um sie für den Aufbau einer lexikographischen bzw. terminologischen Datenbank zu verwenden. Eine Vervielfältigung zu diesem Zweck wird von 51 UrhG nicht mehr gedeckt. Dabei ist unerheblich, ob die Informationen aus Wörterbüchern oder aus fremden Datenbanken übernommen oder aus anderen Texten extrahiert werden.

Zulässig wäre die Vervielfältigung allerdings unter dem Gesichtspunkt der nach 53 Abs. 2 UrhG gestatteten Vervielfältigungsfreiheit zum eigenen (internen) Gebrauch. So dürfen beispielsweise kleine Teile eines Buches, etwa bis zu 20 Datenbankanfragen auszuwerten. Die Vervielfältigungsstücke dürfen jedoch nicht an Dritte weitergegeben werden, 53 Abs. 5 UrhG. Unzulässig ist ferner die Übernahme geschützter Werke aus einer anderen Datenbank in die eigene, um sie Nutzern zum Abruf anzubieten. Denn auch eine solche Vervielfältigung dient nicht dem ausschließlich eigenen Gebrauch [11]; sie ist daher nur mit Zustimmung des Berechtigten erlaubt.

3.3 Verwertungsrechte im Arbeitsverhältnis

Werden lexikographische oder terminologische Informationen im Rahmen eines Arbeitsverhältnisses erstellt, so stehen die Verwertungsrechte an den geschaffenen Werken gemäß 43 UrhG dem Arbeit-

geber zu. Insoweit ist von einer stillschweigenden Übertragung der Verwertungsrechte auszugehen, allerdings nur in dem Umfang, in dem sie für die betriebliche Auswertung erforderlich ist [12]. Benötigt beispielsweise ein Datenbankanbieter die Verwertungsrechte nur für die Zwecke des Online-Datenbankangebotes, so gehen gem. 43 UrhG nicht auch die Rechte zur Herausgabe gedruckter Informationsdienste auf ihn über.

Diese Regelung gilt freilich nicht für freie Mitarbeiter, die auf Grund eines Werkvertrages mit dem Datenbasenproduzenten tätig werden. Sie erhalten eine Vergütung nicht für ihre Tätigkeit als solche sondern für die abgelieferten Ergebnisse. Die freien Mitarbeiter bleiben daher auch grundsätzlich Inhaber der Urheberverwertungsrechte. Der Datenbasenproduzent kann diese Rechte nur auf Grund ausdrücklicher vertraglicher Vereinbarung erwerben, in der auch geregelt werden müßte, in welchem Umfang die Verwertungsrechte auf ihn übergehen. So könnte sich der freie Mitarbeiter z.B. das Recht vorbehalten, auch Dritten, etwa anderen Datenbasenproduzenten oder Verlagen, Nutzungsrechte einzuräumen. Er könnte aber auch seine ausschließlichen eigentumsähnlichen [13] Verwertungsrechte auf den Produzenten vollständig oder beschränkt übertragen mit der Folge, daß insoweit [14] allein dieser über die weitere Nutzung entscheiden kann.

3.4 Überspielung oder Umcodierung der auf Datenträger gespeicherten Informationen

Dem Urheber ist gemäß 16 UrhG das ausschließliche Vervielfältigungsrecht vorbehalten. Dieses Recht beinhaltet die Vervielfältigung eines Werkes gleichwie in welcher Form und in welcher Zahl. Es kommt nicht darauf an, ob die maschinenlesbar gespeicherten Informationen wiederum auf maschinenlesbare Datenträger überspielt werden. Auch gedruckte Texte, die in eine Datenverarbeitungsanlage eingelesen werden, sind vervielfältigt, ebenso wie die Speicherung von Informationen etwa auf einer CD-ROM (Compact Disk-Read Only Memory) oder die Speicherung in veränderter Datenstruktur eine Vervielfältigung darstellt. Dabei ist die zeitliche Dauer der körperlichen Fixierung ohne Belang [15].

Sollen urheberrechtlich geschützte Werke in diesem Sinne vervielfältigt werden, bedarf es daher grundsätzlich der Einwilligung des Berechtigten, es sei denn, die Vervielfältigungsstücke werden zum rein internen Gebrauch angefertigt. 53 Abs. 2 UrhG. Wie ausgeführt, ist in diesen Fällen eine Einwilligung entbehrlich, wenn nur kleine Teile eines Werkes vervielfältigt werden.

3.5 Urheberschutz von Datenbanken

Datenbanken bestehen aus einer Vielzahl strukturierter und nach formalen Kriterien gespeicherter Informationen. Durch die thematische Ausrichtung einer Datenbank wird zugleich die Auswahl der zu speichernden Informationen mitbestimmt. Darüberhinaus wird der Datenbankinhalt durch die Auswahl der Quellen sowie durch die mehr oder weniger vollständige Auswertung dieser Quellen bestimmt. Als solche Quellen kommen z.B. fremde Datenbanken, Wörterbücher, Zeitschriften, Fachbücher usw. in Betracht.

Von daher stellt eine Datenbank dann ein Sammelwerk im Sinne des 4 UrhG dar, wenn die Zusammenstellung der Informationen eine eigenschöpferische, weil nach den Vorstellungen und

dem Verständnis des Datenbasenproduzenten erstellte Leistung beinhaltet [16]. Als solche genießt sie urheberrechtlichen Schutz. Das gleiche gilt für die Fortschreibung, d.h. das periodische, z.B. monatliche Updating, einer Datenbank.

3.6 Übermittlung gespeicherter Informationen auf Abruf

Besonderes Kennzeichen der Online-Nutzung von Datenbanken ist die trägerlose, unkörperliche Übermittlung elektronisch gespeicherter Informationen. Urheberrechtlich relevant ist die Datenübermittlung dann, wenn sie als unkörperliche öffentliche Wiedergabe im Sinne des 15 Abs. 3 UrhG anzusehen ist. Das bedeutet, sie muß für eine Mehrzahl von Personen bestimmt sein und der Kreis dieser Personen darf weder bestimmt abgegrenzt sein noch dürfen die Personen durch gegenseitige Beziehungen oder durch Beziehung zum Veranstalter persönlich untereinander verbunden sein.

Festzustellen ist insofern, daß sich das Datenbankangebot an die Allgemeinheit wendet und von jedem genutzt werden kann, der über ein Datengerät verfügt und zur Datenbanknutzung berechtigt ist. Die Zahl der Nutzer ist prinzipiell unbeschränkt; persönliche Beziehungen unter den Nutzern bzw. zum Datenbankanbieter sind - auch bei geschlossenen Benutzergruppen - eher zufällig. Von daher dürften die Voraussetzungen des urheberrechtlichen Öffentlichkeitsbegriffs für die Datenübermittlung erfüllt sein [17].

Diese Ansicht wird jedoch von der herrschenden Meinung in Literatur und Rechtsprechung nicht geteilt, da die Wiedergabe nicht wie beim Rundfunk zeitgleich eine Vielzahl von Teilnehmern erreicht [18]. Danach ist die Datenübermittlung also urheberrechtlich irrelevant.

Es mehren sich aber die Stimmen, die eine zeitliche Kumulierung fordern, d.h. die Datenbankabfrage als einheitliches Ereignis sehen wollen [19]. Danach richtet sich die Übermittlung der Daten an eine Öffentlichkeit im Sinne des 15 Abs. 3 UrhG.

3.7 Nutzung der abgerufenen Informationen

Läßt der Nutzer sich Abrufinformationen auf seinem Terminal anzeigen, liegt darin eine nicht-öffentliche Wiedergabe, es sei denn, die Abrufinformationen werden etwa im Rahmen einer Demonstration einem größeren Personenkreis vorgeführt. Eine solche Wiedergabe ist gegen Zahlung einer angemessenen Vergütung frei zulässig, 52 Abs. 1 Satz 2 UrhG.

Werden die Informationen hingegen ausgedruckt oder beim Nutzer maschinenlesbar gespeichert (sog. Downloading [20]), so liegt darin eine urheberrechtsrelevante Vervielfältigung der abgerufenen Werke. Diese darf er im Rahmen der gesetzlichen Bestimmungen des 53 UrhG frei nutzen, d.h. er kann sie für den eigenen wissenschaftlichen oder sonstigen eigenen (internen) Gebrauch verwenden. Er darf die Informationen aber ohne Erlaubnis des Berechtigten weder öffentlich wiedergeben noch verbreiten, indem er sie z.B. in Printform oder als Magnetbanddienst der Öffentlichkeit anbietet oder in Verkehr bringt. Aber auch für den eigenen Gebrauch dürfen nur einzelne Vervielfältigungsstücke, d.h. bis zu sieben Exemplare, angefertigt werden.

Die in der Praxis der Literaturdatenbankanbieter üblichen Vereinbarungen unterscheiden allerdings nur zwischen der Verwendung für den eigenen Gebrauch und für geschäftliche, besser gewerbliche

Zwecke und beschränken die Vervielfältigung zum eigenen Gebrauch z.B. auf die Herstellung von Hardcopies, indem sie andere, etwa elektronische Speicherformen untersagen. Lediglich Informationsvermittlern wird in der Regel die Weitergabe recherchierter Informationen an deren Auftraggeber im Rahmen eines konkreten Rechercheauftrags gestattet.

3.8 Rechtslage nach dem internationalen Urheberrecht

International hat das Urheberrecht vor allem durch die Berner Übereinkunft vom 09.09.1886 und das Welturheberrechtsabkommen vom 06.09.1952, beide zuletzt am 24.07.1971 revidiert, weltweite Anerkennung erlangt. Vor allem in der Revidierten Berner Übereinkunft (RBU) ist das Urheberrecht weit entwickelt worden. Ihr gehören u.a. mit Ausnahme Albaniens und Andorras alle europäischen Staaten an. Von den außereuropäischen Ländern möchte ich nur die Sowjetunion und die USA nennen, die bislang nur das Welturheberrechtsabkommen unterzeichnet haben [21].

Wesentlich ist beiden Konventionen das sog. (Inländerprinzip). Dieses besagt, daß die Angehörigen eines Mitgliedslandes in jedem anderen Mitgliedsland die dort geltenden Urheberrechte in Anspruch nehmen können. Darüberhinaus gewährt die Berner Konvention sog. (Mundestrechte), d.h. Angehörige eines Mitgliedsstaates können sich in einem anderen Mitgliedsstaat auch auf die Vorschriften der Berner Konvention berufen. Von Interesse ist dies vor allem dann, soweit diese weitergehende Rechte enthalten.

Beide Konventionen gewähren dem Urheber das Vervielfältigungsrecht, setzen allerdings den nationalen Gesetzgebern für Einschränkungen dieses Rechtes bestimmte Grenzen, Art. 9 Abs. 2 RBU, Art. IV bis Abs. 2 WUA.

Sofern Datenbanken nicht in gedruckter Form oder als Magnetbanddienst vertrieben, sondern nur online als Datenbanken angeboten werden, sind die gespeicherten Werke als nicht-öffentlich anzusehen. Denn als veröffentlicht gelten nur erschene, d.h. in Printform (Art. VI WUA) bzw. auch in anderer Form der Öffentlichkeit angebotene oder in Verkehr gebrachte Werkexemplare (Art. 3 Abs. 3 RBU) [22].

Die Datenübermittlung ist den Konventionen zufolge urheberrechtlich irrelevant, anders aber z.B. in Frankreich, wonach jede nicht private Wiedergabe als öffentliche Wiedergabe angesehen wird [23].

Besonderheiten bestehen für die USA und die UdSSR insofern, als sie der RBU noch nicht beigetreten sind. Die Mundestrechte können dort also nicht geltend gemacht werden. Für die USA kommen als weitere Besonderheiten die dort geltenden Formvorschriften hinzu. Danach ist jedes veröffentlichte Werk mit einem Copyright-Vermerk ("c" im Kreis, Angabe des Berechtigten und des ersten Erscheinungsjahres) zu versehen.

Die lediglich in einer Datenbank maschinenlesbar gespeicherten Informationen brauchen deshalb prinzipiell auch nicht mit einem Copyright-Vermerk versehen zu werden; ein solcher Vermerk ist allerdings erforderlich, soweit die Datenbank auch in Printversion publiziert und in den USA vertrieben werden soll. Darüberhinaus hat der Copyright-Vermerk auch in den übrigen Staaten zumindest eine Warn- und Hinweisfunktion.

4 Schlußbemerkung

Zusammenfassend können wir feststellen:

- (1) Lexikographische und terminologische Informationen sind als Schriftwerke grundsätzlich dem Urheberrecht zugänglich. Geschützt sind Form und Gestaltung der Gedankenführung und des Inhaltes. Aus der Kürze der Informationen, der Natur der Sache sowie der Verwendung der allgemein gebräuchlichen Fachsprache können sich aber Einschränkungen ergeben.
- (2) Die Übernahme terminologischer oder lexikographischer Informationen aus anderen Werken (z.B. Datenbanken, Wörterbüchern und Zeitschriften) in die eigene Datenbank stellt eine erlaubnispflichtige Vervielfältigung dar. Aus der gesetzlich geregelten Zitier- und Vervielfältigungsfreiheit können nur in engen Grenzen Ausnahmen vom Zustimmungserfordernis abgeleitet werden.
- (3) Die Verwertungsrechte an Werken, die in Erfüllung dienstlicher Verpflichtungen in einem Arbeitsverhältnis geschaffen wurden, gehen auf den Arbeitgeber über, soweit die Rechte für die betriebliche Auswertung benötigt werden. Freie Mitarbeiter bleiben regelmäßig Inhaber der Verwertungsrechte an den von ihnen geschaffenen Werken, es sei denn, sie haben ihre Rechte ausdrücklich auf den Auftraggeber übertragen.
- (4) Die Überspielung von maschinenlesbar gespeicherten Werken auf andere Datenträger oder die Speicherung der Werke in einer anderen Datenstruktur stellt grundsätzlich eine erlaubnispflichtige Vervielfältigung dar.
- (5) Datenbanken genießen als Sammelwerke urheberrechtlichen Schutz, wenn die gespeicherten Informationen nach individuellen Kriterien zusammengestellt wurden.
- (6) Die Übermittlung gespeicherter Informationen im Wege der Datenfernverarbeitung vom Anbieter zum Nutzer stellt nach herrschender Lehre keine urheberrechtsrelevante Verwertungshandlung dar. In der Literatur wird aber verstärkt die Datenübermittlung als urheberrechtlich relevante öffentliche Wiedergabe angesehen.
- (7) Die Wiedergabe von abgerufenen Informationen auf dem Terminal des Nutzers ist urheberrechtlich relevant, wenn sie öffentlich geschieht. Ausdrucke und maschinenlesbare Speicherung beim Nutzer stellen eine erlaubnispflichtige Vervielfältigung dar, wenn die Vervielfältigungsstücke über die Grenzen der gesetzlichen Vervielfältigungsfreiheit hinaus genutzt werden.
- (8) International hat das Urheberrecht vor allem durch das Welturheberrechtsabkommen und die Revidierte Berner Übereinkunft Anerkennung erlangt. Beiden Konventionen liegt das Prinzip der Inländerbehandlung zugrunde. Darüberhinaus gewährt die Berner Übereinkunft sog. Mindestrechte. Neben dem Vervielfältigungsrecht ist vor allem der Veröffentlichungsbegriff von besonderer Bedeutung. Die Datenübertragung wird in den Konventionen nicht geregelt. Für die USA ergeben sich Besonderheiten aus den dort geltenden Formvorschriften für veröffentlichte Werke.

Nicht zuletzt dank der beiden Urheberrechtskonventionen hat sich der Gedanke vom Schutz des geistigen Eigentums, wenn auch in unterschiedlicher Intensität, in fast allen Teilen der Erde durchgesetzt. Allerdings bezieht sich das Urheberrecht nur auf die (schöpferische) geistige Leistung. Dementsprechend sind Datenbanken nur begrenzt dem Urheberrecht zugänglich. Das Urheberrecht wird daher dem Anliegen der Informationsanbieter, nämlich dem (generellen) Schutz ihrer Produkte, nur zum Teil gerecht. Dafür sind die lexikographischen und terminologischen Datenbanken ein bereites Beispiel. Eine befriedigende Lösung dieses Problems steht noch aus.

Anmerkungen

- [1] Deutschsprachige Version des am 16. 10. 1986 anlässlich des IV. Forum Information Science and Practice in Saarbrücken gehaltenen Vortrags. Die Vortragsform wurde beibehalten.
- [2] F.K. Fromm, Nordemann, Urheberrecht, 5. Auflage, Stuttgart u.a. 1983, 2. Ann. [8].
- [3] BGH DB (Der Betrieb) 1981, S. 1281; BGHZ (Amtliche Sammlung des Bundesgerichtshofs in Zivilsachen, Band, Seite) 73, S. 292; OLG Hamm GRUR (Gewerblicher Rechtsschutz und Urheberrecht) 1981, S. 131.
- [4] OLG Hamburg GRUR 1978, S. 308; BGH GRUR 1980, S. 231.
- [5] BGH DB 1981, S. 1281. Soweit es um technische Erkenntnisse geht, kommt hinzu, daß deren Schutz nach dem Urheberrechtsgesetz in Widerspruch zu den technischen Schutzrechten träte, vgl. nur BGHZ 73, S. 292.
- [6] OLG Hamm GRUR 1981, S. 131.
- [7] BGH DB 1984, S. 2028; BGH BB (Betriebs-Berater) 1986, S. 2012 f.
- [8] Vgl. dazu BGH BB 1986, S. 2012 f.; Appellationsgericht Paris Ur. v. 16.12.1985 Nr. 2, S. 5; vgl. auch für die Abstracts von Dokumenten J. Mehrings, Information und Dokumentation (IuD) - Ein Stiefkind der Urheberrechtsnovelle?, in: GRUR 1983, S. 287; M. Hackemann, Information und Dokumentation aus urheberrechtlicher Sicht - Einige Anmerkungen zur gegenwärtigen und künftigen Rechtslage, in: GRUR 1982, S. 269.
- [9] Mehrings, Anm. [8], S. 287.
- [10] Vgl. auch BGH NJW (Neue Juristische Wochenschrift) 1972, S. 22306.
- [11] Hackemann, Anm. [8], S. 266.
- [12] Fromm/Nordemann, Anm. [1], 43. Ann. [3b].
- [13] Vgl. BVerfGE (Amtliche Sammlung der Entscheidungen des Bundesverfassungsgerichts, Band, Seite) 31, S. 240 ff.
- [14] Z.B. nur für die Nutzung in elektronischer Form, nicht aber zur Weiterverwendung in Printform, oder nur zur Ausübung in der Bundesrepublik Deutschland, nicht aber in den übrigen europäischen Staaten, oder auch nur zeitlich begrenzt.
- [15] Fromm Nordemann, Anm. 1, 16. Ann. 1]. Vgl. auch J.W. Goebel m. Hackemann J. Scheller, Zum Begriff des Erscheinens beim elektronischen Publizieren, in: GRUR 1986, S. 355 ff.
- [16] Vgl. dazu BGH GRUR 1982, S. 39; BGH GRUR 1980, S. 230; BGH GRUR 1981, S. 521.
- [17] Goebel Hackemann/Scheller, Anm. [15], S. 357 f.
- [18] Vgl. Nachweise bei Hackemann, Anm. [8], S. 271 f.
- [19] P. Katzenberger, Urheberrechtsfragen der elektronischen Textkommunikation, in: GRUR Int. (Gewerblicher Rechtsschutz und Urheberrecht, Internationaler Teil) 1983, S. 904 f. m.w.N.
- [20] Zum Muster eines Downloading-Vertrages, vgl. J.W. Goebel, Die Gestaltung des Downloading-Vertrages, in: CuR (Computer und Recht) 1986, S. 73 ff.

[21] Vgl. im einzelnen M. Hackemann, Urheberrechtliche Aspekte des grenzüberschreitenden Datenverkehrs, in: Deutsche Gesellschaft für Dokumentation (Hrsg.), Deutscher Dokumentartag 1985, München 1986, S. 304 ff., Die USA wollen aber noch 1986 der Berner Konvention beitreten.

[22] Goebel Hackemann Scheller, Anm. [15], S. 360 f.

[23] Hackemann, Anm. [21], S. 314 f.

Martin Hackemann, *Gesellschaft für Information und Dokumentation mbH (GID), Abteilung für Rechtsfragen der Fachkommunikation, 6000 Frankfurt am Main 71, Lyoner Str. 44-48.*

AK "LDV UND NACHBARN"

Ein kurzer Rechenschaftsbericht und ein Appell

Der Arbeitskreis wurde auf der Tagung der LDV-Fittings 1980 in Saarbrücken ins Leben gerufen. Er begann seine Arbeit am 23.1.1981 und gab sich folgendes Arbeitsprogramm:

- 1) Zum Selbstverständnis von LDV/CL, AI/KI, Informationslinguistik (IL) und Informationswissenschaften (IW) und zu den Beziehungen dieser Selbstverständnisse untereinander;
- 2) Erfassung laufender Projekte und Untersuchung des Verhältnisses zwischen Absichtserklärung und Realisierung in praxi für die einzelnen Projekte;
- 3) Feststellung fehlender bzw. unterrepräsentierter Problemfelder (als Ableitung aus den Ergebnissen von 1) und 2));
- 4) Zum Transfer von experimentellen, weitgehend grundlagenforschungsorientierten Systemen hin zu Anwendungs-/praxisorientierten Systemen. (vgl. LDV-Fittings Info 02/81, S. 9)

Aus der Bearbeitung des ersten Punktes des Arbeitsprogramms entstand:

AK "LDV und Nachbarn", Materialien 1, Berlin/Bonn Sept. 1981 mit folgenden Beiträgen:

- Zum Begriff "Künstliche Intelligenz" (Arno Schmidt)
- Zum Begriff "Linguistische Datenverarbeitung" (Hans Dieter Lutz)
- Zum Begriff "Informationslinguistik" (Hans Dieter Lutz)
- Auswahlbibliographie zu den Bereichen LDV, KI, CL, IL und IW (Hans Dieter Lutz)

Diese Arbeiten boten auch den Anlaß, die Tagung der LDV-Fittings 1982 in Koblenz unter dem Titel "LDV und Nachbarn - ein Versuch einer Standortbedingung" auszurichten (vgl. I. Bator, J. Krause, H.D. Lutz (Hrsg.), *Linguistische Datenverarbeitung, Versuch einer Standortbestimmung im Umfeld von Informationslinguistik und Künstlicher Intelligenz*, Tübingen 1982 (= Sprache und Information, Bd. 4)). Nächster Arbeitsschwerpunkt war ab Mitte 1982 die Erstellung einer Dokumentation "Projekte zur maschinellen Verarbeitung sprachlichen Materials". Diese Arbeit erwies sich als sehr aufwendig und teilweise auch notwendig, da sich die eruierten Daten oft als überholt, unzutreffend oder unvollständig erwiesen.

Deshalb wurde ein Aufruf zur Projektdokumentation mit Fragebogenaktion gestartet; das war im August 1984. Veröffentlicht wurde er im LDV-Forum 2 (2), Dez. 1984, S. 26.

Ergebnis - nach langen Mühen und nach zusätzlicher Erhebungsarbeit - war dann die von G. Knorz anläßlich der COLING86 herausgegebene "Map on Current Research Projects in the FRG Related to the Field of Computational Linguistics" (Juli 1986; erscheint in der nächsten Ausgabe von Sprache und Datenverarbeitung, Niemeyer).

Der Punkt "Projekt-Evaluierung" aus 2) wie der Rest des Arbeitsprogramms konnten bislang noch nicht angegangen werden (außerdem wäre ein neuer Anlauf bzgl. Punkt 1) zu machen in Hinblick auf Sprachpsychologie, Kognitionswissenschaft, Informationslinguistik und Kommunikationsforschung). Warum? Der AK, der zu Anfang aus drei Mitgliedern (davon waren zwei aktiv), dann aus vier (von denen drei aktiv mitarbeiten konnten), dann sieben (von denen wiederum zwei aktiv waren) Mitgliedern bestand, ist inzwischen auf zwei Mitglieder geschrumpft, die selbst nicht mehr so aktiv im AK direkt arbeiten können, wie sie es gerne möchten. G. Knorz ist voll ausgelastet mit der Redaktion und Produktion des LDV-Forum, H.D. Lutz organisiert die Arbeitstreffen "LDV-Studiengänge".

Bevor der AK einschlafte, richten die letzten beiden AK-Mohikaner einen Appell an Sie zu überdenken, ob Sie in diesem AK nicht aktiv und damit auch das Fach gestaltend mitwirken wollen. Falls dies eine Perspektive ist, hier unsere Adressen:
Prof. Dr. G. Knorz, Fachhochschule Darmstadt, Fachbereich IuD, Schöfferstr. 8, 6100 Darmstadt
Dr. H.D. Lutz, EWH Rheinland-Pfalz, Abt. Koblenz, Rheinau 3-4, 5400 Koblenz-Oberweith

ARBEITSKREIS SPRACHERKENNUNG, SPRACHGENERIERUNG UND PHONETISCHE DATENBANK

Nachdem der Arbeitskreis als erste große Aufgabe die diesjährige Jahrestagung in Bonn organisiert hatte (siehe dazu die einschlägige Berichterstattung in diesem Heft), ist er gerade dabei, einen Workshop für den Herbst vorzubereiten, auf dem Interessierte die ersten experimentellen Ergebnisse der AK-Arbeit diskutieren sollen. Als einführender Aufsatz kann der Kurzbeitrag "Die Extremwertanalyse im Zeitbereich als Möglichkeit der automatischen Spracherkennung" von P. und T. Schweisthal (in der vorliegenden Ausgabe) angesehen werden.

Nähere Information sind vom AK *Spracherkennung, Sprachgenerierung und phonetische Datenbanken*, K.G. Schweisthal, Institut für Phonetik und sprachliche Kommunikation der Universität München, Schellingstraße 3 V6, D-8000 München 40 anzufordern.

Red.

VORSCHLÄGE ZUR KONTURIERUNG DES FACHES COMPUTERLINGUISTIK

Mit dem folgenden Text stellen Teilnehmer ([1]) der "Arbeitstreffen 'LDV-Studiengänge' das Ergebnis ihrer bisherigen Arbeit zur Diskussion und fordern dazu auf, kritisch dazu Stellung zu nehmen.

Dieser Text ist der erste Teil eines projektierten Arbeitspapiers, das folgende Gliederungspunkte umfassen soll:

0. Einleitung
1. Inhalte und Methoden des Faches Computerlinguistik
2. Fähigkeiten und Ausbildungsprofil von Computerlinguistik
3. Computerlinguistik als Diplomstudiengang
4. Computerlinguistik als
 - (a) Nebenfach
 - (b) Aufbaustudium
 - (c) Zusatzstudium
 - (d) Schwerpunkt
 - (e) Teil eines Studiums in einem anderen Fach

Vorgelegt werden die Ausführungen zu den Gliederungspunkten 0 bis 2. Die Punkte 3 und 4 wurden noch nicht formuliert, weil die Ergebnisse der gewünschten breiten Diskussion des hier vorgelegten Textes berücksichtigt werden sollen. Diskussionsbeiträge werden bis Ende September 1987 erbeten an: *Hans Dieter Lutz, Erziehungswissenschaftliche Hochschule, Abteilung Koblenz, Studiengang Angewandte Informatik, Schwerpunkt Linguistik, Rheinau 3-4, 5400 Koblenz-Oberwerth.*

0. Einleitung

Die Zahl der LDV/CL-Studiengänge nimmt ständig zu. Die Unterschiedlichkeit der fachlichen und institutionellen Einbindung und die Verschiedenartigkeit der Studiengangstypen mit entsprechend verschiedenen Abschlüssen erschwert Überschaubarkeit und Vergleichbarkeit der Studienangebote für

- Lehrende,
- Studienanfänger und fortgeschrittene Studierende,
- Studienberater und
- eventuelle Arbeitgeber.

Auch über Ausbildungsziele, Berufsqualifikation und Berufsprofil von CL-Absolventen bestehen keine festumrissenen, sondern allenfalls vage und differierende Vorstellungen.

Auf Anregung der Arbeitskreise "Ausbildung und Berufsperspektiven" sowie "LDV und Nachbarn" haben Beirat und Vorstand der GLDV im März 1986 Arbeitstreffen interessierter Mitglieder initiiert, deren Ziel es sein sollte, etablierte und im Aufbau begriffene Studiengänge, deren Inhalte, Ausbildungsprofile sowie die sich daraus ergebende Art von Berufsqualifikation zu analysieren. Langfristige Zielvorstellungen für dieser Arbeit waren es,

- Lehr- und Lernziele einzelner Lehrveranstaltungen zu beschreiben,
- fachinhaltliche Minimalvorstellungen (je Studiengangtypus zu entwickeln),

- die "Durchlässigkeit" (Studienort-Wechsel) zwischen einzelnen Studiengängen zu erhöhen oder überhaupt erst zu ermöglichen,
- das Ausbildungsprofil von Studiengängen herauszuarbeiten
- das Profil des Faches CL genauer zu konturieren und damit die Berufsqualifikation eines CL-Absolventen beschreibbar zu machen,
- dadurch zugleich die Überschaubarkeit für die Studienberatung zu verbessern,
- schließlich aber auch das fachspezifische Angebot auf dem Stellenmarkt zu analysieren.

Aus dem ersten Arbeitstreffen (am 30./31.8.1986 in Bonn) gingen zwei Arbeitsgruppen hervor, die sich näher mit den Themenkomplexen "Inhalte und Methoden der Computerlinguistik" (AG1) sowie "Fähigkeiten und Ausbildungsprofil von Computerlinguisten" (AG2) befaßten. Für die gestellte Aufgabe erwies es sich als zweckmäßig, einen Diplomstudiengang Computerlinguistik (ohne eine institutionelle Einbindung) als "tertium comparationis" anzusetzen, zu dem die existierenden oder geplanten Studiengänge dann in Beziehung gesetzt werden können. Einig war man sich auch über die Fachbezeichnung "Computerlinguistik" (CL).

1. Inhalte und Methoden des Faches Computerlinguistik Grundlagen

Computerlinguistik ist weder über eine Aufzählung einzelner Anwendungsgebiete (wie z.B. MÜ, Textverarbeitung, CUU) noch über - von den Nachbar-disziplinen Informatik und Linguistik her motivierte - stratifikationale Einteilungen des Sprachsystems (wie z.B. Phonologie, Morphologie, Syntax, Semantik) zu erfassen, sondern hat ihre eigene zentrale Problemstellung und Zielsetzung (vgl. *Ungeheuer 1971, S. 689*).

Grundsätzlich kann die Zusammenführung von Sprache auf der einen Seite und EDV auf der anderen Seite von zwei verschiedenen Zielsetzungen ausgehen:

- A: EDV wird in den Wissenschaften von der Sprache eingesetzt, um das (linguistische und/oder literaturwissenschaftliche) Wissen über Sprache zu mehren.
- B: Die maschinelle Verarbeitung natürlicher-sprachlich repräsentierter Daten dient zur Verbesserung maschineller Informationssysteme im Sinne eines abstrakten "Informationserschließenden Problemlösungssystems" (I-P- Systems) nach G. Ungeheuer.

Es gibt Auffassungen, die davon ausgehen, daß sich B direkt aus A ergibt. Das hieße z.B., daß ein zur Erforschung syntaktischer sprachlicher Strukturen entwickeltes Syntax-Analyse- Programm(system) unmittelbar auch zur Verbesserung maschineller Informationssysteme eingesetzt werden kann. In diesem Fall ist B keine eigenständige Zielsetzung von A, sondern die externe Anwendung der in A entwickelten Instrumente und Erkenntnisse. Als eine vermittelnde Begriffsbestimmung ergibt sich:

(CL1: Computerlinguistik ist die Wissenschaft von der maschinellen Verarbeitung natürlicher Sprache unter Einbeziehung linguistischen Wissens.)

Computerlinguistik in diesem Sinne umfaßt nicht allein die Implementierung von Lösungswegen für sprachliche Problemstellungen, sondern auch deren

Konzeption unter gezielter Berücksichtigung und angemessener Aufbereitung linguistischen und informatischen Wissens. Die Entwicklung des Lösungsweges wird bestimmt durch Forschungs- oder Anwendungsinteresse. Die Forschung ist mit dem Ziel zu betreiben, Voraussetzungen für anwendungsorientierte CL-Verfahren zu liefern.

Andere Auffassungen betonen die Eigenständigkeit von B sowohl in praktischer wie in theoretischer Hinsicht. Dabei geht man davon aus, daß eine Verbesserung maschineller Informationssysteme (im weitesten Sinne) mit Hilfe verbesserter Software Erkenntnisse aus der Linguistik, der Informatik, der KI-Forschung, der Informationswissenschaft, der Kognitionswissenschaft und anderer Disziplinen voraussetzt, die zu einer neuen Gesamtschau integriert werden müssen. Bei dieser Auffassung verändert sich notwendigerweise die theoretisch- und die praktisch-orientierte Betrachtungsweise aufgrund des anderen Umfeldes und der veränderten Zielsetzung.

Eine Computerlinguistik im Sinne von B, die sich nicht darin erschöpft, den Rechner nur als technisches Hilfsmittel für sprachwissenschaftliche Arbeiten einzusetzen, sondern die ihn vorrangig in seiner Funktion als Kommunikationsmedium sieht, muß von einer Theorie der Mensch-Maschine-Interaktion (MMI) ausgehen, die selbst wiederum zu einer Theorie der Mensch-Mensch-Kommunikation in Beziehung zu setzen ist.

Daraus ergibt sich, daß eine derartige Computerlinguistik weder eine Angewandte Linguistik noch eine Angewandte Informatik sein kann, sondern daß sie eine eigene Disziplin ist, die sich mit sprachlichen Äußerungen in deren Funktion des Mit-Teilens, also mit verbaler Kommunikation beschäftigt, auf deren einen Seite der Mensch und auf deren anderer Seite der Computer stehen.

Eine Computerlinguistik, welche die Interaktion Mensch-Maschine in den Mittelpunkt stellt, hat auch die Aufgabe, die sich aus dieser Interaktion selbst ergebenden Probleme zu diskutieren. Sie hat danach zu fragen, welche Typen von MMI in Abhängigkeit von Aufgabenstellungen in spezifischen Situationen und bezüglich der Anwendungsgebiete für den Menschen festgestellt werden können und ob diese Typen sich durch eine jeweils ihnen eigene Verbalität und Diskursivität kennzeichnen lassen, so daß im Umkehrschluß aus diesen spezifischen, habituellen Formen sprachlicher Äußerungen "Äußerungssequenzen zwischen Mensch und Computer Forderungen an die konversationelle und kooperative Kompetenz und Performanz des Computers abgeleitet werden können. Erst auf dieser Basis läßt sich Software entwickeln, die den Anforderungen einer natürlichsprachlich fundierten MMI (Ungeheuer 1971) in den jeweiligen Anwendungen zu genügen vermag.

Eine so verstandene Computerlinguistik ist praxisorientierte Grundlagenforschung, die ihre Anwendungen (nicht) außerhalb ihres Gebietes ansiedelt. Dieses Verständnis von Computerlinguistik führt zu folgender Begriffbestimmung:

(CL2: Computerlinguistik ist die Wissenschaft von der systematischen Verarbeitung natürlichsprachlicher Äußerungen im Rahmen der Mensch-Maschine-Interaktion.)

Natürlichsprachliche Mensch-Maschine-Interaktion hat hier nicht den Status eines Anwendungsgebietes unter vielen, sondern ist der zentrale Aspekt, unter dem jegliche Anwendung zu sehen ist.

Für CL1 wie CL2 gilt gleichermaßen: Analog etwa zur Informatik (mit ihren Nachbardisziplinen Mathematik und Physik) bezieht die Computerlinguistik einige ihrer Methoden und Inhalte aus ihren Nachbardisziplinen Linguistik und Informatik, die sie ihrer Zielsetzung entsprechend in ihren eigenen Methoden- und Inhalte-Kanon integriert. In diesem Sinne sind die folgenden Gruppierungen und Detailausführungen zu verstehen.

"VORFELD LINGUISTIK"

- 1) Grundlegende Methoden empirischen Arbeitens und Kenntnis ihrer Ergebnisse (aus deskriptiver Linguistik, quantitativer Linguistik und kontrastiver Linguistik)
- 2) Phonologie- und Syntaxmodelle
- 3) Semantik (lexikalische Semantik, Merkmalssemantik, Bedeutungstheorien ...)
- 4) Pragmatik (Sprechakt-Theorie, Textgrammatiken ...)
- 5) Theoriegeleitete Beschreibung ausgewählter sprachlicher Phänomene (Nominalphrasen, Quantifikation, Koordination, ...)
- 6) Intellektuelle Validierung von (Teil-)Theorien des Sprachsystems

"VORFELD INFORMATIK"

- 1) Grundlegende mathematische Methoden (Algebra, Graphentheorie, formale Logik, Modelltheorie, Statistik); informatische Grundlagen (Architektur eines Computers, Betriebssysteme, Theorie und Anwendung von Programmen, Programmdokumentation, Datenstrukturen, Algorithmentheorie)
- 2) Theorie formaler Sprachen, Automatentheorie
- 3) DB-Modelle; Wissensrepräsentationssprachen
- 4) Problemlösungsverfahren (Suchen, Planen, Dekomposition, schnelles Lernen)
- 5) Systemarchitekturen, praktischer Umgang mit Software-Tools
- 6) Testtheorie (Software-Validierung)

ZENTRALE CL-BEREICHE

- 1) Theorie der Mensch-Maschine-Interaktion (MMI)
 - Grundlagen aus Interaktionstheorie, Kommunikationsforschung und Kognitionswissenschaft
 - Simulation der Mensch-Maschine-Kommunikation und Besonderheiten der MMI gegenüber der Mensch-Mensch-Kommunikation
 - natürlichsprachlich fundierte MMI in Abhängigkeit von Aufgabenstellungen
 - gesellschaftliche Auswirkungen und Technologiefolgenabschätzung
- 2) Parsing-Verfahren, Parsing-Strategien und (grammatische) Spezifikationssprachen
- 3) Beschreibungsformalismen (ATN, semantische Netze, CD, conceptual graphs, frames, scripts ...) und Verarbeitungsprozeduren
- 4) Benutzermodelle (Zielverfolgungs-, Planungs-, Interessen-Konstrukte), Adaptivität, Argumentationsmodelle, Einbettung in Informationserschließende Problemlösungssysteme (I-P-Systeme); Transfer-Problemik (Überführung der theoriegeleiteten Beschreibung in konkrete, implementierte I-P-Systeme)

- 5) Architekturen von natürlichsprachlich fundierten I-P-Systemen praktischer Umgang mit CL-spezifischen Software-Tools
- 6) Systemevaluierung (Effektivität, Effizienz und empirische Adäquatheit implementierter I-P-Systeme)

2. Fähigkeiten und Ausbildungsprofil von Computerlinguisten

Vorbemerkung:

Der Inhalt dieses Kapitels ist auf dem Hintergrund eines CL-Hauptfachstudiums konzipiert.

Die "Fähigkeiten des Computerlinguisten" (Abschnitt 2.1.) beziehen sich auf seine Kompetenz nach Abschluß des Studiums.

Abschnitt 2.2. ist *nicht curricular* zu lesen.

Die Bezeichnung "Computerlinguist" steht stellvertretend für "Computerlinguist" und "Computerlinguistin".

2.1. Fähigkeiten des Computerlinguisten

Der Computerlinguist muß vorhandenes CL-Wissen adäquat einsetzen können. Das bedeutet:

- 2.1.1. Der Computerlinguist muß in der Lage sein zu entscheiden, ob ein sprachliches Problem auf dem gegenwärtigen Stand des Know-How mithilfe des Rechners sinnvoll zu lösen ist.
- 2.1.2. Der Computerlinguist muß die Fähigkeit haben, offene Problemfelder zu erkennen und neue Problemlösungen zu suchen.
- 2.1.3. Der Computerlinguist muß in der Lage sein, sprachliche Problemstellungen so aufzubereiten, daß sie sich mit dem Computer lösen lassen. Das bedeutet:
 - Er muß mit fachspezifischen Verfahren eine Lösung finden und dabei gezielt das vorhandene sprachliche und linguistische Wissen einsetzen.
 - In einem zweiten Schritt muß der Lösungsweg in einer für die maschinelle Verarbeitung geeigneten Form dargestellt werden.
 - Er muß den Lösungsweg in einem dritten Schritt selbständig ein Programm(-System) umsetzen können.
- 2.1.4. Der Computerlinguist muß in der Lage sein, Lösungen in d CL zu bewerten und diese Bewertung explizit zu begründen. Das setzt die Kenntnis von spezifischen Eigenschaften sowie von Vor- und Nachteilen unterschiedlicher Lösungswege voraus und die Kenntnis des innovativen Elementes bezüglich vorangehender Lösungen.
- 2.1.5. Der Computerlinguist muß in Software-Entwicklungsprojekte zwischen Linguisten und Informatikern vermitteln können. Das beinhaltet u.a.:
 - daß er terminologische Unterschiede der beteiligten Fachsprachen berücksichtigt und
 - daß er in der Lage ist, dem Programmierer den Weg der Programmierung kompetent zu erläutern.
- 2.1.6. Der Computerlinguist muß weiterhin die Fähigkeit haben, die sprachlich fundierten Teilaspekte von Anwenderproblemen zu erkennen, in Kooperation mit dem Anwen-

der zu beurteilen und gegebenenfalls eine Lösung zuzuführen.

- 2.1.7. Das Ausbildungsziel muß auch darin bestehen, den CL-Absolventen in den Stand zu setzen, das spezielle Berufsbild zu vermitteln.

2.2. Ausbildungsprofil

Das Ausbildungsprofil eines Computerlinguisten setzt sich zusammen aus

- a) Grundwissen bezüglich der Computerlinguistik mit ihren Vorfeldern Linguistik und Informatik und
- b) Vertiefungswissen, das in der Regel induziert wird durch einen Anwendungsbereich, der wiederum die Hinzunahme von Anwendungswissen und Inhalten eines "Nebengebietes" erfordern Anwendungswissen und Inhalten eines "Nebengebietes" erfordern kann.

Dieses Vertiefungswissen ist auf mögliche Berufsfelder bezogen.

Grundwissen und Vertiefungswissen werden ergänzt durch praktisches Erfahrungswissen (interne und externe Praktika, Übungen, u.ä.).

Vertiefungswissen kann in den folgenden Anwendungsbereichen (exemplarische, nicht vollständige Liste) erworben werden:

- Maschinelle Textverarbeitung
- Maschinelle Lexikographie / Maschinengestützte Lexikographie
- Information und Dokumentation
- Maschinelle Übersetzung / Maschinengestützte Übersetzung
- Natürlichsprachliche Frage-Antwort-
- Computerunterstützter Unterricht
- Entwicklung CL-spezifischer Tools

Grund- und Vertiefungswissen gemeinsam setzen den CL-Absolventen in den Stand, die Anforderungen des Fähigkeitenkataloges zu erfüllen.

Anmerkungen

[1] Mitarbeit haben bisher: I.S. Batori, R. Drewek, B. Endres-Niggemeier, A. Frantzke, T. Gardner, U. Hahn, P. Hellwig, M. Heyn, L. Hitznerberger, Kh. Jakob, B. Kelle, G. Knorz, W. Kreitmeyer, H.D. Lutz, W. Löser, M. Lutz-Hensel, P. Ovenhausen, H.P. Pütz, W. Putschke, U. Reyle, B. Schaefer, K.D. Schmitz, H. Schnelle, G. Schweisshal, G. Willke, K. Wothke

Literatur

Ungeheuer, Gerold 1971: Linguistische Datenverarbeitung - die Realität und eine Konzeption. In: IBM-Nachrichten 21, H. 206, S. 688-694 (Wiederabgedruckt in: DERS., Sprache und Kommunikation. [=KFP B 13]. Hamburg 1972, 191-197.)

Mitteilungen aus der GLDV

AUS DER ARBEIT DER GLDV

Mitgliederversammlung am 5.3.87 in Bonn

Erstmals in der Geschichte der GLDV war die Mitgliederversammlung 1987 nicht beschlußfähig, weil statt der erforderlichen 67 nur etwa 30 Mitglieder anwesend waren. Mit der Einladung zu der dann erforderlichen zweiten Mitgliederversammlung wurde das Protokoll an alle GLDV-Mitglieder verschickt. Es wurde eine Satzungskommission gebildet, die der Mitgliederversammlung 1989 eine Neufassung der Satzung zur Beschlußfassung vorlegen soll.

Beiratssitzung am 4.3.87 in Bonn

In seiner Sitzung vom 4.3.87 in Bonn befaßte sich der Beirat der GLDV mit folgenden Themen:

- Gutachterwahlen bei der DFG
- Planung eines DFG-Schwerpunktes "Verarbeitung sprachlich repräsentierter Information"
- Pläne zur Gründung einer Sektion "Computerlinguistik" in der DGFS
- LDV-Forum und Planung eines LDV-Newsletters
- Arbeitsprogramm der GLDV 1987/8

Die GLDV hat sich teilweise erfolgreich dafür eingesetzt, daß die DGFS und die GAL bei der Gutachterwahl in der DFG Computerlinguisten nominieren. Sie unterstützt diejenigen GLDV-Mitglieder, die auf Listen von Nachbarvereinigungen kandidieren. Im übrigen ist eine langfristige Aktivierungsstrategie erforderlich, um eine angemessene Berücksichtigung der Computerlinguistik in den DFG-Gremien zu erreichen.

Die Pläne zu einem DFG-Schwerpunkt "Verarbeitung sprachlich repräsentierter Information" wurden mit DFG-Vertretern erörtert. Ein entsprechender Antrag könnte in der Herbstsitzung 1988 des DFG-Senats behandelt werden. Die Arbeitsgruppe in der GLDV strebt ein größeres Arbeitstreffen aller interessierten Kolleg(inn)en an.

Wie sich die Gründung einer Sektion Computerlinguistik in der DGFS auf die Arbeit der GLDV auswirkt, ist in vielen Punkten nicht klar. Der Vorstand führt ein Gespräch mit den interessierten Computerlinguisten in der DGFS, um Kooperationsmöglichkeiten zu erkunden und für die Computerlinguistik schädliche Effekte zu vermeiden. Die GLDV erhöht im übrigen weiter ihre Attraktivität. Ein gewisses Defizit in der theoretischen Arbeit muß wettgemacht werden.

Ein Redaktionsstatut des LDV-Forums liegt im Entwurf vor. Der geplante LDV-Newsletter wird unverzüglich in Angriff genommen. Zumindest für den Anfang wird es eine Netz- und eine Papierversion geben.

Als Thema für die Jahrestagung 1988 schlägt der Beirat "Computerlinguistik und ihre theoretischen Grundlagen" vor. Als Orte kommen Stuttgart und Tübingen besonders in Frage, auch Saarbrücken ist möglich.

Mitgliederversammlung am 27.4.87 in Frankfurt

Bei der Mitgliederversammlung der GLDV am 27.4.87 in Frankfurt, die wegen der Beschlußfähigkeit der Mitgliederversammlung am 5.3.1987 in Bonn erforderlich geworden war, waren 75 Mitglieder anwesend oder vertreten. Der Vorstand berichtete über die Entwicklungen seit der letzten Mitgliederversammlung. Die Jahrestagung 1988 ("Computerlinguistik und ihre theoretischen Grundlagen") findet in Saarbrücken statt. Der mit der DGFS abgestimmte Termin ist die Woche vom 7.-12.3.1988. Die Tagung wird in Zusammenarbeit mit der Fachgruppe "Naturlich-sprachliche Systeme" der GL durchgeführt und eventuell mit der neugegründeten Sektion Computerlinguistik der DGFS durchgeführt.

Im Herbst 1988 soll eine Diskussion über die Lehrinhalte von Computerlinguistik Studiengängen stattfinden, eventuell in Zusammenarbeit mit der DGFS. Die GLDV arbeitet bei der Gestaltung der diesjährigen GAL-Tagung mit dem Rahmenthema "Angewandte Linguistik und Computer" mit. Die Pläne für eine DFG-Schwerpunkt "Verarbeitung sprachlich repräsentierter Information" werden unterstützt.

Die Wahlen zu Vorstand und Beirat erbrachten folgende Ergebnisse: Der Vorstand wurde mit 70 Stimmen bei 5 Enthaltungen entlastet. Im ersten Wahlgang wurden Brigitte Endrey-Niggemeyer (69 Stimmen) als Vorsitzende, Burkhard Schander (72 Stimmen) als Schatzmeister und Christine Schneider (70 Stimmen) als Schriftführerin wiedergewählt. Die Wahl des zweiten Vorsitzenden ergab im ersten Wahlgang

Klaus-Gunther Schwesithal	24 Stimmen
Johann Haller	19 Stimmen
Ursula Klenk	18 Stimmen
Wolfgang Putschke	5 Stimmen

In der laut Satzung erforderlichen Stichwahl setzte sich Klaus-Gunther Schwesithal mit 40 Stimmen gegenüber Johann Haller mit 20 Stimmen endgültig durch.

In den Beirat der GLDV wurden gewählt

Jürgen Krause	(59 Stimmen)
Istvan Baton	(57 Stimmen)
Hans-Dieter Lutz	(55 Stimmen)
Peter Hellwig	(52 Stimmen)
Rainer Kühlen	(52 Stimmen)
Christopher Habel	(48 Stimmen)
Raimund Drewiek	(42 Stimmen) oder
Dietmar Rosner	(42 Stimmen)

Dietmar Rösner und Raimund Drewek entscheiden selbst, wer von beiden zugunsten des anderen zurücktritt.

Brigitte Endres-Niggemeyer, FH Hannover

NEUE MITGLIEDER IN DER GLDV

Folgende Personen wurden auf der Vorstandssitzung im März 87 neu als Mitglieder in die GLDV aufgenommen:

- Frank Dingeldein, Mainz
- Klaus-Jürgen Engelberg, Stuttgart
- Clivia Gassner, Heidelberg
- Roland Hauser, Berlin
- Matias Heidemann, Freiburg
- Dr. Michael Hess, Zürich
- Dr. Bernhard Kelle, Freiburg
- Gregor Meder, Essen
- Prof. Dr. Georg F. Meier, Olching
- Marie Anne Moreau, Paris
- Guillaume Schiltz, Freiburg
- Elke Spielmanns-Rome, Bonn
- Gabriele Stallwitz, Dormitz
- Petra Steiner, Trier
- Holger Trojahn, Trier

Folgende Personen wurden auf der Vorstandssitzung im April 87 neu als Mitglieder in die GLDV aufgenommen:

- Thomas Barth, Heidelberg
- Stephanie Bohnert, Mainz
- Christoph Caspary, Frankfurt
- Kuno Dunhölder, Mannheim
- Christoph Gutknecht, Hamburg
- Konrad Jablonski, Delbrück
- Ute Jekosch, Herne
- Ernst Lutz, Darmstadt
- Heike Westermann, Frankfurt

Herr Wolfgang Mausberg ist aus der GLDV ausgetreten.

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.

DAS "LDV-FORUM" IM ABONNEMENT

Auch Nicht-Mitglieder der GLDV können jetzt das LDV-Forum regelmäßig beziehen. Dazu ist es notwendig, formlos ein Abonnement bei der Redaktion

Redaktion des LDV-Forum, Prof. Dr. G. Knorz, Fachhochschule Darmstadt, Fachbereich IuD, Schöfflerstraße 8, D-6100 Darmstadt

zu bestellen und den Preis der Einzel exemplare jeweils für das laufende Jahr auf das Konto

Gesellschaft für Linguistische Datenverarbeitung / LDV-Forum, Kto.-Nr. 554 090, BLZ 508 501 50,

Sparkasse Darmstadt

zu überweisen (Vorauskauf!). Für diese DM 30,- pro Jahr gehen Ihnen dann die beiden Hefte unmittelbar nach Erscheinen (ohne weitere Versandkosten) per Post zu. An der Tatsache, daß Sie den aktuellen Newsletter (wird gerade vorbereitet, siehe dazu das Editorial), den Rabatt auf die Reihe Linguistische Datenverarbeitung (OLMS) und vieles andere versäumen, sind Sie dann allerdings selbst schuld! Dazu müßten Sie dann doch Mitglied der GLDV werden (...was ich Ihnen in jedem Falle empfehlen kann).

Gerhard Knorz

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 hoffentlich längst geschehen) möglichst umgehend tun!

Burkhard Schaefer, Fachbereich 3, Universität Essen Gesamthochschule, Postfach 103764, D-4300 Essen 1

GUTACHTERWAHLEN BEI DER DFG

Der GLDV-Kollege Prof. Dr. W. Lenders kandidiert bei den Gutachterwahlen der DFG auf der Liste der Gesellschaft für Angewandte Linguistik (GAL). Die wahlberechtigten GLDV-Mitglieder können mit ihrem Votum dazu beitragen, daß möglichst fachlich einschlägige Kollegen über die Computerlinguistik-Anträge befinden können, die an die DFG gerichtet werden.

Brigitte Endres-Niggemeyer, FH Hannover

EDITORIAL	1
THEMENSCHWERPUNKT MASCHINELLE ÜBERSETZUNG	
Zum Themenschwerpunkt Maschinelle Übersetzung	4
Haller, J.: Forschungssysteme für maschinelle Übersetzung - ein Überblick	5
Haller, J.: EUROTRA - gegenwärtiger Stand und Planung 1987/88	8
Haller, J.; Reyle, U.: Statusseminar zu EUROTRA-D und Begleitforschung im Februar 1987	10
Schmidt, P.: Syntaktische Analyse deutscher Sätze im EUROTRA-Formalismus - Konfigurationale und relationale Struktur	12
Zelinsky-Wibbelt, C.: Die Überführung von relationalen Strukturbeschreibungen in die Interface-Struktur-Überlegungen zur EUROTRA Interface-Struktur	15
Steiner, E.: Grundprinzipien und Vorschläge für eine semantische Beschreibung von Argumentstrukturen - Überlegungen zur EUROTRA Interface-Struktur	17
Zelinsky-Wibbelt, C.: Lexikalische Disambiguierung mit semantischen Merkmalen - Überlegungen zur EUROTRA Interface-Struktur	19
Schütz, J.: EUROTRA-Software-Entwicklung	21
Der Aufbau mehrsprachiger Lexika mit dem Lexikonsystem MULI, Brustkern, J.	24
Semantik und Generierung - Bericht über den Zweiten Externen Workshop der EUROTRA-D-Begleitforschung, Hauenschild, C.; Busemann, S.	26
FACHBEITRÄGE	
Kipke, U.; Wille, R.: Formale Begriffsanalyse - erläutert an einem Wortfeld	31
Phonetische Beiträge zur maschinellen Spracherkennung:	
Schweisthal, P.; Schweisthal, Th.: Die Extremwertanalyse im Zeitbereich als Möglichkeit der automatischen Spracherkennung	37
Erklärung einiger technischer Grundbegriffe zur automatischen Spracherkennung. AK Sprach-erkennung, Sprachgenerierung und Phonetische Datenbanken	39
GLDV-JAHRESTAGUNG 1987	
Analyse und Synthese gesprochener Sprache	40
Bemerkungen zu den Vorträgen der Jahrestagung 1987 der GLDV, Meier, G.	41
TAGUNGSBERICHTE	43
Sprachwissenschaft im Computerzeitalter (43), Computereinsatz in der praktischen Lexikographie (45), Anwendung von Repräsentationstechniken (47), Anwenderworkshop des AUTOCAT-Projekts (48), Künstliche Intelligenz - Sprachforschung (50)	
NACHRICHTEN	51
Chancen für Philologen (51), Nichtnumerische Datenverarbeitung in den Geisteswissenschaften (51) Schwerpunkt Maschinelle Sprachverarbeitung an den Universitäten Stuttgart und Tübingen (52), Forschungszentrum für Künstliche Intelligenz (53), Offener Brief zur Einrichtung des KI-Zentrums (54), Parallelverarbeitungsmodelle für Problemlösungen in der KI (54)	
REZENSIONEN	
Lenders, W.; Willée, G.: Linguistische Datenverarbeitung - ein Lehrbuch, Schwanke, M.	55
Lustig, G. (Hrsg.): Automatische Indexierung zwischen Forschung und Anwendung, Schneider, Ch.	56
PUBLIKATIONEN	57
Neuere Literatur zur Linguistischen Datenverarbeitung (57), FLF - Frankfurter Linguistische Forschungen (65), Projekt SEMSYN - New Publications (18)	
VERANSTALTUNGEN	67
Veranstaltungskalender (67), Neue Ansätze in der computerlinguistischen Morphologie und Phonetik (70), Linguistik und Computer (GAL) (70), Theoretische Grundlagen der Computerlinguistik (70), Int. Conference on Data Bases in the Humanities and Social Sciences (70)	
ARBEITSKREISE	71
Rechtliche Probleme bei der Nutzung von lexikographischen und terminologischen Datenbanken für wissenschaftliche Zwecke (71), AK "LDV und Nachbarn" (75), Arbeitskreis Spracherkennung, Sprachgenerierung und phonetische Datenbanken (75), Vorschläge zur Konturierung des Faches Computerlinguistik (76), Kontaktadressen der Arbeitskreise (7)	
MITTEILUNGEN AUS DER GLDV	79
Aus der Arbeit der GLDV (79), Der Schatzmeister informiert, bittet, mahnt ... (80), Neue Mitglieder in der GLDV (80), Das LDV-Forum im Abonnement (80), Gutachterwahlen bei der DFG (80)	